TESIS DOCTORAL

2021

STATISTICAL MODELLING OF IN-GAME PLAYER BEHAVIOUR

ANA FERNÁNDEZ DEL RÍO

PROGRAMA DE DOCTORADO EN CIENCIAS

ELKA KOROUTCHEVA
ÁFRICA PERIÁÑEZ

ANA FERNÁNDEZ DEL RÍO

STATISTICAL MODELLING OF IN-GAME PLAYER BEHAVIOUR

PHD THESIS

STATISTICAL MODELLING OF IN-GAME PLAYER BEHAVIOUR

A dissertation submitted to attain the degree of

DOCTOR OF SCIENCES of UNED

presented by

ANA FERNÁNDEZ DEL RÍO
MSc. Statistical Physics of Complex Systems

Supervisors:

Dr. E. Koroutcheva
Dr. Á. Periáñez

2021

Ana Fernández del Río: *Statistical modelling of in-game player behaviour,* 2021

A mis padres, fuente inagotable de amor y aliento.

ABSTRACT

This thesis deals with the statistical modelling of in-game player behaviour. More generally, it lies within the framework of building a scientific body of knowledge around human affairs.

Video games allow for the display of many interesting traits concerning human behaviour within a controlled setup. They are nowadays played online, and each action of every player recorded, generating incredibly rich and detailed datasets. Therefore, in-game player activity is an ideal playground to put hypothesis concerning human behaviour at large to test.

Part of this work is concerned with the social character of games and how this affects decisions players are constantly making as they play. Borrowing methodology and tools from statistical physics, a formal theoretical approach is proposed as a framework to qualitatively understand the processes at play and provide insights into how player and choice interaction affect the average outcome of decision-making processes.

A more significant part of this work follows a data-driven mindset, covering several statistical and machine learning algorithms applied to the predictive modelling of different quantities of interest in the game. Player engagement and purchasing behaviour are the main focus, to which player conversion (from non-paying user to paying user), player attrition or churn, and purchase churn (when paying users cease to purchase) are used as proxies. They are extensively studied at different scales (or levels of aggregation) in the game. Results at different scales are relevant for different purposes, and can be used to complement and enhance each other, as will be discussed. Individual player behavioural predictions for the phenomena of interest are generated using decision forests, survival ensembles, and deep learning, and their performance compared. Several time series models are explored to predict group behaviour. The use of these predictions in player profiling is also discussed, as is a machine learning item recommendation system.

The research presented here has immediate practical applications. Understanding how players behave and why allows studios to design more engaging games and provides tools to optimise game planning. It opens the door to personalisation, as games can be developed and planned to cater to individual player tastes. From a more fundamental and ambitious perspec-

tive, this work intends to be a small contribution to laying the foundations of a mathematical understanding of human behaviour and societies.

RESUMEN

El objeto de esta tesis es la modelación estadística del comportamiento de los jugadores de videojuegos (dentro del juego). De manera más general, se inscribe en el marco de la construcción de un cuerpo científico de conocimientos sobre la actividad humana.

En los videojuegos se dan muchos rasgos interesantes del comportamiento humano dentro de una configuración controlada. Hoy en día se juegan en línea y se registra cada acción de cada jugador, lo que genera conjuntos de datos increíblemente ricos y detallados. El estudio de la actividad de los jugadores es, por lo tanto, una buena forma de poner a prueba algunas hipótesis sobre el comportamiento humano en general.

Parte de este trabajo analiza como el carácter social de muchos juegos afecta las decisiones que los jugadores toman constantemente mientras juegan. Tomando prestada metodología y herramientas de la física estadística, se propone un enfoque teórico formal como marco para comprender cualitativamente los procesos que intervenienen. Es posible así proporcionar información sobre cómo las interacciones, tanto entre jugadores como entre distinas decisiones, afecta al resultado colectivo de estos procesos de toma de decisiones.

Una parte más significativa de este trabajo se centra en el análisis y modelación a partir de datos. Se analiza la validez de distintos algoritmos estadísticos y de aprendizaje automático para predecir variables de interés en el juego. El nivel de actividad e implicación del jugador, así como las transacciones económicas que realiza dentro del juego, son el centro de este estudio. Éste se lleva a cabo a través de la modelación de la conversión (que ocurre cuando usuarios deciden gastar dinero -real o virtual- dentro del juego), el abandono del juego y el cese en el gasto. Estas cantidades se analizan en todo detalle a diferentes escalas (o niveles de agregación) en el juego. Los resultados a diferentes escalas son relevantes para diferentes propósitos y se complemetan y enriquecen entre sí. Las predicciones de comportamiento de los jugadores individuales para los fenómenos de interés se generan utilizando bosques de decisión, colectividades de modelos de supervivencia y aprendizaje profundo. Para predecir comportaminetos a nivel de grupo se emplean distintos modelos de series temporales. También se analiza el uso de estas predicciones en la creación de perfiles de jugadores, al igual que un sistema de recomendación de objetos.

Los resultados aquí descritos tienen una aplicación práctica directa, ya que para los estudios de videojuegos es de gran utilidad comprender el comportamiento y las motivaciones de sus jugadores. Esto les permite diseñar juegos más atractivos y optimizar la planificación de los juegos ya comercializados. Señala además el camino hacia una mayor personalización del contenido ofrecido, ya que distintos elementos del mismo juego se pueden adaptar para satisfacer los gustos individuales de los jugadores. Desde una perspectiva más fundamental y ambiciosa, este trabajo pretende ser una pequeña contribución al estudio ciéntifico y matemático del comportamiento y las sociedades humanas.

Las principales contribuciones originales de esta tesis se resumen a continuación. En la lista de publicaciones 6 se pueden consultar los detalles de dónde y cuando ha aparecido el contenido con anterioridad.

Contenido publicado previamente como primera autora:

1. En el capítulo 4 se describe un marco en el que modelos de espacio de estados se usan para predecir conversión, abandono y cese de gasto en el juego, y que permite cuantificar el efecto, tanto de elementos de la planificación del juego (eventos dentro del juego, promociones...), como de otros externos (festivos, lanzamiento de nuevos juegos de la competencia...) [73].

2. El capítulo 7 hace uso de predicciones, para cada jugador, en días, horas de juego, progresión en el juego y gasto total hasta el abandono para elaborar perfiles de jugadores y comprender mejor elementos de la dinámica del juego [72].

3. El capítulo 9 es un estudio, mediante el uso de modelos y herramientas de física estadística, del comportamiento colectivo respecto a procesos de toma de decisiones interdependientes en videojuegos con interacciones sociales [74].

Contenido publicado previamente como segunda o tercera autora:

1. En el capítulo 6 (sección 6.1) se predice el potencial de los jugadores para convertirse en usuarios de pago, así como los días, el tiempo de juego y el nivel en que la conversión tendrá lugar, utilizando modelos de supervivencia [133].

2. En los capítulos 5 (sección 5.3) y 6 (sección 6.4) se analiza el impacto de diferentes perfiles de abandono y cese de gasto en el desempeño de modelos de clasificación binaria y modelos de supervivencia para la predicción de abandono y cese de gasto [132].

3. El capítulo 6 (sección 6.5) evalúa el uso de perceptrones multicapa y de redes neuronales convolucionales para predecir el gasto total esperado de cada jugador [51].

Trabajo no publicado con anterioridad al que la autora contribuyó:

1. En el capítulo 8 se describe un sistema de recomendación de objetos para videojuegos con una dimensionalidad elevada (gran número de objetos entre los que elegir). Dicho sistema utiliza una combinación de métodos de agrupación, colectividades de árboles extremadamente randomizados y filtrado colaborativo. Tan sólo se describe la metodología, no se presentan resultados concretos, ya que no es posible publicar con los datos empleados en su desarrollo.

2. En el capítulo 9 sección 9.5 se presenta el estudio de comportamiento colectivo en procesos de toma de decisiones interdependientes para el caso en que las poblaciones son heterogéneas en sus preferencias. Este trabajó no se ha publicado con anterioridad pero sí fue presentado en dos congresos.

AGRADECIMIENTOS / ACKNOWLEDGEMENTS

This project has taken much longer than usual (and than I would recommend). As a result, I have lived in five cities and had as many different jobs in the meantime. I have always been extraordinarily lucky in coming across amazing people throughout my life who have been a constant source of support, growth and inspiration. To sum up: huge thanks to many different people are definitely due. I could fill many pages with them, but I will do my best to keep this short and on point.

Quienes más agradecimientos merecen aquí son mis directoras de tesis. Las dos son mucho más para mí que directoras de tesis. Elka, muchísimas gracias por no haber perdido la paciencia y haberme sabido dar siempre nuevas oportunidades. Si hubiese elegido a cualquier otra persona para guiarme en este camino, casi seguro que este día no habría llegado nunca. África, muchas gracias por empujarme a acabar este proyecto y por todo lo demás.

I am also particularly grateful to all my coauthors other than my advisors (Anna, Pei Pei, Shi Hui, Javier). Without them, this thesis would not exist, or at least not in its current form. Gracias especialmente a Anna por sus ánimos hasta el final. A Javier además por haber sido un apoyo imprescindible para arrancar este proyecto. Espero que la sorpresa de que realmente le ponga fin a esta tesis sea el mejor agradecimiento.

I could also not be more thankful to the non-academic bosses I have had since starting this thesis. Roland, deine Ermutigung und Unterstützung haben dieses Projekt wahrscheinlich am Leben erhalten. Vielen Dank von ganzem Herzen! Y de nuevo toca dar las gracias a África en este otro papel, por no haber dejado nunca de abrirme puertas y de cambiarme la vida.

I have also had the best colleagues ever, both in and out of academia. I will refrain from making a long list of names, but you know who you are. You have all shaped me into the scientist and professional I have become. For the countless interesting discussions (on work, on science, on life) and networking-team-building-drinks, for the many tips and advises shared, for all the skills learnt: many thanks to you all! Special thanks to Dexian and (again) Anna for so many shared experiences across three cities, for the joy that working with you guys brings me and for your continued support over the last few years.

Entrando ya en el terreno personal, si por algo soy afortunada (y mira que he tenido suerte yo en cosas) es por la familia que me ha tocado. Por ser amor y ser refugio: gracias, gracias, gracias. Especialmente a mis padres y a mi hermano, porque *el sol que brilló sobre mi infancia me privó del rencor para el resto de mis días*. Pero también por su inestimable colaboración a mi tía y a mis tíos (que han sido como segundos padres), a mis primas (que son más bien hermanas) y a mis abuelos (que aunque ya no estén estarían más orgullosos que nadie). A Lucas y Gabs, porque aunque quede cursi, hacen brillar el sol sobre mi adultez como pocas otras cosas. A Eva por haberse convertido en una más. Si con alguien tengo ganas de celebrar, esta tesis en particular y la vida en general, es con todos vosotros.

Como no estar también eternamante agradecida a mi familia elegida. Que por algo la habré escogido, digo yo. Gracias por la complicidad cotidiana, presente y pasada. Gracias por hacerme las penas más llevaderas, por las alegrías compartidas, por que nos quiten lo bailao y por lo que te rondaré morena. Por hacerme, en fin, la vida mucho más bonita y a mí mucho más yo.

Gracias a las que llevan aquí desde siempre, casi literalmente: Lu, Rox, Blanqui, Sus, Alex, Yai, Mara (por estricto orden de aparición en mi vida creo). Todas ellas señoras modernas a tope de power. Porque *hasta en el infierno están las amigas*, y porque que os voy a contar yo a vosotras de mí que no sepáis.

Gracias a las que se me cruzaron hace media vida (y aquí siguen). Aquel núcleo familiar tan particular y semifuncional de la calle León: María, Pame, Lu (otra vez), Natxo, Esther. A los que entraban y salían (y ya siempre van a estar): Sara, Gal, Matteo, Fouad, África (nuevamente). A mis sister-wives Vicente y Andreea, aunque entonces aún no sospechase que nos convertiríamos en comando. A Álvaro por haberme acompañado hasta llegar a esa etapa. A Alberto por haberme enseñado tanto.

A todos los demás que llegasteis o andabáis ya por allí alegrándome la vida: Ernesto, Sergi, Rubén, Temes, Iván, Ana Bea, Diego, Judith, Mario, Meggy, las Grecas, Jorge, Ana, Laura, Jesús, la benasquepípol. A David por haberse encontrado a Blanca, y por Dani, Eva y Pau.

Gracias como no a las que me salvaron la vida y hasta la felicidad en Frankfurt: África (otra vez), Javi, Just, Andreu y Ari, por haber sido siempre casa. Thanks, Gopika (and Christian!), for all the fun bits too of course, but very specially for having taken care of me at my lowest. A Walter por hacer que hasta de los últimos meses allí sacase nuevas incorporaciones para la familia (y por las velitas que tanto han ayudado a que termine esta tesis).

A Miguel, Badia y Pilar por haber sido un trozo tan grande y tan bonito de mi vida en esa época. A William y Andy por haber sido un soplo de aire fresco. Thanks Harpreet for the path we shared. An Michelle und Nico: vielen Dank für deine Begleitung und Unterstützung, damals und heute.

Gracias también a las últimas incorporaciones. Meeting Isis would have already made moving across the world worthwhile. I owe you, together with Yukako and Tea, some of my best memories in Tokyo. Thanks to you all (and Hiroshi and Thomas), I have people to come back to once we are allowed to visit again. And thank you Nick for Infinity! Gracias también Gerard por los ánimos en la recta final, y por la alegría de que siga entrando gente nueva y bonita en mi vida.

Gracias por último a los que perdí por el camino pero igual siguen muy presentes.

CONTENTS

LIST OF TABLES

NOTATION

FREQUENTLY USED ACRONYMS

AI	Artificial intelligence
ANN	Artificial neural network
AoI	*Age of Ishtaria*
AUC	Area under the curve
CDF	Cumulative distribution function
CF	Collaborative filtering
CNN	Convolutional neural network
DAU	Daily active users
DL	Deep learning
$DMLP$	Deep multilayer perceptron
DNN	Deep neural network
ERT	Extremely randomized trees
IBS	Integrated Brier score
IID	Identically independently distributed
IWA	Idiosyncratic willingness to adopt
IWP	Idiosyncratic willingness to pay
$LSTM$	Long short-term memory
LTV	Lifetime value
ML	Machine learning
$non-PU$	Non paying user
PDF	Probability distribution function
PU	paying user
$RFIM$	Random Field Ising Model
RNN	Recurrent neural network
RSF	Random Survival Forest

FREQUENTLY USED SYMBOLS

F	Free energy
f	Free energy density
H	Hamiltonian
h_i^a or H_i^a	Opinion field or IWA in choice a
J_a	Social coupling strength in choice a
k or K	Choice coupling
s_i, t_i	Spin/choice variables that can take values -1 or 1
w_i, u_i	Spin/choice variables that can take values 0 or 1
Z	Partition function
μ_a	Fraction of adopters in choice a

PHYSICAL CONSTANTS

K_B	Boltzmann's constant, $K_B = 1.380\,649\,\mathrm{J\,K^{-1}}$

1

INTRODUCTION

If you can't give me poetry, can't you give me poetical science?

— Ada Lovelace

This thesis deals with player activity in video games. The context is that of understanding human behaviour and social dynamics with a scientific approach, in what is sometimes referred to as *sociophysics*. Player behaviour is indeed human behaviour, and a particularly useful and interesting example to study, as will be discussed. Besides, the sort of analysis presented here have immediate practical applications: the statistical modelling results that will be described can be used to develop more engaging games, in the interest of both studios and players.

The content of this thesis can be broadly divided into two categories. Most of its chapters dwell on analysing and predicting player behaviour quantitatively. That is, within a *data driven* mindset, the results of applying different *statistical* and *machine learning (ML)* models to video game datasets are described, and their possible uses discussed in detail.

The last part of this thesis however, has a fundamentally different approach. Instead of data, simplified models of reality and the processes at play are the starting point. The tools of statistical mechanics are then put to use, with the goal of qualitatively (rather than quantitatively) understanding the mechanisms giving rise to interesting collective properties. It focuses on the role of player and choice interactions in decision making processes, and could therefore fit into what can be labelled *statistical physics of choice or opinion dynamics*.

Statistical mechanics is not only a theory of matter, but also a framework and a set of tools that can be used to study the aggregate characteristics of systems made up of many smaller constituents. It is particularly useful to study properties emerging through interaction. After it was developed and successfully used to understand the microscopic origins of the thermodynamic theory, and to bring further insights into the behaviour of gases, liquids and solids, it has found applications in diverse *foreign fields*, ranging from biophysics to neuroscience. This is now a field of research on its own,

the study of *complex systems*. It provides too, as will be discussed, a very useful framework in which to study human groups.

The quest to build a more scientific body of knowledge for the social sciences -albeit not new- is increasingly popular in the literature, and has gained a lot of momentum in the last couple of decades, with the surge of increasingly rich datasets and computational capacity. This has enabled the development, on the one hand, of ever more sophisticated and effective ML algorithms to extract the relevant information from the available data; and on the other, the possibility to run complex simulations in an attempt to understand the essential characteristics giving rise to the observed properties.

This chapter briefly introduces the general themes running through this work. In section 1.1 the case for a more *physics like* body of knowledge for the social sciences is made, and its history and state of the art succinctly presented. The specific dataset that will be used repeatedly throughout this thesis presented in section 1.2. This introductory chapter finishes with an outline of the thesis content, highlighting which parts are original contributions, in section 1.3.

1.1 STATISTICAL MODELLING OF HUMAN BEHAVIOUR

This work intends to make the case for, and take another step towards, what can be denominated as *physics of society* or *socioeconomic physics*, in particular in regards to *human behavioural science*.

From an epistemological point of view, this is mainly a matter of methodology and approach, being tantamount to using the *scientific method*. This entails building a formalised body of knowledge about reality, and most critically, its *empirical* nature, in that any theory developed should be able to make predictions that can to be confronted with the observed reality, in order to either verify or falsify its hypotheses.

The scientific method is the backbone of all scientific disciplines, including many approaches (particularly quantitative) within the social sciences. The use of the word *physics* does by no mean intend to discredit these, or claim any exclusive ownership over the methodology. It rather refers to the more or less direct application of models and tools from statistical mechanics and condensed matter physics to understand human systems.

The key lies in the paradigmatic revolution underwent by physics from a *deterministic* and *mechanistic* view to an *statistical mechanical* one during the first half of the 20^{th} century. Uncertainty has a central place in statistical

mechanical theories, and the tools and methods developed by physicists to study matter within this approach are specifically designed to deal with systems about which there is incomplete knowledge.

After succeeding in providing first principle explanations of the laws of thermodynamics for idealised systems, the statistical mechanical framework was then used to further delve into the nature of matter. Effective models for different phenomena of interest were built, in an attempt to understand non-ideal systems and explain their properties. The mathematical tools available were refined, polished and expanded, as the theoretical understanding underpinning them grew. The study of phase transitions and critical phenomena became a subject of its own. Applications and interrelations with other theoretical disciplines were found, and are being explored to this day, ranging from information theory, game theory, and nonlinear dynamics and chaos, to graph theory and geometry and topology.

In parallel, the development and widespread use of computers, the continued increase in cheap computing power, and the consequent improvement of algorithms (in particular of Monte Carlo methods), made available a different set of tools to study the same type of problems: computer simulations. Hypotheses in the model's definitions could be relaxed, and their properties still studied methodically and precisely, even for systems that had previously been intractable.

It was soon obvious that the tools and methodologies developed could be used to study problems not pertaining matter. By the beginning of this century, the study of complex systems was a well established discipline, outreaching non physical fields such as graph theory, behavioural and evolutionary biology, neuroscience, genetics, and, in precisely what concerns this work, social sciences and human behaviour.

In short, statistical physics therefore provides the tools that allow for a mathematical characterisation and exploration of human systems, with properties that can then be rigorously deducted, and the level of abstraction necessary to build a consistent body of knowledge around these systems. It provides a way of building *theoretical models* of human behaviour, from something resembling *first principles* (or basic assumptions on how the system under study operates). As with all physical models of reality, these will involve hypotheses, idealisations and simplifications depending on the intended use of the model. As with all physical theories of reality, their validity will lie ultimately in how accurately they are able to describe and predict observations of the phenomenon modelled.

The analysis of the available observations regarding human and social behaviour is therefore the other critical part to this approach. Carrying out actual experiments to put the hypotheses of social dynamics models to test is normally very difficult or outright impossible, both due to practical viability issues and ethical concerns. The study of collected data consequently plays the role of the *experimental* counterpart of the statistical mechanical theoretical models in sociophysics. Note that building models using data has been a traditional approach in the quantitative social sciences.

This area of research has also undergone an explosive growth, both in the literature and its practical applications. The *digital transformation* human life has undergone is responsible for the production of enormous amounts of data pertaining virtually all aspects of the human experience. The *computer revolution* already described, not only provides the means to store, process and analyse this data, it also promotes the appearance of new and increasingly rich datasets to study, as smaller and more powerful devices acquire a role in an increasing number of human activities. From our training progress, to with whom, how often and in what language and register we communicate, to how we choose to present ourselves (or not) in personal and professional networks, to how we go about our (online) dating lives, our shopping lists, or our mobility patterns, digital datasets contain a wealth of information about human behaviour that can be unlocked with the correct tools.

Statistical and machine learning, and their brighter cousin *artificial intelligence (AI)*, constitute arguably the field that has experienced a more drastic growth during the 21^{st} century, both in terms of research produced, and its impact in everyday life applications. They provide ways to uncover patterns, trends and correlations in the datasets, beyond the descriptive analysis and hypothesis testing afforded by classical statistics. As in the case of the statistical mechanical tools, ML has also been used with a lot of success in fields unrelated to human behaviour and organisation. From image recognition and neuroscience, to weather prediction, and even theoretical physics, ML is being of extraordinary help in solving problems formerly inaccessible through classical numerical methods, even with the most powerful computers available.

Data science can be defined as the quest to extract knowledge from the data, be that through the use of frequentist or bayesian statistics, or of statistical machine learning algorithms. It can be used in a utilitarian and functional manner, to describe, predict or classify for particular datasets in practical setups. It can also contribute to conceptually better understand

basic general properties of the systems under study, which should match the basic assumptions for the theoretical models described above. In this sense, it can be rightfully considered the *empirical branch of sociophysics* when applied to human related datasets.

The borders between the theoretical and empirical arms of any discipline are typically rather diffuse, as they share the end goal of describing the observed reality as accurately as possible, and should feedback into each other. Good empirical science makes use of all the theoretical knowledge available. Good theoretical science looks for models having the latest experimental results on mind. This is also the case with the formal and data driven approaches to modelling human behaviour that have been described. Data models can not only *get inspired* by statistical mechanical methods -such as the use of mean field theory, for example-, some allow for or require assumptions about the underlying dynamics of the system under study. Theoretical models should try to reproduce meaningful correlations discovered through data science, and help shed light on the causation lying at their origin. In other words, data science uncovers *what* is happening, while the statistical mechanical formal approach is more interested in the *why*.

The idea is to find interesting properties in the data, to then build models of reality that will recreate some of the characteristics observed, to then check their predictions against real data again. Then go back from data to models, modifying the initial hypotheses to make their predictions increasingly close to reality. The mathematical characterisation of the system gives the formalised body of knowledge, while the data driven approach the empirical nature, that are the pillars of the scientific method.

In fact, the close links between machine learning, information theory and statistical mechanics have been clear for decades. Statistical physics can be used to theoretically analyse the performance of machine learning algorithms. As has been described above, ML is already being used to solve problems in theoretical physics, including statistical physics, such as the automatic detection of phases of matter. A paradigmatic example can be found in *neural networks (NNs)*. Statistical physics provides very apt tools for the formal study of *artificial neural networks (ANNs)*. These can in turn be used to solve ML problems, and are able to *learn* very non linear relations in datasets. Particularly so when considering many layers, referred to as *deep learning (DL)*, which has been critical to the success of ML methods in areas such as natural language processing or computer vision (among many others). Not only can ANNs be studied using statistical mechanical

tools, Boltzmann machines (a type of stochastic recurrent neural network) for example, are a direct application to ML of a Sherrington-Kirkpatrick model (a stochastic Ising model first formulated to describe spin glasses).

Without leaving the realm of statistical physics, the combined data science-statistical mechanical approach proposed can be illustrated using black body as an analogy to any particular human behavioural process. Thermal radiation was first described by Wien's distribution law for short wavelengths (large frequencies) and by Rayleigh-Jeans law for long wavelengths (small frequencies). Although they both involve certain classical physics and thermodynamic principles in their deduction, they can be rightfully said to be empirical laws, in that they were found in an attempt to describe experimental results. They both failed to be universal enough as to work for all the spectrum, but provided accurate predictions for a large range of frequencies. These would be equivalent to some data science based results, which could be using some assumption on the governing dynamics of the system, and which would be producing relatively accurate predictions at least under certain conditions. The black body problem was finally solved by Planck's law in 1900, which was derived heuristically by assuming the oscillations were quantified, in what can be considered the kick-off of both statistical mechanics and quantum physics. This would be equivalent to finding which key mechanism(s) in the theoretical modelling is giving rise to the data driven results obtained.

Of course, the talk here about first principles in relation to human behaviour, does not intend to convey that someone's actions can be predicted exactly, or that people inexorably follow some rules, as is the case of thermal radiation in the example given above. Humans are not atoms or photons. What is argued here, is that it is sometimes useful to approach the study of their behaviour as if, *in some sense*, they behaved as such. Specially in what concerns aggregated or averaged properties, many of the complicated details are irrelevant to some processes. As in the ideal gas, where the shape of its molecules might be irrelevant, and the velocity at which each one is moving impossible to determine. Their temperature will still be related to their average speed, and the temperature to the pressure, volume and the number of molecules, following the same relation for all gases.

The black body example intends only to illustrate the methodology followed. This includes idealised models of a real phenomenon (the theoretical black body is a a perfect absorber and emitter and is in perfect thermodynamic equilibrium). Experimental efforts in order to observe data in real systems close enough to the theoretical assumptions (experiments measur-

ing thermal radiation), and in finding equations that could reproduce the results (leading to Wien's and Rayleigh-Jeans laws before the phenomenon was correctly understood), are also key. Finally, theoretical efforts in arriving at this empirical equations from the physical knowledge readily available, with some additional hypotheses, closes the loop (in this example, Planck's law).

The study of problems concerning human behaviour using the scientific method will involve far cruder idealisations (such as humans as rational agents making choices to maximise some abstraction). The experimental efforts in this case can correspond to actual experiments (as carried out with small groups in certain behavioural science labs), or more frequently the careful curation and analysis of data related to the problem under study (for example, measures of the actual choices humans are making). Statistical or machine learning models can in some sense play the role of the empirical laws. Ideally, there will be constant feedback loops between the data-driven and conceptual approaches which would refine the understanding of the phenomenon under study, even if it remains imperfect.

All this will be applied in this work in the context of video games, with the data-driven results for a particular title (*Age of Ishtaria*, described in section 1.2) used to illustrate the quantitative approach. Video games present a particularly good example to study for three reasons. First, most video games are nowadays played online. Every action every player makes is recorded, generating extremely rich high quality datasets that constitute an ideal playground to analyse human behaviour from a data driven point of view. Second, games can in many aspects mimic life, and they definitely engage different human features concerning skills, psyche, thinking processes and decision making. In this sense, they can also be used to shed light on general human behavioural problems, and to study human traits of interest such as strategic thinking, association, competition or confrontation. Last but not least, understanding how players behave and why, can be used to develop better games, which is in the interest of both video game studios and users.

Section 1.1.1 gives a brief historical perspective of scientific approaches to the understanding of human affairs. Section 1.1.2 gives an overview of some theoretical approaches to problems of the social sciences, while 1.1.3 gives a similar survey for data driven studies. Note that in some (the best) cases works comprise (at least partially) both theoretical and data oriented elements, and thus the assigning to one or the other sections is sometimes somewhat artificial. In both cases they intend to be illustrative of interesting

ongoing research, but they are by no means comprehensive or thorough accounts.

1.1.1 *Brief historical perspective*

The enterprise of building a scientific body of knowledge concerned with human matters is everything but new. The purpose of this section is to illustrate (rather than to give a detailed account) this long lived ongoing pursuit through some interesting examples, as well as to highlight the interrelation between the development of social and natural sciences that existed, and is rarely acknowledged. More comprehensive pictures can be found in [12, 13, 112], on which what follows is largely based.

Note that, throughout this thesis, the word *science* is used as synonym of *modern western science*, as it emerged from what is sometimes referred to as the *scientific revolution* during the European enlightenment. The history narrated here is thus focused in what happened in the western world. This does not necessarily mean that a shared approach towards knowledge about human affairs and the natural world did not exist too in other cultures, and of course a lot could be learned from analysing these too. These have regrettably disappeared, or are understudied and accounts on them not easily accessible.

As when discussing natural sciences, much can be said about a rigorous approach to the study of human society by the classical Greek. As the knowledge they gathered was then lost and obscured for many centuries to come, the starting point of this narrative will be the 17^{th} century and the beginning of the *Age of Enlightenment*.

Thomas Hobbes (1588-1679) can be credited for the *conceptual origin* of a *physics of society*. He was the first to talk about something like natural first principles for the individuals composing society, and attempted to deduce the best form of government using these in *De cive* (*On the citizen*, 1642) and *Leviathan* (1651). It comes as no surprise that Hobbes had close links to Francis Bacon (he served as his secretary), and to the circle of French mechanistic philosophers -included Marin Mersenne (1558-1655) and Pierre Gassendi (1592-1655), colleagues of Descartes-, or that he travelled to Florence to meet Galileo.

The *data driven approach origin* was contemporary to the conceptual one and can be attributed to one of Hobbes disciples: William Petty (1623-1687). This founding member of the Royal Society was probably the first to suggest that the study of the fundamental laws (in a physical Newtonian sense)

that govern human systems (and which he named *political arithmetic*) had necessarily to entail quantitative measures. This led to the first collections of demographic data (births and deaths). By the end of the 18^{th} century, it had become very popular to look for trends in these type of datasets, activity that by then already received the name of *statistics*. This new field was however not considered to have any connections to mathematics.

The first mathematical characterisations of human problems also had to wait until the 18^{th} century, that saw the *birth of the theoretical approach* described in this work. Under D'Alambert's influence, the French mathematician Marie Jean Antoine Nicolas de Caritat de Condorcet (1743-1794) started to apply probability theory to voting. He arrived at interesting results, still relevant today. Condorcet's jury theorem states that the probability that the majority votes the correct decision will improve with the voting group's size, whenever each of its members are more likely than not to make the correct choice. In Condorcet's paradox, he showed that majority preference becomes intransitive with three or more options. Both results were collected in his *Essai sur l´application de l´analyse a la probabilité des décisions rendues à la pluralité des voix* (*Essay on the Application of Analysis to the Probability of Majority Decisions*, 1785). Other known thinkers of the 18*th* century which showed interest in a scientific approach to the study of human affairs were Baron de Montesquieu, David Hume, François Quesnay and Adam Smith.

The story became particularly exciting in the 19^{th} century, when data and maths were put together to analyse human systems, yielding what can be described as the *first physical (or physics like) theories* of human behaviour. The French mathematician and astronomer Pierre-Simon Laplace (1749-1827) and his pupil Siméon Denis Poisson (1781-1840), had used the Gaussian curve to fit astronomical measurement errors, in connection to probability theory. It was then found that it could fit social and demographic data too, both by Laplace himself, and by the also French mathematician Joseph Fourier (1768-1830), who was at the time director of the *Bureau de Statistique* of the *Département de la Seine*. The Belgian astronomer Adolphe Quetelet (1796-1874) was then to develop his *social mechanics* framework, after visiting the Royal Observatory, and becoming greatly impressed by Laplace's work. It consisted of an statistics based approach to analyse social processes that had an impact, among others, in Jeremy Bentham, John Stuart Mill or Karl Marx.

Auguste Comte (1798-1857), considered by many the father of *sociology*, was the first to coin the term *social physics*. He begun by applying his

epistemological perspective of positivism to mathematics and the natural sciences (physics, chemistry and biology). He then moved to the social sciences, and used the social physics label (which was later also used by Quetelet) for what was after to become known as sociology. His follower Henry Thomas Buckle (1821-1862), who was a great admirer of Quetelet's work, was the first to make the case for a *science of history* in *History of Civilisation in England*, where a large number of regularities in demographic data are compiled.

An unexpected twist in the plot came when the influence between the study of humans and the physical world was reversed. The work on demographic data by Laplace, Quetelet, Buckle an others was to inspire physicists such as James Clerk Maxwell (1831−1879) and Ludwig Boltzmann (1844−1906), two of the fathers of statistical physics. They were studying the new field of thermodynamics, which appeared after the industrial revolution, with the goal of increasing engine efficiency. They both explicitly mentioned the analogy between molecules in a gas and groups of humans. Laplace eloquently acknowledged this when he wrote "*those uniformities which we observe in our experiments with quantities of matter containing millions of millions of molecules are uniformities of the same kind as those explained by Laplace and wondered at by Buckle arising from the slumping together of multitudes of causes each of which is by no means uniform with the others*". He was the first to use the same Gaussian curve for the velocities of the molecules in a gas, developing the kinetic theory of gases, which began the statistical mechanical revolution in physics.

During most of the 20^{th} century, the social and the natural sciences developed independently. Physics was transformed by statistical mechanics, quantum mechanics, and the theories of special and general relativity. This brought about a unified quantum theory for all forces except gravity, an explanation of chemistry out of physical first principles, and an incredibly detailed picture of the constituents of matter down to subatomic (and subnuclear) level.

There were meanwhile also interesting things going on in the social sciences. The sociologist Vilfredo Pareto (1848–1923) introduced power laws to explain wealth distribution in 1897. George Kingsley Zipf (1902–1950) was a Northamerican linguist and philologist who studied statistical regularities in different languages. He later went on to collect data of diverse origins (demographic datasets, travel statistics, marriage date, war casualties...), and to show how power laws could also be used to explain them. By the middle of the century, there was also already a well formulated neoclassical

microeconomic theory in terms of rational agents maximising their expected utility.

In the last decades of the 20^{th} century, as has already been discussed, statistical mechanics started being used more and more frequently to study problems in foreign fields, giving rise to the study of complex systems as a discipline. Basic learning neural networks and other machine learning algorithms had been around for the whole of the second half of the century, with the psychologist Frank Rosenblatt's (1928–1971) introduction of the perceptron in 1957. The 1980s saw the birth of what can be called *statistical physics of machine learning*, with the pioneering work of John Hopfield (born 1933) -who introduced the Hopfield network-, Leslie Gabriel Valiant (born 1949) -and his theory of the learnable-, and Elizabeth Jane Gardner (1957-1988) -shifting the focus to the dynamics of the connections rather than the units themselves. Simultaneously, quantitative social science researchers were finding analogies between some of their problem formulations and statistical mechanics. Meanwhile some physicists such as Serge Galam were beginning to use statistical mechanical frameworks to study some social systems.

The 21^{st} century has seen an explosion of both machine learning applications, and of physics like research about social systems, as access to computational capacity and interesting datasets continues to grow. By 2010, with the arrival of deep learning, the huge potential of ANNs was out of question, and began to be routinely exploited. Sections 1.1.2 and 1.1.3 present a collection of current lines of research (more theoretical and more data driven respectively) that can fall under the umbrella term of sociophysics, to illustrate the state of the art.

1.1.2 *Theoretical approaches: related works*

This section gives an overview of some related active areas of research analysing human affairs from a complex systems point of view, in what can be referred to as statistical physics of human behaviour. It deals with works in which the focus is in building theoretical models of the system under study, rather than in their quantitative predictive modelling. For a similar overview of data driven approaches see section 1.1.3. The distinction is of course sometimes artificial, as many could fit both categories. Discrete choice theory will not be considered in this section, as it will be discussed in much more detail in chapter 2 section 2.10. The collection of references is

by no means extensive, and it intends only to illustrate interesting research going on.

General overviews of statistical physics like approaches to the modelling of human behaviour are given in [14, 266]. Surveys on different social dynamics models can be found in [46] and [188]. A review of the application of concepts and tools from statistical physics to social and political behaviour can be found in [112], and an exploration of various agent simulation models of political landscapes and election results is presented in [175]. The use of evolutionary game theory to model human behaviour is described in detail in [118]. In [13], Ball gives a detailed non technical account of what was the state of the art in 2004.

Within the social sciences, economics is the discipline where formal and quantitative approaches encountered less suspicion, and more rapidly became a respected mainstream branch of research. From Adam Smith's market theory in the 18^{th} century to our days, classical economic theory has remained captive of the idea of a market in equilibrium and Gaussian statistics. Interestingly, the known discrepancies with observed data (in particular in regards to the *fat tails* of many relevant fluctuation distributions), have meant no obstacle for the widespread *belief* in its postulates, or their role in supporting certain policies in the real world. The introduction of statistical mechanical frameworks has enriched classical microeconomic theory, particularly through the introduction of interactions. Although still far from providing a complete a well rounded understanding of economic processes, it has definitely shed light on the origin of some of the discrepancies between real data and predictions of the conventional approaches. Surveys in *econophysics*, as this approach is sometimes referred to, can be found in [41, 94, 185, 186, 262, 264].

Opinion dynamics is the study of how opinions spread, and is intimately related to discrete choice theory. The object under study is the same, but the spotlight here is on the dynamic evolution of the system (rather than its states of statistical equilibrium). A comprehensive review of basic opinion dynamic models can be found in [46], while [260] is a relatively recent survey. The simplest possible model is probably the *voter model*, which is equivalent to zero temperature Glauber dynamics in one dimensional lattices, or to random walkers that coalesce upon encounter.The *majority rule model* is another popular simple approach introduced by Galam. Together with threshold dynamics, it has been used to explain the *Trump phenomenon* in [113]. Galam models of opinion dynamics [111] are studied in the case of asymmetric contrarians (agents that tend to contradict rather than follow

the prevalent choice in the discussion group) in [114]. *Social impact theory* refers to another collection of opinion dynamic models [146].

Continuous bounded confidence models [182], introduced simultaneously by Deffuant [70] and by Krause and Hegselman [140], are used in [289] to understand the emergence and development of normalised (i.e., adopted by a relatively large part of the population) extremist views. The use of CODA (continuous opinions and observed discrete actions) models to understand the emergence of extremist individuals in networks of different topologies is addressed in [189], finding the presence of these will be pretty ubiquitous, although it can be controlled in certain types of networks, in particular by allowing individuals to change their position in the network.

Echo chambers, fake news spread, and polarisation, are analysed using a mean field approach in [190], and in complex networks in [18] (which reproduces qualitatively empirical observations of debates observed in Twitter). A model of information diffusion in social media taking with a competing large number of items of varying quality is presented in [214]. In [45] a continuous-time Markov process is used to model collective decision-making, when individuals can change their opinions to increase their own fitness value, but also due to social interactions. This is (as will be clear in chapter 2 section 2.10), a very close framework to the one that will be used in chapter 9 to study interdependent choices in video games, within the context of discrete choice theory.

Axelrod's model vector version of opinion dynamics type models that can be used to explain dissemination, acquisition and disappearance of cultural traits in different communities [7, 47]. Other models of *social diffusion* are based on epidemic dynamics. Examples range from the turn of the century works by Campbell and Ormerod with regards to the spread of crime [43] or the prevalence of marriage [218], to the more recent [115] which models the emergence of radicalisation, highlighting the role of social integration in preventing it.

Cooperation and group growth have also been subject to a lot of interest. An already *classical* paper is [9], in which cooperation is explained in terms of evolutionary game theory. The interaction of cooperators and free-riders, and their impact in the growth of the communities they belong to, was studied in [184]. In [187], social systems are a living evolutionary ensemble, with individuals that have different strategies and can choose to cooperate with others and form groups based on common interests. The mathematical framework employed to study such systems borrows tools from game and

kinetic theory. Urban gentrification is studied using the Schelling model in [219, 220].

A modelling framework in which the power laws (Zipf's, Heaps' and Taylor's laws), ubiquitous in many human related datasets, emerge naturally, is described in [277]. Power accretion in social systems was studied in [240] using agent simulations, showing that taxation can be used to prevent the naturally appearing growing inequality, and that the roughness and Shannon entropy of the power distributions can be used as useful complements to more traditional measures of inequality such as the Gini index or the Lorenz curve.

Other interesting works regarding human activities are concerned with crowd and traffic dynamics [141, 142], in which a very similar picture as the van der Waals theory of solids appear, with different phases depending on the mobility or flow of the agents.

1.1.3 *Data driven approaches: related works*

Along the lines of section 1.1.2, with the focus now on data driven research, this section outlines some current hot topics under exploration. This section covers the use of statistical and machine learning models that systematically look for correlations in the data (excluding the methodologies that will be used in this thesis, including deep learning, as these will be described in more detail in chapter 2 and are therefore not covered here). It covers, too, the use of models that could have very well been included in the previous section (section 1.1.2). As was discussed there, the distinction can be rather artificial (particularly for well rounded research). The criterion used to include works here has been simply to have a strong focus in explaining readily available or collected data).

Again here, the works cited are chosen with the intention of giving an idea of interesting interfaces between the qualitative approaches described in 1.1.2 and real data, and is by no means complete. A general overview of statistical learning theory discussing generalisation in the context of algorithmic approaches to function estimation can be found in [282]. A detailed discussion on data preparation is [231].

An interesting *early* example is [11], and agent modelling approach to business growth, which is used to correctly fit the data of both company size and growth rate of about twenty million US firms in 1997. This scaling behaviour in the growth of companies had been previously observed in [263]. Also remarkable in its simplicity while accurately explaining ob-

served results is [26], in which a simple *Sznajd model* (in which opinion is modelled as a spin chain) is used to explain voting outcome regularities. Another interesting example from last century is the *landscape model* devised by Axelrod and Bennet to explain alliance formation as an energy minimisation problem [8, 10]. They used it to (almost) accurately *predict* company alliances in regards to Unix standardisation in the 1980s, and political alliances in Europe prior to the Second World War.

The analysis of social network data has been the protagonist of very diverse analysis. In [162] data from three billion Facebook users and their opinions on vaccines considered. Mass-action modelling was used to successfully reproduce the evolution of pro- and anti-vaccination clusters. In [5] twitter messages are used to evaluate the appeal of extremism in the US, finding that text-based psychological indicators support the existence of psychological differences between left- and right-wing activists (*moral foundation hypothesis*), and extremist users distinguishing themselves from the rest in four of the five big personality traits (openness, conscientiousness, extraversion, agreeableness and neuroticism).

In [122] the propagation of conspiracy theories is considered, using the 9/11 terrorist attack as an example. Their spread evolution is modelled using information theory and entropy, analysing online comments to related news or blog posts, and showing that the evolution of entropy measures too the degree of penetration of the conspiracy theory.

The understanding of urbanism has also received a lot of interest, with [205] being a very recent example, which proposes a maximum entropy, nonlinear, generative model of cities, and uses it to predict the evolution of French towns. Gentrification during the last decades is studied for New York, London and Tokyo in [281]

Remarkably, [280] presents a first principles approach to understand the birth and evolution of social networks, and is then used to account for the structure of mentions between Twitter users, co-authorship of the American Physical Society and mobile-phone-call network (see [212, 213] for introductory reviews on complex network theory and applications).

1.2 AGE OF ISHTARIA

Data analysed throughout this thesis comes from the game *Age of Ishtaria (AoI)*, which is a Japanese mobile role-playing card freemium game developed by Silicon Studio and available worldwide. This game's data is explored in the papers with original work contributing to this thesis [51,

72, 73, 132, 133, 224], as well as in previous work [28, 52, 131, 167, 224, 225, 239].

The game was launched on September 25, 2014. It is still currently available for download, but data only until May 9, 2017 was available, i.e., nearly the first two years of the game's life can be studied with the dataset available. Only data for Japanese players was considered.

As will be clear from chapter 3 onward, most of the work in this thesis bases in characterising players by when and how long they play, how they progress (level-up) throughout the game, and when and how much they spend, which can be further summarised into giving, for each player at any given moment in time, how many days have gone by since their first login into the game (lifetime days), how long they have played in total since (accumulated playtime), total in-game progression since (level) and total expenditure since. The possibility of missing or noisy values in these variables in basically negligible: even if there was some technical problem preventing the recording of some of the actions, any single action recorded is enough to know the player has connected. Players will typically complain if level-ups or purchases are not recorded (and thus not effective), so these are reduced to the minimum too. Playtime variable allows for more noise than the others, but is still kept to the minimum. Session length is computed ideally as the difference between login and logoff time. It happens often however, that users do not bother to log off. Therefore, if after any action there is a period of more than 5 minutes of inactivity, the session is considered to have finished after that action. Similarly, if there is no session active through a log in, but the player logs any action, then that action marks too the beginning of a new session.

In the period available for study there were a total of 2107166 players, with 33194 still considered active at the end of period (more details on what is meant bu active player can be found in chapter 3 section 3.2). Of those 33194 did at least one in-app purchase. While there are peaks of nearly fifty thousand *Daily active users (DAU)*, i.e., of different users playing on a given day, typical values are normally in the range of ten to twenty thousand. Periods with higher values are believed to have been do to aggressive new user acquisition campaigns. Although it is known that there were both marketing or new user acquisition campaigns (i.e., outside the game to get people to try it and hopefully continue playing) and promotion campaigns (discounts or the like offered to players inside the game to promote spending) often throughout the period considered, no additional information (dates, duration, details ...) is available. In chapter

4 time series modelling techniques are applied to try and discover when these campaigns were taking place.

Information about in-game events that took place during the period under study is available. For each event, the start and end date, together with the event type and a measure of the impact (in playtime and/or purchases) that particular event was expected to have are available. The latter is an integer between 0 and 4 (both included), with 0 representing no noticeable impact expected and 4 the highest possible effect. Though numerical, this characteristic is better understood as a categorical qualitative measure of the expected outcome. Every event falls into one of the following categories or event types: Adveniment, Battle Arena, Battle Event, Call to Arms, Duel Arena, Gacha, Gift Event, Giant Break, Item Collection, Mission Event, Mission Bingo, Poll Event, Raid Battle, Raid Boss and Raid Event. Gacha is a type of game monetization strategy very popular in Asia, in which the player pays a fixed amount of money for an item that is different every time an assigned by chance. Being in fact more of a gamble than a purchase, they are very strongly regulated. The equivalent in Europe and America are loot boxes, that started being introduced and becoming popular much later than in Asia, and where regulations are also starting to pick on.

Plots of the daily time series of non-PUs, PUs and churned players can be found in figure 4.1, and of the transitions between these groups in 4.2. See chapter 4 for more details.

All quantitative plots of chapters 3 to 7 use the AoI dataset.

1.3 CONTRIBUTION AND OUTLINE OF THIS THESIS

This thesis deals with in-game player behaviour, with the dataset for AoI described in 1.2 as *experimental observations*. Most of it is concerned with a practical understanding of what is going on from the studios point of view, i.e., in exploiting the data collected from players to make predictions on how relevant quantities in the game are going to evolve, or how individual players are going to behave.

Although other problems such as that of item recommendation (see chapter 8) will be tackled , most content revolves around two key features, particularly in free2play games: whether users are playing or not, and whether they are spending or not. These are investigated through what is referred to throughout the text as *churn* (associated to active users that quit the game), *player conversion* (meaning players that are not spending begin doing so), and *purchase churn* (when the opposite happens, and purchasing

players stop spending). These can be studied at individual or aggregated level, and determine the degree of success in freemium games.

The last chapter before the conclusions (chapter 9) describes a theoretical framework in which to study the qualitative collective outcome of, either players of two types making choices under social influence (which has different intensity within than without the group), or a single group of players simultaneously deciding on two related matters, also with certain pressure to conform to their peers. The focus is in investigating a formal setup that can be applied to many different problems concerning video games (particularly very social ones), and different examples will be given. It will be stressed however, that the case of two simultaneous interrelated choices can be used to qualitatively understand precisely the active/non-active and paying/non-paying nature of the user which is thoroughly investigated quantitatively for AoI throughout the first part of the thesis.

The thesis is organised as follows. All models and methodologies to be used will be briefly presented in chapter 2 for later reference. Chapter 3 introduces some basic definitions and methods to divide players into active or not, and active spenders or not, which will be used throughout the thesis. Chapter 4 then analyses conversion, churn and purchase churn at very aggregated levels using a time series approach. The interest then turns to individual player churn in chapter 5, and to individual conversion and purchase churn in chapter 6, which also introduces methods to predict the total expenditure in the game players will have. The use of the predictions from the previous chapters to group players in meaningful ways is presented in chapter 7, and an item recommendation system in chapter 8. The use of statistical physics to draw qualitative conclusions on collective outcomes of interdependent decision processes (of particular interest to study playing or not, and purchasing or not) is then discussed in chapter 9. The thesis closes with some general conclusions in chapter 10.

Chapters 4 to 9 contain original content and are designed to be somewhat self contained. They all begin with some introductory paragraphs describing the problem to be addressed, the organisation of the chapter, and how much of it and to what extent is an original contribution of this thesis. They all finish with a summary of the more relevant results presented and concluding remarks when pertinent.

2

MODELS AND TOOLS

Meine Methoden sind wirklich Methoden des Arbeitens und Denkens; deshalb haben sie sich überall anonym eingeschlichen.

— Emmy Noether

This section will provide a review of the different models and tools used in this work to understand and predict player behaviour. Only the fundamental aspects required for a correct understanding of the following chapters will be presented, together with references in which the interested reader can find more details.

2.1 STATE SPACE MODELS

State Space Models (SSM) refers to a broad category of time series models in which an stochastic dynamic process is characterised by two equations: the *state transition equation* that describes the evolution of the so called *latent state* (unobservable directly), and the *observation model*, which is also probabilistic in nature, and establishes the relation between the observations and the latent state [33, 39, 136, 255]. One of the earliest examples studied in depth, and still in wide use today in numerous fields is the *Kalman Filter (KF)* [164].

Every SSM is therefore determined by the equations $p(l_t|l_{t-1})$ (state transition) and $p(z_t|l_t)$ (observation model), where $l_t \in \mathbb{R}^L$ is the latent state at time t and $z_t \in \mathbb{R}$ the observed state at time t. In particular, any linear SSM can be expressed as:

$$l_t = T_t l_{t-1} + c_t + R_t \eta_t \tag{2.1}$$

$$z_t = D_t l_t + d_t + \epsilon_t \tag{2.2}$$

where T_t is the *transition matrix*, c_t the *latent state intercept*, R_t the *selection matrix*, D_t the *design matrix* and d_t the *observation intercept*. The terms η_t and ϵ_t represent random innovations that are typically considered to be normally distributed, i.e,

$$\eta_t = \mathcal{N}(0, \Sigma_t^s) \tag{2.3}$$
$$\epsilon_t = \mathcal{N}(0, \Sigma_t^o) \tag{2.4}$$

where Σ_t^s is the *state covariance matrix* and Σ_t^o is the *observation covariance matrix*.

When two latent states $l_{1,t}$ and $l_{2,t}$ satisfy:

$$z_t = D_1 l_{1,t} + D_2 l_{2,t} \tag{2.5}$$
$$l_{1,t} = T_1 l_{1,t-1} + \eta_{1,t} \tag{2.6}$$
$$l_{2,t} = T_2 l_{2,t-1} + \eta_{2,t} \tag{2.7}$$

their governing equations 2.5 can be rewritten as:

$$z_t = (D_1 D_2) \begin{pmatrix} l_{1,t} \\ l_{2,t} \end{pmatrix} + \epsilon_t \tag{2.8}$$
$$\begin{pmatrix} l_{1,t} \\ l_{2,t} \end{pmatrix} = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} \begin{pmatrix} z_{1,t-1} \\ z_{2,t-2} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \tag{2.9}$$

I.e., when the time series of observations can be related to a linear combination of two latent states with linear states transition equations, the resulting model is also a linear SSM. Hence, any two linear SSMs can be combined to form a new one. This will be relevant in our case, as in chapter 4, the performance of two classical time series models (that can be expressed as SSMs as will be seen in the next subsections) will be compared. In both cases, the stochastic time series modelling is combined with a linear regression to explanatory variables (which can also be expressed as an SSM as will be soon described).

A lot of different filters and smoothers can be formulated as linear SSMs [33, 39, 136, 255]. The next three subsections describe how this is the case for the three instances of big model families that will be used in this thesis (in chapter 4 to be specific): linear regression, autoregressive integrated moving average and structural time series models.

2.1.1 *Linear regression*

A regression to time varying exogenous explanatory variables $z_t = \sum_i \beta_i x_t^i$ is a simple way to try an model the deterministic behaviour of a time series

z_t, with x_t^i for $i = 1, \ldots, n$ the n explanatory variables (sometimes referred to as *covariates* or *regressors*), and β_i their corresponding parameters to be estimated. This simple model can also be expressed as an SSM by setting $T_t = d_t = R_t = 0$ and $c_t = \sum_i \beta_i x_t^i$ in equations 2.1. See chapter 3 in [159] for a detailed explanation and discussion of applications of linear regression in statistical learning.

To perform forecasts, besides the parameters estimated using the historic values available (trainig dataset), the projected values of the explanatory values x_t^i into the future are needed. Predictions can then be carried out by simply multiplying each covariate times its associated paramter and summing the result.

2.1.2 *Autoregressive integrated moving average (ARIMA)*

The best known expression for an *Autoregressive Moving Average (ARMA)* process is [33]:

$$\begin{aligned} z_t &= \alpha + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + && (2.10)\\ &\quad \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t && (2.11)\\ \epsilon_t &= \mathcal{N}(0, \sigma^2) && (2.12) \end{aligned}$$

where $\phi_1, \ldots \phi_p$ the autoregressive parameters, $\theta_1, \ldots \theta_q$ the moving average parameters and α the model's intercept. This is an stochastic time series model in which each observation has a weighted dependence on the previous p observations (and the process is said to have autoregressive (AR) polynomial of order p) and on the previous q noise realisations (moving average (MA) polynomial of order q). The parsimonious nature of these models arises because even with a relatively small number of parameters (low p and q), each time step can be made to depend on a large (virtually infinite) number of the previous values of the observation time series.

An ARMA model can be expressed in SSM formulation (equation 2.1) yielding [39]:

$$\begin{aligned} y_t &= (1, 0, \ldots, 0) l_t && (2.13)\\ l_t &= \begin{pmatrix} \phi_1 & 1 & 0 & \ldots & 0 \\ \phi_2 & 0 & 1 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \\ \phi_r & 0 & 0 & \ldots & 0 \end{pmatrix} l_{t-1} + \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_r \end{pmatrix} && (2.14) \end{aligned}$$

where $r = max(p, q+1)$, $\theta_i = 0$ for $q < i \leq r$ and $\phi_i = 0$ for $p < i \leq r$, and $l_t^T = (y_t, y_{t-1}, \dots y_{t-p})$.

ARMA models are easily extended into *Autoregressive Integrated Moving Average (ARIMA)* models by allowing the observation time series z_t to be differenced (i.e., subtracting to each observation the value of the previous time step). An ARIMA model of order (p, d, q) is thus one with AR polynomial of order p, MA polynomial of order q, and d differences taken in the original series, typically in order to make it stationary and/or reduce its variance. This concept can be extended to differences and polynomials of lag different than one, yielding *seasonal ARIMAs* or *SARIMA*.

Once the parameters of the ARIMA have been estimated using the training dataset, forecasting can be performed recursively for consecutive times into the future by applying equation 2.10, first to the last real observed values, then to the previously predicted ones.

The combination with a linear regression to covariates to account for some deterministic behaviour is sometimes referred to as SARIMAX modelling. If the series modelled is not transformed, the resulting model in the explanatory variables is *additive*, in that the parameter estimated for each covariate can be understood as the increase (or decrease if negative) in the modelled time series for each unit of increment in the covariate. If the series is log-transformed, the model is then multiplicative, and the parameters are better understood as elasticities, i.e., proportionality constants between relative increments in the explained and explanatory series. *Residuals* is the name given to the unexplained part of the time series after adjusting the model corresponding to the random noise in the equations, i.e., they should be normally distributed with mean zero.

2.1.3 *Unobserved components or structural time series*

Unobserved Components (UC) or *Structural Time Series* models are SSMs where the observation is explicitly expressed in terms of a trend, cycle and/or seasonal dependency (each of which can be stochastic or deterministic in nature) [138, 139]:

$$z_t = \mu_t + \gamma_t + c_t + \epsilon_t \tag{2.15}$$

where μ_t is the *trend* component, γ_t is the *seasonal* component, c_t the *cyclic* component and ϵ_t a random shock $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$.

The trend component is given by:

$$\mu_{t+1} = \mu_t + \nu_t + \eta_{t+1} \quad (2.16)$$
$$\nu_{t+1} = \nu_t + \zeta_{t+1} \quad (2.17)$$

where η_t and ζ_t represent white noise (normally distributed with zero mean) with variances parameters to be estimated. Depending on which elements of these equations are zero, the trend term is referred to as local linear (stochastic) trend (no null terms), smooth trend ($\eta_t = 0$), local (stochastic) level and deterministic trend ($\zeta_t = 0$), deterministic trend ($\eta_t = \zeta_t = 0$), local (stochastic) level ($\nu_t = \zeta_t = 0$) or constant term ($\nu_t = \zeta_t = \eta_t = 0$). Note the local level model is simply a random walk. It will appear repeatedly in chapter 4.

The seasonal part (or parts) can be expressed as:

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + w_t \quad (2.18)$$

where w_t is random noise with zero mean and variance estimated as an additional parameter. It captures a fixed seasonality of the time series (for example $s = 7$ for behaviour repeating itself each week for a daily time series).

The cyclic term captures repeated behaviour over longer unspecified periods of time:

$$c_{t+1} = c_t cos\lambda_c + c_t^* sin\lambda_c + u_t \quad (2.19)$$
$$c_{t+1}^* = -c_t sin\lambda_c + c_t^* cos\lambda_c + u_t^* \quad (2.20)$$

where u_t is also normally distributed with mean zero and estimated variance. The cyclic frequency λ is also estimated as a parameter.

As for the ARIMA models, once all parameters have been estimated, forecasts can be performed recursively into the future.

ARIMA and UC models are closely related and equivalencies can be found among them. UC models allow for more randomness (through the possibility of including more than one noise term), hence normally both parameter spaces are not identical even for *equivalent* models (in that for example some only versions with constraints between the noise terms of the UC side will be able to be expressed as an ARIMA). This is normally expressed using the term *reduced*, and the ARIMA of order (0, 1, 1) is, for example, the reduced model of a local level UC model.

2.2 STATISTICAL TESTS AND ESTIMATORS

This section compiles the different classical statistical tests and estimators used in chapter 4 of this thesis.

2.2.1 *Augmented Dickey-Fuller test*

The *augmented Dickey-Fuller (ADF) test* is a hypothesis testing used to evaluate stationarity of a time series after removing autocorrelation, and is described in [56]. The null hypothesis is here that the series is correctly characterised by a unit root, i.e., that $\gamma = 0$ for:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \ldots + \delta_{p-1} \Delta y_{t-p-1} + \epsilon_t \tag{2.21}$$

with α, β, γ and δ_i constants, and p the lag of the AR process. The alternative hypothesis states that $\gamma < 0$. The critical values are found using a Dickey-Fuller distribution, to which the relevant statistic, computed as the average γ divided by its standard deviation, can be compared.

2.2.2 *Ljung-Box test*

The *Ljung-Box test* is a hypothesis testing used to evaluate independence of a time series, and is described in [181]. The null hypothesis is here that any observed correlations are as result of sampling. If the null hypothesis holds, the statistic Q needs to be greater than the chi-squared distribution with h degrees of freedom, where h is the lag under test, Q is given by the expression

$$Q = n(n+2) \sum_{k=1}^{h} \frac{\rho_k^2}{n-k} \tag{2.22}$$

and n is the sample size.

2.2.3 *Jarque-Bera test*

The *Jarque-Bera test* is a hypothesis testing used to evaluate normality of a time series, and is described in [160]. The null hypothesis is here that the skewness and kurtosis of the sample matches, i.e. zero skewness and zero excess kurtosis. In this case, the statistic JB is close to zero and follows a

chi-squared distribution with two degrees of freedom asymptotically, with JB given by

$$JB = \frac{n}{6}(S^2 + \frac{1}{4}(K-3)^2) \tag{2.23}$$

n the sample size, and S and K its skewness and kurtosis, i.e. its third and fourth order central moments, respectively.

2.2.4 *Akaike information criterion*

The *Akaike information criterion (AIC)* is an estimator of in-sample prediction error and is described in [3]. It is given by

$$AIC = 2k - 2\ln(max\{L\}) \tag{2.24}$$

where k is the number of estimated parameters, L is the likelihood function of the model, and $max\{L\}$ denotes its maximum value. This estimator is linked to goodness of fit but also to the number of degrees of freedom, penalising the introduction of additional parameters. It gives no information in absolute terms, but can be very useful for model selection, favouring those model definitions with lower AIC values.

2.2.5 *Bayesian information criterion*

The *bayesian information criterion (BIC)* is described in [249], and is another estimator of goodness of fit, very closely related to the AIC described in section 2.2.4. AS the AIC, it also discourages overfitting by penalising additional parameters. It is given by the expression

$$BIC = k\ln(n) - 2\ln(max\{L\}) \tag{2.25}$$

where n is the sample size, k the number of estimated parameters, L is the likelihood function of the model, and $max\{L\}$ denotes its maximum value. As was the case for the AIC, lower values are linked to better model definitions in terms of information theory.

2.2.6 *Hannan-Quinn information criterion*

The *Hannan-Quinn information criterion (HQIC)* is an additional model selection index in the lines of the AIC (see section 2.2.4) and BIC (see section

2.2.5), and is described in [137]. Lower values are more desirable, and it is computes as

$$HQC = -2max\{\ln(L)\} + 2k\ln(\ln(n)) \tag{2.26}$$

where n is the sample size, k the number of estimated parameters, L is the likelihood function of the model, and $max\{\ln(L)\}$ denotes maximum value of the log-likelihood.

2.2.7 *Parameter significance and z-scores*

Standard scores or z-scores describe where a given score is in relation to the mean, measured in standard deviations, i.e. $z = \frac{x-\mu}{\sigma}$, where μ is the mean of the population and σ its standard deviation. Standardising the estimated parameters for a model can be an aid in the interpretation of the relative contribution of each of them. The significance of each $z - score$ can be evaluated through hypothesis testing, with the null hypothesis being that there is no correlation between the covariate and the variable being modelled. Z-test is the analogue of the Student's t-test when using z-scores, and the associated p-values give the probabilities of the observed data being compatible with the null hypothesis. This means the lower the p-value, the less likely it is that the modelled variable would be as observed, if it was uncorrelated to the regressor. Hence, it can be used to assess parameter significance. More details can be found in [172].

2.3 DECISION TREES AND FORESTS

Classification and regression trees (CART) were introduced in [36]. They are predictive modelling technique in which the labelled training dataset is recursively split in two at the tree's *nodes*. Each node uses the values of one of the features for the partitioning, and the idea is to automatically find patterns of differences in the modelled variable that are explained in terms of differences in the features. The final nodes are populated by individuals with the same value of the target variable (in classification, or very close in regression). This makes possible to give a probabilistic prediction outside of the training set, depending on the values of the features. See for example chapter 8 in [159] for a general discussion of tree-based methods in statistical learning.

Ensemble learning refers to the use of a collection of learning algorithms (as opposed to a single one) to tackle a classification or regression problem.

Typically different instances of the same model are used, involving different parameters or selections of the training dataset. They usually improve model performance by removing biases, and provide generally more robust solutions. Additionally, they are trivially parallelizable, so while they do require additional computation capacity, they can be efficiently implemented (or as efficiently as the original algorithm). An ensemble of decision trees is sometimes referred to as *forest*.

When they are used for binary classification, a good verification metric is given by the *area under the receiver operating characteristic curve* (AUC) which is briefly described in section 2.9.

2.3.1 *Random forest*

Random forests were first introduced by Breiman in [35] and are probably the better known and more widely used ensemble tree technique. Each tree is trained using a bootstrap sample of the total set. Selection of the split variable and point are done in the same step. The selection of splitting variable at each node is done at random (hence the name). The split point is then chosen as that maximising the Gini impurity measure [36] (or a similar splitting criteria) which can be expressed as

$$I_G(p) \sum_{i=1}^{J} (p_i \sum_{k \neq i} p_k) = 1 - \sum_{i=1}^{J} p_i^2 \tag{2.27}$$

for J the number of classes and p_i the fraction of entities belonging to class i. These models tend to be biased as they favour variables with many possible splitting points. They have been extensively used for virtually all types of classifications and regressions, with some uncommon examples ranging from remote sensing [23] to gene expression [75].

2.3.2 *Conditional inference forests*

Conditional inference trees are described in [147], which describes how the use of conditional inference procedures in the recursive partitioning can solve the selection bias (towards covariates with many split points) problem, without negatively impacting overall model performance. The significance of the association between target variable and features is assessed using a chi-squared distribution, and a criterion is given to stop the process also based in sound hypothesis testing.

2.3.3 *Extremely randomised trees*

A *extremely randomised tree (ERT)* is another variation of a decision tree introduced in [117], featuring strong randomisation in the choice of both the covariate to use for the partitioning, and the split point, at each node. With appropriate parameter tuning, ensembles of ERTs do not degrade performance, while being significantly more efficient computationally. This makes them particularly useful for very large datasets.

2.3.4 *Gradient boosting*

Gradient boosting, of which *XGboost* [55] is a particular implementation, is a technique that can be used with decision trees. It was first introduced by Friedman [101, 102] and in Mason et al. [191, 192], and combines *boosting* (a bias and variance reduction meta-algorithm) and *gradient descent* (method to iteratively compute the local minimum of a function). The idea is to find a function of the covariates $F(x)$ as close to the target variable y as possible, i.e., of finding the minimum of a loss function $L(y, F(x))$ using gradient descent. Decision trees (or other weak learners) are used at each iteration to fit the pseudo-residuals (partial derivatives of the loss function with respect to the covariate function).

2.4 SURVIVAL ANALYSIS

Survival Analysis [58] is a general framework to study time-to-event regression problems. These methodologies were originally devised within the medical and biological fields, where the event of interest was death or organ failure [150], hence the name. These kind of problems are characteristically *censored*, i.e., data is incomplete or partially labelled. Some survival models are also able to handle *competing risks* [230], i.e., situations where there are other events which could impede the observation, or alter the probability of the event of interest under study.

Survival models yield a survival probability curve for each individual, i.e., the probability at each time point (past and future), of the individual still being alive (probability of the event of interest not having taken place). From these, a single time prediction can be computed if it is needed. Typically, the prediction of survival time is given by the median of the survival probability curve, i.e., the event of interest or death is considered to happen when survival probability drops below 0,5.

The *hazard function* $h(t)$ is defined as the ratio between the probability density function $P(t)$ to the survival function $S(t)$

$$h(t) = \frac{P(t)}{S(t)} \tag{2.28}$$

with different survival models accounting for different ways of defining and/or computing the survival function $S(t)$.

Survival model validation is often performed through error curves that highlight the dependence of the error with the value itself of the variable, and taking into account the probabilistic nature of the predictions. The *integrated Brier score* (IBS), briefly described in section 2.9, summarises the information of the whole curve in a single score ans is thus a convenient metric.

2.4.1 *Cox regression*

The *Cox proportional hazards model* or *Cox regression* [63, 64] is a semi parametric survival approach that assumes the relationship between covariates and hazard to be multiplicative

$$h(t|X_i) = h_0(t)e^{\beta X_i(t)} \tag{2.29}$$

with the survival function then expressed as:

$$S(t|X_i) = e^{-h_0(t)e^{\beta X_i(t)}} \tag{2.30}$$

where X_i is the matrix with all the covariates corresponding to individual i.

The likelihood function can be expressed in terms of the conditional hazards in equation 2.29, and its maximum computed numerically using the Newton-Raphson algorithm. The hazard function is not assumed to follow any particular distribution, but there is still a fixed relationship between the target variable and the covariates. Another inconvenient is that this method does not scale well to be very large datasets, although this is at least partially fixed by the *regularised Cox regression* [202]. Cox regression is typically used as a baseline for any survival methodology proposed.

2.4.2 *Random survival forest*

Random survival forests are the survival extension of the random forest algorithm defined in section 2.3.1, and is described in detail in [157]. It is a fully non-parametric approach that uses tree-based Nelson-Aalen estimators:

$$H_n(t, X_i) = \int_0^t \frac{T_n(dt, X_i)}{Q_n(t, X_i)} \tag{2.31}$$

where X_i is the matrix with all the covariates corresponding to individual i. T_n gives the number of uncensored events, and Q_n the total number of individuals at risk. The survival function can then be expressed as

$$S(t|X_i) = e^{-\frac{1}{N}\sum_{n=1}^{N} H_n(t,X_i)} \tag{2.32}$$

where N is the number of ensemble members.

Random survival forests allow for the introduction of competing risks [156], i.e., of events (different form the event of interest) that can affect the probability of observing the outcome. For example, when survival analysis is used to predict time to PU conversion (chapter 6 section 6.1), churn can be added as a competing factor (to lack of interest in purchasing). This amounts to understanding a non occurring phenomenon (a player not becoming PU) as being due to either not having interest or means to purchase or to having already quitted the game. Each node then becomes event specific, and for each event:

$$H_n(t, X_i) = \int_0^t \frac{T_{nj}(dt, X_i)}{Q_n(t, X_i)} = \sum_{k=1}^{m(t)} \frac{d_{nj}(t_k, X_i)}{Q_n(t_k, X_i)} \tag{2.33}$$

where $m(t) = maxk : t_k \leq t$ and $d_{nj}(t_k)0\sum_{i=1}^{M} I(T_i = t_k, \delta_i 0j)$ is the number of type-j events at time t_k for all individuals i, with I being the corresponding event indicator, and $d_n = \sum_j \delta_{jn}(t_k)$ the total number of events taking place at t_k.

2.4.3 *Conditional inference survival ensembles*

Conditional inference survival ensembles are the survival extension of conditional inference forests introduced in section 2.3.2. They have many similarities to the random survival forest described in 2.4.2, but use conditional inference trees [147] as base learners. The splitting is performed in two steps. First the variable more correlated with the output is chosen as split variable, then the optimal splitting point is computed based on two-sample linear statistics. They use weighted Kaplan-Meier estimates [149], and have survival function of the form

$$S(t|X_i) = \prod \left(1 - \frac{\sum_{n=1}^{N} T_n(dt, X_i)}{\sum_{n=1}^{N} Q_n(t, X_i)}\right) \tag{2.34}$$

where X_i is the matrix with all the covariates corresponding to individual i, T_n the number of uncensored events, and Q_n the total number of individuals at risk.

2.5 LIFETIME VALUE PROBABILISTIC MODELS

Lifetime customer value (LTV) (also referred to sometimes as *customer lifetime value, CLV*) can be defined as the estimation of the total amount of revenue that will be generated by a customer or user. It is of particular interest in non-contractual, continuous (in that purchases can be made at any time) business relations, where there is typically a very large variability between the income generated by different users. This is exactly the setting of free-to-play games, where players do not need to pay to play, but are free to purchase for in-game items or privileges (such as playing ad free), generating in many cases most of the game's revenue.

Traditional models aimed at giving LTV predictions basically assume that each customer (player in this case) will continue with identical purchasing patterns as have been displayed in the past (until they churn), in what are sometimes called "buy till you die" or "buy till you defect" (BTYD) models [247]. The more widespread probabilistic models used to this effect fall into the *RFM* category [89], acronym for recency, frequency and monetary value, as predictions for each individual are based only on when the last time they purchased was, how often they have purchased in the past, and how large their purchases have been. They operate under the general assumption that an individual is more likely to purchase again, the more recently, frequently and with larger value they have purchased in the past.

2.5.1 *Pareto/NBD*

The most popular of these models, often used as benchmark for any LTV estimation method proposed, is the so called *Pareto/NBD* model. The name refers to the two parametric distributions combined in generating the predictions. The Pareto distribution models the drop-out or churn process, and is used to classify customers into those that are still active and those that are not. NBD stands for Negative Binomial Distribution, which is how the purchasing frequency is parameterised [88, 89]. These models can also

include an additional submodel (parametric distribution) to predict the amount that will be spent per transaction, with a gamma-gamma model being a typical choice [89]. A simpler approach such as simply taking the average past value can also be used. Other possibilities that have been explored in the literature include the use of cohorts [174] or of logistic regressions [193].

Using a Pareto distribution for the drop-out rate is equivalent to considering that player lifetimes are exponentially distributes, with each user having their own churn or dropout rate. These dropout rates should vary independently across the total player population following a Gamma distribution. The continuous mixture of the exponential distributions of all players generates the Pareto one.

The use of NBD as an stochastic model for repeated purchases was first introduced in [82] in the context of consumer good purchases (in which the drop-out models in this case the possibility of brand switching). It rests under the assumption that purchases for each customer are independent and Poisson distributed with constant mean, and that this individual mean purchasing rate across the customer population follows, as was the case for the individual dropout rates, a Gamma distribution [83]. In this case, the NBD results from the continuous mixture of Poisson distributions for all customers.

The Gamma distribution (that models both player churn rate and purchasing frequency heterogeneity) in its shape and scale parameter parameterisation has probability density function:

$$f(x;k,\theta) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)} \tag{2.35}$$

where k is the shape parameter, θ the scale parameter, and $\Gamma(K)$ the gamma function evaluated at k. The maximum likelihood function of the Pareto/NBD model can therefore be written in terms of the four parameters (two scale and two shape) of the two gamma functions characterising both processes (churn and purchase frequency) across the population. These parameters will be estimated using the RFM data available: only three pieces of information from each player are needed: the length of the observed period, the number of transactions in that period, and the time of their last purchase. Once the parameters have been estimated (maximising the maximum likelihood function), LTV predictions for each player (conditional on their transaction history) can be produced.

2.5.2 *Other parametric models*

Other parametric models use the same approach as the Pareto/NBD model, i.e., parameterising both the drop out probability and the number of purchases and combining both distributions, substituting the Pareto and negative binomial distributions by others, in an attempt to either make the estimation of parameters more efficient, and/or to improve accuracy.

Replacing the Pareto distribution by a beta-geometric (BG) one, for example, simplifies dramatically the computations needed, while typically displaying similar performance to the Pareto benchmark [90]. It does not only improve computation efficiency, but is also much simpler to implement and makes the parameter search more robust. The only conceptual difference with respect to using the Pareto distribution is in that players are assumed to churn immediately after their last purchase (the Pareto allows for this to happen at any moment between the last purchase and the observation time). It does however rely partially in flawed logic, as players who only have purchased once will always be considered as active. This is solved using Markov-Bernouilli Geometric (MBG) distribution, which allows for zero repeat purchasers [16].

Not long after the NBD was first proposed to describe repeat purchase behaviour, and even if it was shown to successfully represent purchasing histograms in many empirical cases studied, observations appeared where the interpurchasing times appeared to be more regular than those Poisson distributed [134, 143, 176]. One alternative is replacing the Poisson's exponential on the individual interpurchasing times by a Gamma distribution with parameter a positive integer, the so called Erlang distribution [49, 143]. This is normally referred to in the marketing literature as the CNBD (condensed negative binomial distribution) purchasing model. Its dynamics are explored from a theoretical point of view in [248]. The varying degrees of regularity across costumers can be accounted for by using a mixture of Gamma distributions instead of the NBD, in what is normally referred to as Pareto/GGG models [228]. Another alternative explored has been the log-normal [176]. See [290] for a statistic on purchase regularity across consumers, that can be computed with only two interpurchase times per user.

2.6 DEEP LEARNING

Deep learning [178] refers to the family of ML models that make use of deep neural networks (DNN) and representation learning. *Artificial Neural Networks (ANN)* are systems inspired in biological neural networks, made up by connected artificial neurons that can process and transmit signals downstream (input(s) to output(s)). Neurons can have a state and an associated weight that is updated during the learning process. The particularity of ANNs is that they learn from examples without instructions on how to do this. In this sense, they are black box devices: they predict the output given the inputs, without any detail on how they have arrived to that result. The *deep* in the DNN simply refers to architectures where there have many layers of neurons between the input and output layers. Representation learning refers to systems that do not require manual feature engineering, as they are themselves capable of detecting or selecting the information needed from the raw data.

The concept of ANNs goes back to the seminal work of McCulloch and Pitts in 1943 [194], with multilayer perceptron being introduced by Ivakhnenko and Lapa in 1967 [158]. After several spikes of popularity, the deep learning revolution started at the beginning of the 2010s, when the computational capacity, data availability, and sophistication of the architectures and learning algorithms (backpropagation, ...) allowed for one breakthrough after the other. DNNs can currently match and outperform human experts in many tasks in which this was unimaginable a decade ago (image classification, board and video game playing, ...), and are known to provide the best ML solutions (in terms of accuracy, albeit not of explainability), with extraordinary performance, to a huge range of problems, from speech [130] and image recognition [87], to natural language processing, drug design, fraud detection and even classification of particle physics experiment results [57], genomics [297] or electronic health records [232].

In predictive analytics for video games, deep learning has been used successfully to tackle a variety of problems. See for example Kim et al. [168] for a churn prediction example, or Guitart et al. for in-game event simulation for AoI [131]. In Sifa et al. [257] deep multilayer perceptron (DMLP) networks were used to predict expenditure in the game within the next year.

2.6.1 *Multilayer perceptron*

The *deep multilayer perceptron (DMLP)* is a class of feedforward network with layers made up of neurons with nonlinear activation functions. It is perhaps the simpler of DNNs. All neurons in each layer are connected to all neurons in the previous and next layer, starting with the input layer, going through several hidden layers, and resulting in the output layer, as is shown schematically in figure 2.1. The learning process is carried out in multiple iterations that receive the name of *epochs*, in which the gradient descent algorithm is used to update the weights between nodes in order to minimise the cost function (for example, the root mean square error). See for example [24] or [246] for more details.

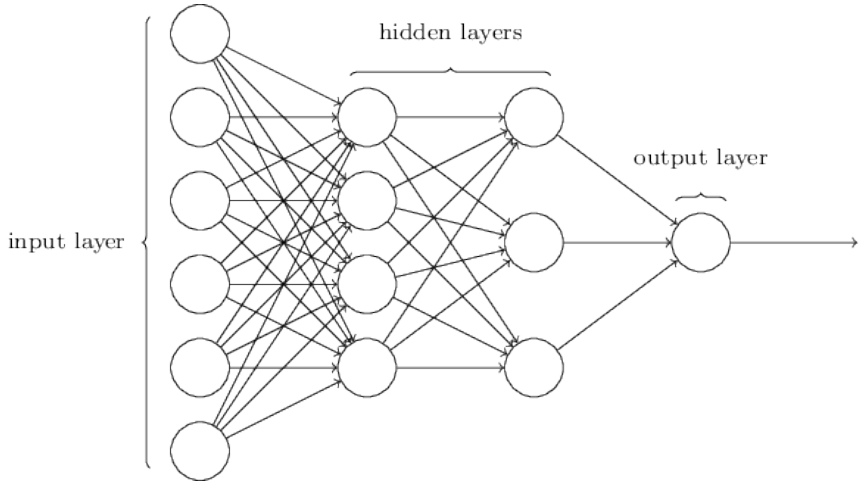


FIGURE 2.1: The structure of the multilayer perceptron network. Source: computersciencewiki.org

2.6.2 *Convolutional neural network*

A *convolutional neural network (CNN)* is a DNN architecture based on the animal visual cortex, and is a regularised version of the DMLP. they are typically formed by several convolutional layers, followed by pooling [245] and several fully connected layers, as shown schematically in figure 2.2. In the convolutional layers, filters or kernels are applied locally (to several adjoining inputs), generating successive abstracted feature maps. This means

that each neuron is connected only to a local region of the previous layer (instead of all neurons in the previous layer as is the case of the DMLP). This is what makes them particularly well suited to learn from inputs where this local connectivity is crucial, such as in image processing or time series data. They also need substantially less (or none at all) feature engineering

See for example [177, 179, 269, 279] for more details. Examples of their application to time series problems [177] include, for example, human activity identification from sensor data [299], system components useful remaining useful life estimation [241], or stock price prediction [278] or energy consumption [268, 304].

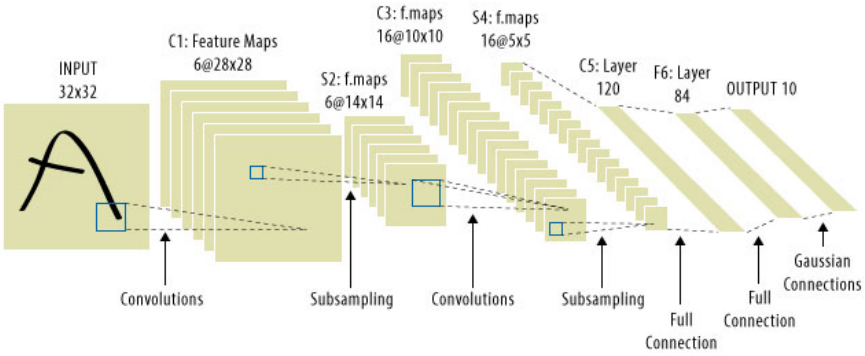


FIGURE 2.2: The structure of the convolutional neural network.

2.6.3 *Long-short term memory*

A *Long-short term memory (LSTM)* network is a DNN-like architecture particularly well suited to deal with sequential data, and was introduced by Hochreiter and Schmidhuber in [144] (which has become the most cited DL research paper of the 20^{th} century with over 26k citations). They are a type of *recurrent neural network (RNN)* designed to deal with long range dependencies in time series data. The information flow in the LSTM is controlled by its three gates. The input gate controls the new information getting it, the output gate the output activation, while the forget gate regulates which new information arriving is kept and which is discarded. This fixes, or at least minimises, the problem with exploding or vanishing gradients that can easily appear in simple RNN through backpropagation.

LSTMs are the cornerstone of natural language processing (NLP) [300], and everything language related. They have been successfully used for language translation and transcription [15, 173, 180, 296]. They can also be used to tackle basically any problem traditionally approached with classical time series forecasting, from traffic flow [298] or human trajectories [4], to stock prices [234]. See [254] for a detailed and pedagogical presentation of its fundamentals.

2.6.4 *Weight initialisation*

An adequate weight initialisation of weights is key to effective learning with NNs. Zero initial weights (no initialisation) can lead to very large or small loss gradients that could hinder convergence. Random initialisation can present similar problems, as too large or too small weights will prevent the NN from learning well. A method widely used today, which is also the method of choice in this thesis, was introduced in [120] by Glorot and Bengio. It is typically referred to as *Xavier initialisation*, and basically consists in sampling the initial weights from a uniform distribution bounded between $\pm\sqrt{(}6)/\sqrt{(}n_i + n_{i+1})$, with n_i and n_{i+1} the incoming and outgoing connections to the layer respectively. This choice makes the variance relatively stable across all layers and through the successive activation and backpropagation.

2.6.5 *Training algorithms*

There are many approaches to learning the optimal weights for the NN, most of them based in gradient descent [237]. The method used both in chapter 6 section 6.5 and in chapter 7 is *adaptive stochastic gradient descent optimisation* (Adam), as described in [169]. This method works well with large datasets and parameter spaces, is computationally efficient, with relatively low memory requirements, and its readily available for used in the deep learning framework used (see section 2.11).

2.7 CLUSTERING METHODS

Clustering algorithms aim at grouping objects according to their *similarity*. To decide on the similarity of the objects or items considered, they need to be defined using a set of properties or characteristics, that can be both numerical and categorical. These span the dimension space in which the

items *live*, and where a distance measure can be defined and used to look which objects are closer and which ones lay far apart. Clustering methods are an example of unsupervised learning, as there is no labelled data, and the algorithms used need to look for the structures in the data that make more sense.

Common distance measures used include the Euclidean distance, the Manhattan distance, the Hamming distance or the Levenshtein distance. The distance used in all cases throughout this work was *Gower's distance* [126], as it allows to use dichotomous (or binary), qualitative and quantitative (in the terminology used in the original paper, the three types are encompassed by numerical or categorical). It is defined as:

$$S_{ij} = \frac{\sum_{k=1}^{r} s_{ijk} w_k}{\sum_{k=i}^{r} \delta_{ijk} w_k} \tag{2.36}$$

where S_{ij} is the *similarity coefficient* between items i and j, s_{ijk} a score that quantifies the similarity of items i and j with regards to characteristic k (of which there are r), δ_{ijk} is either 1 (when items i and j can be compared using characteristic k) or 0 (otherwise), and $w_k \geq 0$ the weight given to the k characteristic (which is zero if and only if the the k property can not be compared for items i and j).

The following subsections briefly describe four algorithms that are used in this thesis: k-means, k-medoids, DBSCAN and HDBSCAN. The first two are centroid base methods, in that items are associated to cluster depending of their distance to a point of the dimension space that is the *centre* of that cluster. The last two methods are density based, in that clusters are defined depending on the local density of items in the dimension space. Each type of algorithm comes with its set of advantages and disadvantages that will be briefly discussed below, and the best choice is normally very problem dependent. A general detailed introduction to the subject of clustering can be found in [165], which also presents the centroid based methods. Clustering methodologies are also discussed in chapter 10 of [159].

2.7.1 *K-means*

The centroid around items or elements are clustered is the mean of the cluster. K-means proceeds by minimising the total squared error. Some of its weakness are that the it is unable to identify clusters of different shapes (it uses only the distance to a fixed point), and that the number of clusters needs to be identified beforehand. The latter can be of critical importance in

that, as all items are assigned to one of the clusters, it is unable to identify outliers, and can sometime detect spurious (unstable) clusters.

2.7.2 *K-medoids*

K-medoids works in a very similar way to K-means, but instead of clustering around the cluster average, it uses as centroid the element belonging to the cluster that is closest to the average. While more robust that k-means to outliers and noise, it still suffers from most of the same weakness. namely, it can not detect clusters with shapes not tending to be spherical, and the number of cluster needs to be predetermined (or the algorithm ran for many different number of clusters and then the optimal according to some validation metric selected).

2.7.3 *DBSCAN*

Density based spatial clustering of applications with noise (DBSCAN) [86] is a density based algorithm, which means that it scans the dimension space to look for areas where the density of items is greater, instead of looking only to the distance of items to a centroid. It can thus detect clusters of any shape and identify outliers, and the optimal number of clusters is found in the process instead of having to be given beforehand. It does however need two parameters to be manually fixed to run the algorithm: the neighbourhood radius (in which to look for items *close enough* to be considered as belonging to the same cluster), and the minimum number of items to be considered as defining a cluster.

2.7.4 *HDBSCAN*

HDBSCAN [44] is the hierarchical version of DBSCAN (the H stands for hierarchical), and has all of its advantages and in addition only one parameter needs to be fixed before running the algorithm: the minimum number of items in a cluster.

2.8 COLLABORATIVE FILTERING

Collaborative filtering (CF) is a broad term referring to a collection of recommendation algorithms. It was designed for systems were users can give

feedback on which content they like and which they do not (i.e., ratings), and then generate recommendations using only this user-rating matrix. This avoids then the need of building user or content features, which makes the method particularly well suited for cases with a lot of users and or possible recommendations. They can also be adapted to be used with implicit feedback, i.e., when there are no explicit ratings, some characteristic relating users and content (e.g. number of purchases, views, clicks...) can be used to define some implicit rating to build the user-rating matrix. CF algorithms basically aim at filling the blanks in the user-rating matrix, thus predicting how much each user would like items they have not rated/for which there is no implicit metric defined. See [267] for a *classical* review, and [54] for a more recent one.

The following subsections describe three of the most widely used CF algorithms: item-item nearest neighbour models, latent factor models and bayesian personalised ranking.

2.8.1 *Item-item nearest neighbour models*

In the spirit of clustering methods, an item-item similarity matrix is computed, where each item is represented by the collection (*vector*) of ratings from the different users. For this, a distance metric in the item space needs to be defined in terms of the (implicit or explicit) ratings, with cosine (cosine of the angle between the two item vectors), term frequency-inverse document frequency (TF-IDF, the frequency of occurrence of a rating scaled using how often it appears across users) or BM25 (another ranking method that uses the scaled occurrence rate, also taking into account the number of ratings associated to each user) being typical choices. This matrix is then used to both find items close to each given item, and find players with similar responses to the similar items, and then use this information to predict user's missing responses using weighted computations. Nearest neighbour methods make it very easy to give the reasoning under every recommendation, but have no flexibility to introduce a measure of confidence of implicit preferences assumed.

2.8.2 *Latent factors models (matrix factorisation)*

This algorithm performs matrix factorisation to express the response matrix as a product of a user and item matrix. If there are n users and m items, the response or rating matrix is nxm. The user matrix will then be nxk

and the item kxm, with k the number of latent (unobserved) factors that are assumed to govern the preference of users for items. This number is a modelling choice and breaks the original space of the problem into two of much smaller dimension that can be used to compute missing scores. Matrix factorisation is often carried out using the singular value decomposition (SVD) or the alternating least squares (ALS) methods. See [151, 272] for more details.

Matrix factorisation approaches have many advantages and they can be used directly over implicit feedback data. Overfitting is easy to avoid by selecting an adequate regularisation model, in which parameters are often learn by stochastic gradient descent. Compared to nearest neighbouring methods, the impact of missing data is smaller and they produce better results. It is on top of all that very fast, as it does not require estimating many parameters and does not involve lengthy computations, rendering it practical for many large scale applications. Its main disadvantage is that optimisation is carried out regarding only one item.

2.8.3 *Bayesian personalised ranking*

This method focuses on item-item ranking to output a personalised item ranking for each user. Optimisation is done for each user, and it can also deal adequately with missing values. It however always assume that a user will always prefer an item for which there already is a (positive) response than all those with no previous history (non-observed items), which is a very strong limitation depending on the intended use of the system. See [233] for details.

2.9 VALIDATION METRICS

Different metrics can be more or less useful depending on the context when trying to assess goodness of fit for estimations and predictions. For regression problems, given N entities of interest (players in all cases in this thesis), each with predicted or estimated value p_i, and observed or real value o_i:

- Mentions to *percentage error* in this work refer to the group deviation defined as:

$$\%error = \frac{\sum_{i=1}^{N}(p_i - o_i)}{No_{max}} \tag{2.37}$$

- The *mean absolute error (MAE)* is defined as:

$$MAE = \frac{\sum_{i=1}^{N} |p_i - o_i|}{N} \tag{2.38}$$

  I.e., it is the average value of all absolute errors.

- The *root mean squared error (RMSE)* is defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (p_i - o_i)}{N}} \tag{2.39}$$

  It does thus give an idea of the spread of the errors and is scale-dependent.

- The *root mean squared logarithmic error (RMSLE)* is

$$RMSLE = \sqrt{\frac{\sum_{i=1}^{N} (log(p_i + 1) - log(o_i + 1))}{N}} \tag{2.40}$$

  It is also scale-dependent, but does not over penalise large differences when both observation and prediction are huge. It also penalises more under predicted values than over predicted ones (in that for $y = log(x)$ there are larger Δys corresponding to the same Δys for smaller x value)

- The *normalised root mean squared error (MRMSE)* is

$$NRMSE = \frac{RMSE}{o_{max} - o_{min}} \tag{2.41}$$

  It is more appropriate than RMSE to compare datasets at difference scales.

- The *symmetric mean absolute percentage error (SMAPE)* is

$$SMAPE = \frac{100\%}{N} \frac{\sum_{i=1}^{N} |p_i - o_i|}{(|p_i| + |o_i|)/2} \tag{2.42}$$

  This accuracy measure is based on relative errors. It is invariant under linear rescaling and is not sensitive to outliers.

See [154] for an interesting discussion on some of these metrics.

For binary classification problems, the *receiver operating characteristic curve* (ROC) *area under the curve* (AUC) is a good metric to compare model performance [34]. The ROC curve is a plot of the true positive rate against the false positive rate for different thresholds. It thus contains information about both sensitivity and specificity. It can be used to decide upon the best threshold, and to compare performance of different models with respect to the aforementioned thresholds. To summarise goodness of classification in a single statistic, the area is generally a good option. See for example [34, 93] for more details.

The *Brier score* (BS) [37] is used in probabilistic forecasting to measure accuracy, and is normally defined as:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2 \tag{2.43}$$

where p_i is here the probability predicted for the entity (player) i, and o_i the observed binary value (0 if the event does not happen, 1 if it does). As in the metrics described above, N is the number of entities of interest (players in this case). In the case of survival models, one can define a BS at each time step. The *integrated Brier score* (IBS) will then give an summary statistic reflecting the overall accuracy in the probabilistic predictions across all times and can be written as

$$IBS = \frac{1}{max\{t_i\}}\int_0^{max\{t_i\}} BS(t)dt \tag{2.44}$$

where $BS(t)$ is the (time dependent) Brier score, and $max\{t_i\}$ the final time step considered. See for example [128, 203] for more details.

2.10 DISCRETE CHOICES AND THE ISING MODEL

The name of *discrete choice theory*, and the rigorous and systematic development of a consistent framework for its study, can probably be rightly attributed to economists such as Blume, Brock and Durlauf in the last decade of the 20^{th} century, who already noted its links to some statistical physics models when taking social interactions into account [31, 38, 79]. The choice making process is described by agents or individuals whose decision making process is aimed at maximising their payoffs, as characterised by a certain utility function. Similar setups as those used in this work however had already been introduced earlier on by Föllmer [97] and

Granovetter [129], and the Ising model had also already been explicitly proposed in similar contexts by some physicists [42, 107, 109]. Work previous to the aforementioned, such as that of the pioneer in statistical modelling and Nobel prize award winner McFadden [195, 196], typically did not take social interactions into consideration. There are some remarkable exceptions, such as the seminal paper by the too Nobel Laureate Schelling on racial segregation [242] (analysed from a statistical physics point of view in [67, 116, 170]) among other of his work [243, 244]. Worth noting are also Becker's contributions [21, 22] as one of the first authors to explicitly introduce these concepts to the study of sociology.

Although these approaches have been explored earlier and more extensively in the field of economics, it was in the context of the study of some sociological problems that the effect of these interactions was first introduced. This basically amounts to considering a social term in each individual's utility, by which it is also dependent on the perceived choices from other individuals. This is sometimes referred to in the literature as social utility with *externalities* or *interdependence*. Both terms are used differently in this section and in chapter 9, where the term interdependence makes reference to relations between different choices for each individual, or between the same choice in different groups, and externalities to factors unrelated to social or choice interaction.

The deep relation between some models of condensed matter and some formulations of discrete choice theory with social interactions is hardly surprising upon some reflection. It were actually the social scientists building the framework that first noted these similarities and exploited the tools of statistical physics to study their problems of interest [31, 38, 79], as will be outlined in next section 2.10.1. But it is in any case straight forward to use models of ferromagnetism to mimic individuals with a tendency to align their opinions (spins), with some of the earliest works in this sense having been contributed by the physical community [42, 107, 109]. Not only is it interesting to think what the social sciences models associated to well known models of condensed matter could be (and the wealth of knowledge ready available about them put to use immediately). The tools of statistical physics, designed to understand how the macroscopic (or collective in the socioeconomic context) characteristics emerge from the interactions of the micro constituents of the system (individual players in our particular case), can be applied to choice theory problems with no equivalence to a condensed matter model. An example of many such applications can be found in [208], where the *maximum entropy principle* is invoked in the discussion

of which functions are well suited to formally study market organisation in a non interaction framework, and arguing in favour of the logistic one.

The direct relation with Ising type models, and the discrete choice framework discussed in this section, is specifically designed to deal with binary choices, i.e., decisions for which there are only two possibilities. A lot of interesting problems obviously do not fall into this category, both in the realm of general social sciences (which party to vote, which brand of a particular product to buy, which degree to study at university, what type of transport to use, how many children to have, in which city or neighbourhood to live...), and in that of video games that is our specific object of interest (in which of the events that is going on to take part, which item to purchase, which of several actions to take, which of several opponents to face ...). The tools of statistical mechanics are suitable too to study these problems, and there are in fact models form condensed matter physics which could mimic several discrete choices (such as the Heisenberg [199] or Potts model [294]). They are however technically much more convoluted. As will be described with more detail in chapter 9, the aim of using these models in this thesis is that of acquiring qualitative insights about choice making in social video games (or social contexts generally speaking), so focusing on binary choices seems a good place to start. Further more, many interesting problems are indeed binary in nature, and many of the non binary can be reframed as such and still provide many interesting insights about the original problem. Again, examples abound in the social sciences: whether to vote or abstain, what to vote for in a no/yes referendum, whether to buy a certain product or not, whether to go to university or not, whether to use public transport or not, whether to have children or not, whether to live in urban or rural areas..., and more generally speaking, whether to believe or not in any particular precept, or to adopt or not any particular lifestyle trait. In the video game realm there are also countless examples of great interest, from whether to purchase any particular item, take part in any particular event, face any particular opponent, take any particular action..., to the even more basic choices, decisive for the success of any game, and that will be studied quantitatively in detail throughout this thesis, of whether to continue playing or not, and whether to purchase in a free2play game or not.

Indeed, the discrete choice theory framework has been used to study a wide variety of problems in the social and economic sciences, such as that of demand [124, 125, 188, 201, 287, 288], election results [26, 98], crime [119], fashion [209], musical choices [32], rumours [105] or political opinions [107,

201, 271]. It basically consists in treating individuals as agents whose decision making process is always aimed at maximising and individual utility function, that can have both a deterministic and an stochastic part. Although they do typically assume rational behaviour (which has proofed to be questionable when trying to understand certain human behavioural patterns), in that all agents are making choices *in their best interest*, there is room in both the utility definition and possible stochastic (uncertain) nature of the model to capture at least some of this apparent *irrationality*. Outside the realm of decision theory in a strict sense, similar setups have been used to study other socioeconomic questions which could also be relevant in the study of in-game behaviour. These include social learning, public goods games and resource allocation, and hierarchical structures and coalition formation among others [31, 46, 97, 111, 123, 170, 184, 188, 208, 217, 287]. Another particularly interesting related approach is the use of a diluted Blume-Capel model of 3-state sites, defined on different complex network topologies, to understand cooperation and organisation [95].

In this section the relationship between socioeconomic utility scenarios and statistical mechanical formulations as they were first discovered in the discrete choice theory framework are outlined in subsection 2.10.1. The very well known behaviour of the simplest model to be used when considering social interactions, the *Ising model*, is summarised and formulated in terms of choices in video games in subsection 2.10.2. Similarly, the consequences of using an extension of the Ising model that includes randomness in the individual preferences, the *random field Ising model (RFIM)* is briefly described in subsection 2.10.3.

2.10.1 *The socioeconomic utility scenario and statistical mechanics*

Consider a group of N individuals or agents, and let us consider the binary choice $s_i = \pm 1$ they need to make, where i denotes the individual making the choice. The binary choice problem then consists in defining a utility function V_i, that each individual will attempt to minimise through their decision making. Very generally, this could be any function depending on the choice of the given individual s_i, its belief or knowledge of the rest of the group's choices $\vec{s}$: $E_i(\vec{s})$, their personal preferences, characteristics or circumstances h_i, and a random shock $\epsilon_i(s_i)$ that can depend both on the individual and the particular choice they make. Considering this random noise to be additive yields an individual utility that can be expressed as:

$$V(s_i, h_i, E_i(\vec{s}), \epsilon_i(s_i)) = U(s_i, h_i, E_i(\vec{s})) + \epsilon_i(s_i) \tag{2.45}$$

where U will be denoted the *deterministic utility* and $\epsilon_i(s_i)$ as the *random utility*.

Following Durlauf, Brock and Blume [31, 38, 79], further additive assumptions allow to rewrite as the utility as:

$$V(s_i, h_i, E_i(\vec{s}), \epsilon_i(s_i)) = u(s_i, h_i) + S(s_i, E_i(\vec{s})) + \epsilon_i(s_i) \tag{2.46}$$

Here u is the so called *private deterministic utility*, which will be considered in what follows to be given by $u(s_i, h_i) = h_i s_i$, and $\epsilon_i(s_i)$ is the *private random utility*. The individual preference h_i gives the deterministic difference in payoffs in absence of social interactions, and is sometimes referred to as *idiosyncratic willingness to adopt (IWA)*[1]. The *social deterministic utility S* considered in [31, 38, 79] can be expressed as:

$$S(s_i, E_i(\vec{s})) = E_i\left(\sum_{j \neq i} \frac{J_{ij}}{2}(s_i - s_j)^2\right) \tag{2.47}$$

where $J_{ij} \geq 0$ is the strength of the coupling to individual j choice, i.e., the desire of the agent i to align its opinion to that of j.

The problem to be solved can be thus expressed as:

$$max_{s_i \in \{-1,1\}}\ h_i s_i + E_i\left(\sum_{j \neq i} \frac{J_{ij}}{2}(s_i - s_j)^2\right) + \epsilon_i(s_i) \tag{2.48}$$

As will be soon discussed, this can be interpreted as a ferromagnetism model where the first term is equivalent to an external field, the second to spin interactions, and the last one is random noise.

Now, assuming a logistic distribution for the difference of the random payoff terms $\epsilon_i(-1) - \epsilon_i(1)$

$$P(\epsilon_i(-1) - \epsilon_i(1) \leq z) = \frac{1}{1 + e^{-\beta_i z}} \tag{2.49}$$

where $\beta_i > 0$, and taking into account that $s_i^2 = 1$ the social utility term can be rewritten:

1 In demand contexts, it is sometimes useful to express the IWA as $h_i = b_i - p$, where p is the price of the item or product, and b_i the *idiosyncratic willingness to pay (IWP)* of individual i. This allows for the study of the *demand curves* or dependence of the demand with the price [124, 125, 251, 287].

$$S(s_i, J_{ij}, E_i(s_j)) = \sum_{j \neq i} J_{ij}(s_i E_i(s_j) - 1) \tag{2.50}$$

the probability of each individual's choice conditioned to its own preferences and its perceptions of the rest can be expressed as

$$P(s_i|h_i, E_i(s_j)\ \forall j \neq i) = P(V(s_i = 1) > P(V(s_i = -1))) \tag{2.51}$$

$$= e^{(\beta_i h_i s_i + \sum_{j \neq i} \beta_i J_{ij} s_i E_i(s_j))} \tag{2.52}$$

The expectation value of the choice of the system $s = \frac{1}{N}\sum_i s_i$ can be computed as the possible values 1 and -1 multiplied by their respective probabilities, which can be computed using equation 2.51. The expected value of each individual's choice is thus

$$< s_i >= \tanh\left(\beta_i\left(h_i s_i + \sum_{j \neq i} J_{ij} s_i E_i(s_j)\right)\right) \tag{2.53}$$

yielding a set of N equations that characterise the system. This is equivalent to the Curie-Weiss model's equation of state

$$s = \tanh(\beta(h + J_0 s)) \tag{2.54}$$

after setting β to be constant over all agents by redefining the rest of the parameters, whenever $E_i(s_j) = s$, $J_{ij} = \frac{J_0}{N}$ and $h_i = h$ for all agents. That is, the system of N individuals making a binary choice, when each of them is equally influenced by all other members of the group, and where all members of the group have identical preferences, and their subjective expectations for the rest match the mathematical expectation (what is sometimes referred to as *rational expectations* in the social sciences literature), is completely equivalent to the mean field Ising model of ferromagnetism. The spin is then analogous to the binary choice, J_0 their total interaction, the identical preference or deterministic private utility is an external field h (which will be referred to as IWA or *opinion field*), and the expected choice is equivalent to the physical system's magnetisation. In this setting, maximising total utility is equivalent to minimising free energy, and the individual utility V is analogous to each particle's energy with changed sign. Private random utility is deeply related to the temperature. Using a constant $\beta = \frac{1}{K_B T}$ is equivalent to studying the Ising model in statistical equilibrium, i.e., to using the canonical ensemble. The use of this approach

is thus not really intended to study the dynamics of any such system (unless this is varying slowly from a state of thermodynamical equilibrium to another). Dynamical approaches to discrete choice making, sometimes referred to as *opinion dynamics*, exist and are of a lot of interest (see for example [26, 98, 105, 184, 209, 251, 265, 270, 271, 274, 287, 288, 295], or [46, 111] for reviews).

The system is represented by the mean field model because of the described rational expectations assumption and because there is identical interaction with all other members. As in the study of magnetisation, this is equivalent to making each individual interact with the average choice, i.e., with their accurate perception of how the group is divided between both options. This is certainly the case for many problems concerning video games in which players virtually interact and are thus exposed to the purchasing and playing decisions of their peers. The mean field approximation is well suited for the study of many interesting discrete choice theory problems where a tendency to mimic a general trend wants to be modelled, with the advantage of been analytically tractable.It has thus been profusely used in this context [32, 38, 79, 80, 104, 107, 109, 111, 124, 125, 208, 209, 217, 251]. There are many interesting works exploring what happens when these conditions are relaxed, be that in regards to the rational expectations themselves [60, 119, 217, 251], the use of non identical couplings [79, 85, 287], or through interaction with some local neighbourhood [26, 38, 79, 97, 98, 119, 123, 201, 265, 270, 271, 288], the use of dynamic ones [111, 274], or of interactions in a complex network [26, 123, 295]. This work is limited to the mean choice approximation only. It should however, as shown by ferromagnetism models, give a reasonable approximation in many cases where this does not hold (at least far enough from critical regions).

The less realistic approximation certainly appears to be that of constant IWA or constant external field, as it allows no room to encode the preference heterogeneity present in any group. Examples of non constant IWAs can be found in [38, 60, 104, 111, 124, 251]. While some randomness will be introduced in section 9.5 of chapter 9 (as an extension to interdependent choices of [124]), most of the discrete choice related results in this thesis refer to the homogeneous population case described by constant h, and a lot of it to the $h = 0$ case. It is an interesting starting point, particularly when studying two interdependent choices, as it allows to identify the regions of the parameter space where social and/or choice interaction can make a difference. In regards to the zero field case, this describes the case where individuals do not have any particular preference, and any payoff

is related to social gains. They could be loosely understood as describing *fashions* or *traditions*. In video game contexts, they can however be useful to understand situations in which the number of players opting for one particular option (taking part or not in an event for example), is indeed what provides most (if not all) of the attractiveness to other players. This situation can be described as corresponding to *homogeneous* (all agents have equal IWA and social utility term) *unbiased* (in that they do not have a favourite option outside what social interaction dictates) populations.

It is worth mentioning that the reasoning above holds for populations of any size. It is indeed an statistical model, so computed expectation values will be more accurate the larger the population is, but in no moment did the thermodynamic limit play a role in the deductions of this section. In the study that will follow for interdependent choices in chapter 9, however, the large population limit will be used, and thus finite size effects not consider (see [104, 109, 111] for a study of finite size effects in similar setups).

2.10.2 *The Ising model*

This section compiles well known results of the Ising model in mean field approximation, solved exactly in the 1930s, and discusses its implications when used to study binary choices. Results presented in this section were published during the first half of the 20^{th} century and can be found in many statistical physics textbooks. More details can be found, for example, in [19], while [40] provides a historical review. Work pointing out its implications when used to study problems in the social sciences context, which will also be summarised in this section, include [38, 79, 80, 104, 107, 109, 209].

The Hamiltonian of a single infinite range Ising model with constant external field for N particles with $s_i \pm 1$ is

$$H = -\frac{1}{N}\left(\sum_{(i,j)} J_{ij} s_i s_j\right) - \sum_i h_i s_i = -\frac{1}{2N}\left(J_0 \sum_{i \neq j} s_i s_j\right) - h\sum_i s_i \qquad (2.55)$$

where the sums on i are over all N agents, sums on (i,j) over all possible $\frac{N(N-1)}{2}$ different pairs of agents ($1 \leq i < j \leq N$) and sums on $i \neq j$ over all pairs ($1 \leq i \leq N, 1 \leq j \leq N, i \neq j$). The spin s_i represent agent i choice ($s_i = +1$, decide in favour; $s_i = -1$, decide against), J_{ij} is the spin coupling (or that between agent i's decision and agent j's decision), and $h_i = h$ a constant external field. We will be considering identical coupling between

all individuals ($J_{ij} = J_0$ for all i, j). Note that this Hamiltonian is extensive (both terms scale as $\sim N$).

As has been discussed in the previous section 2.10, this is equivalent to considering homogeneous populations, in that all individuals are therefore subject to identical opinion fields, and are subject to social interactions of the same sign and strength with the same agents (all the rest). Heterogeneity can be introduced in both the coupling and/or the field through the use of random fields and/or spin glass models, that have also been extensively studied in the physical literature [29, 59, 121, 200, 222, 301].

In the mean field approximation equation 2.55 takes the form:

$$H \approx \frac{1}{2} J s^2 - \left(\frac{1}{N} J s + h \right) \sum_i s_i \tag{2.56}$$

where $s = \frac{1}{N} \sum_{i=1}^{N} s_i$ is the average magnetisation or choice, and $J = \sum_{j(\neq i)} J_{ij} = (N-1) J_0$. In the thermodynamic limit this yields

$$H \approx \frac{N}{2} J_0 s^2 - (J_0 s + h) \sum_i s_i \tag{2.57}$$

To compute the model's partition function in the canonical ensemble, the 2^N possible spin configurations is weighted with the corresponding Boltzmann distribution

$$Z = Tr e^{-\beta H} = e^{-\frac{\beta}{2} J s^2} \left(2 \cosh \left(\beta (\frac{J}{N} s + h) \right) \right)^N \tag{2.58}$$

The free energy can therefore be written as

$$F = \frac{1}{2} J s^2 - N \beta \ln \left(2 \cosh \left(\beta (\frac{J}{N} s + h) \right) \right) \tag{2.59}$$

and the free energy density of the system

$$f = \frac{1}{2N} J s^2 - \frac{1}{\beta} \ln \left(2 \cosh \left(\beta (\frac{J}{N} s + h \right) \right) \tag{2.60}$$

which in the thermodynamic limit can be expressed as

$$f = \frac{1}{2} J_0 s^2 - \frac{1}{\beta} \ln \left(2 \cosh \left(\beta \left(J_0 s + h \right) \right) \right) \tag{2.61}$$

The order parameter of this system is the average choice (or magnetisation) s. In the *paramagnetic* or *unpolarized phase* $s = 0$ and there is no *order*

or choice alignment in the group. Note that we are discussing statistical equilibrium, so individuals can change their opinions, but on average half of the population will be deciding against and half in favour. In the only other possible phase, the *ferromagnetic* or *polarised phase*[2], $s = \pm m$ with $0 < m \leq 1$. In this case the individuals to align their choices, and order emerges.

Minimising the free energy yields the equation of state (first arrived at by Bragg and William in 1934):

$$s = \tanh[\beta(J_0 s + h)] \tag{2.62}$$

for $J_0 \neq 0$. At high temperatures ($\beta \to 0$) the system will be in paramagnetic (no prevalence of a particular option) even in the presence of large fields or preferences h. For $T = 0$ however, $\beta \to \infty$ and the whole population will align their choices ($s = \pm 1$). For all other temperatures the system can only be in paramagnetic state if $h = 0$.

There is a continuous second order phase transition that can be studied by linearized the equation of state 2.62 for $h = 0$ and $s << 1$

$$s = \beta J_0 s + O(s^3) \tag{2.63}$$

yielding therefore $T_c = \frac{J_0}{k_B}$ (for constant J_0) or $J_{0c} = \frac{T}{k_B}$ (for constant T) as critical values for $h = 0$. For $J_0\beta < 1$ the only stable state is paramagnetic, which becomes unstable (a local maximum of the free energy) for $J_0\beta < 1$, where two physically equivalent (equally probable) with finite spontaneous magnetisation or choice alignment (of same absolute value and opposing sign) are the equilibrium states. Magnetisation is plotted $J_0\beta$ in black for the $h = 0$ case in figure 2.3 (both plots).

Now let us consider the $h \neq 0$ case. For $J_0\beta < 1$ the minimum of the free energy 2.60 moves from $s = 0$ to a finite value of absolute value $0 < m \leq 1$ with sign that of the h (ferromagnetic phase). Here there is no phase transition in strict sense, but rather a continuous change through the values of the average choice when varying h, with $s = 0$ corresponding to $h = 0$. Figure 2.4 plot (a) shows how the magnetisation or average choice varies with h for $J_0\beta < 1$.

For $J_0\beta < 1$ but low enough absolute values of h, the equation of state 2.62 has now three critical points, two of them local minima of the free energy with a local maximum in between. The multiple equilibria regime

2 For the case of negative coupling J, which will not be considered here, it would be *anti-ferromagnetic phase*.

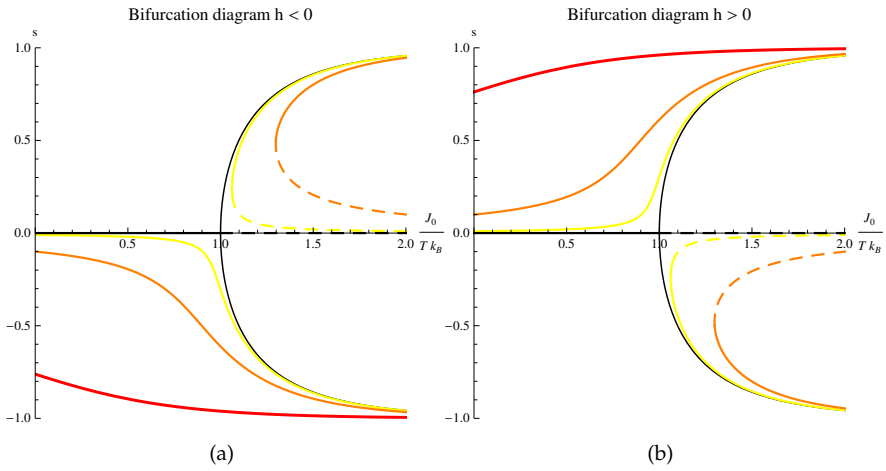


FIGURE 2.3: Average magnetisation is plotted against βJ_0 for (a) Different negative values of h (b) Different positive values of h. Black line $h = 0$, yellow $h = 0.01$, orange $h = 0.1$ and red $h = 1$. Dashed lines represent non stable critical points. Own elaboration. The image has previously appeared in [71].

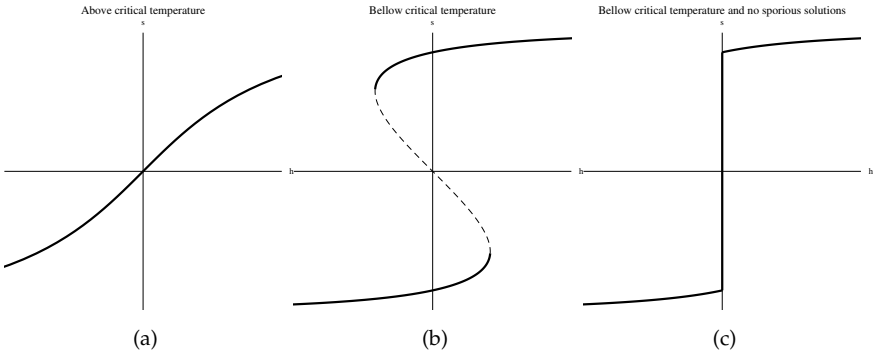


FIGURE 2.4: Average magnetisation is drawn against h for (a) $J_0 < k_B T$ and (b) $J_0 > k_B T$ at a fixed finite temperature and coupling. (c) Shows (b) removing metastable or spurious solutions. Own elaboration. The image has previously appeared in [71].

is not broken (for weak IWA) but rather shifted, and while both local minima have still opposing s signs, they are now not symmetric and have different absolute values. Unlike in the $h = 0$ case however, these states are not physically equivalent. Only one of them is the absolute minimum of the free energy 2.60 (that with s aligned with h) and thus the actual *ground state* of the system, while the other is a so called *metastable state*. This means that a system at finite temperature (that allows for fluctuations) will always eventually end up in the ground state, which is the attractor for a larger region of the $s - h$ plane. Depending however, on where the system was prior to the onset of the field, and the relative magnitude of this and the temperature, the system could end up in a metastable state for a relatively long time, so their existence can actually be very relevant. The dependence of a system's state on its previous history is known as *hysteresis*. This implies that, in this situation, changing in the system can not be undone by reversing the process. For large enough values of the field h however, there is only one possible state, always ferromagnetic, with magnetisation of the same sign as the field, and higher absolute value than that of the $h = 0$ case. This situation is depicted in figure 2.4 plot (b), while plot (c) shows the same relation removing spurious solutions (metastable states that will eventually decay to the ground state depicted).

Bellow the critical temperature, the average magnetisation or choice can be considered to made up by some due to the action of the field or IWA h, and some *spontaneous magnetisation* (*spontaneous choice alignment* in the decision context) emerging from the interactions

$$s_0 = \lim_{h \to 0^+} s(h, T) = \tanh(\beta J_0 m_0) \tag{2.64}$$

which behaves identically as the total magnetisation when $h = 0$ and becomes negligible at high enough temperatures (statistical fluctuations are too large for significant alignment to occur) or fields (that then dominate the choice alignment).

Figure 2.3 shows the dependence of the magnetisation s with $J_0\beta$ for different values of positive (a) and negative (b) fields, and illustrates the previous discussion. For $h = 0$ the system is either in paramagnetic state or in one of the two equally probable ferromagnetic states. For large enough fields there is only one solution, with magnetisation of the same sign as the field, an absolute value larger the lower the temperature is. For low enough values of h there are tow possible states, but in this case one of them is metastable.

The *phase diagram* of the mean field Ising model is therefore as follows. The parameter space is three dimensional h, J_0, K_BT. For $h = 0$, the (J_0, K_BT) cross section has a second order phase transition across the line $J_0 = K_BT$. For $K_BT > J_0$ the system is in paramagnetic state, and for $K_BT < J_0$ in ferromagnetic states with two physically equivalent possible states with different sign and same absolute value of the average magnetisation. Moving to the $h \neq 0$ sections, the system is always in ferromagnetic state with a single magnetisation ground state of the same sign as the field, although for sections of low enough h there is an additional metastable state. When looking at the other two sections $h - K_BT$ and $h - J_0$ are also ferromagnetic except for a paramagnetic segment for $h = 0$ delimited by the critical point $J_0 = K_BT$. Across this segment, there is a first order phase transition in h as the magnetisation changes abruptly. Above the critical temperature however, the change is smooth and there is no phase transition.

The key features of the Ising model from an statistical physics point of view, i.e., those emerging from the interactions between the constituents, are the existence of both first and second order transitions, and of metastability and hysteresis. These have profound implications when discussing the binary choices under social influence equivalent:

- *Microeconomic specification of the model that may not uniquely determine its macroeconomic properties*, as there is more than one possible collective state for weak enough private deterministic utilities (although only equally probable and thus stable for zero private deterministic utilities). If using this approach to study demand, this means that there can be both a high and low demand possible states associated to the same price.

- *Regions where social utility counts and regions where it does not*. There are distinct phases depending on the parameter values. In some only private deterministic utility rules the outcome, while in others social influence (spontaneous magnetisation) can have a decisive impact the decision making process.

It is also worth stressing the role played by the $\beta = 1/K_BT$ as some sort of *social permeability* [107], in that it is the inverse *socioeconomic temperature* T. The latter accounts for the possibility of statistical fluctuations, hence codifying the uncertainty about individual choices. It is a measure of how likely it is for an individual to make a choice in contradiction to what their deterministic utility is dictating as providing a better payoff. Depending on the intended use of the model and/or the system it is meant

to represent, it can have several interpretations. It could simply reflect the lack of knowledge about all details affecting the individual's decision process. It could have however a more transcendental meaning in terms of *free will* or a fundamental impossibility in predicting individual choices. If we consider all rational elements are included in the utility, then it measures, in a very literal sense, the probability individuals have of making irrational choices. It introduces randomness in the system (varying with time or *annealed disorder* in condensed matter phrasing) even when couplings an IWAs are constant. As the introduction of more and more randomness tends to make the problems more difficult to formally understand, many of the works considering heterogeneous private deterministic utilities (fixed in time or *quenched disorder* in statistical physics language), do so fixing $T = 0$, i.e., in the deterministic or purely rational case. These correspond to Nash equilibria [210, 211] when describing socieconomic phenomena [3]. This will also be the case of section 9.5 chapter 9 in this thesis.

Note that while the K_B will be explicitly written when it appears, the use of dimensionless spin or decision variables s_i is equivalent to the natural units choice $h = K_B = 1$ in the ferromagnetic counterpart. As a result, in the discrete choice model utility, the utility, social coupling and opinion fields are expressed in the same units. For many problems of interest these will be dimensionless, and the payoffs can be thought of as representing abstract qualities such as happiness, well being, satisfaction or reputation. For other problems, they will represent concrete surpluses or deficits, for example in many in practically all systems studying demand. In the case of video games, depending on the specific use, units such as experience points, in-game lives, playtime, in-game (as well as real) currency... could be of interest depending on the decision under study.

The metastable states have an interesting interpretation in binary choice contexts. It represents a collective state that can persist (at least for a while) in time due to individuals picking a choice that is actually not in their best interest, and this due to social interactions with their peers and a previous history where that choice was in fact the best option. Durlauf and Brock used this to reconcile both typical explanations of *social pathologies* (say, crime or school dropout prevalence), that of economic fundamentals (i.e., people for which crime is indeed the *best option* from a utility point of view) and social norms (i.e., where the situation is so pervasive in their

3 The Nash equilibrium of a non-cooperative game (concept stemming from game theory) is the solution in (static) equilibrium where all players are aware of the choice of strategy of the rest, and no player has anything to gain from changing their own strategy

environment that they continue to pick this option by imitation even if they would be better off deciding against) [79].

In the context of video games it can similarly be used to understand player behaviour. Imagine there are is an event in which players can decide take part and face other players, and that in doing so they can in principle earn higher rewards or progress more rapidly in the game than by not taking part in it. In this situation, it is to be expected than more than half of the players will be taking part in the event (and a larger fraction the smaller the statistical fluctuations considered). Imagine this situation changes slowly over the time, and the rewards associated tot taking part in the event (be that in terms of in-game currency, experience points or other) diminish until they are below than the payoff associated to playing without participating in the event. While the difference is still not too large (i.e., for weak IWA), it is possible that a large part of the players will still decide to join the event and perpetuate this collectively reinforced option that is now no longer in their best interest. Note that this social utility term can mimic the tendency to imitate (a player decides if most people are choosing this it must be better without analysing it properly), or can be due to a real payoff (players need to face other players to earn rewards anyway, and so participating may mean a greater variety of players of which to obtain these). Even in the latter case (and even more so in the former) are the players not choosing wisely. Even if due to the amount of players in the event the rewards are higher for each individual than if they do not take part, if most of them decided against (and thus socially reinforcing the option of not participating), the rewards would be even higher. Note however that the metastable character of this state anticipates that, eventually, enough players will change their mind to make the situation collapse and the collective outcome change drastically (first order phase transition).

2.10.3 *The random field Ising model*

As has been already noted, the introduction of randomness in the external field (affecting the private utility term) or coupling (affecting the social utility term), can be used to model the varying individual preferences and interactions across the population. Models of ferromagnetism including this type of randomness or *quenched disorder* have already been extensively studied in physics [59, 121, 200, 222, 301].

In this section, the implications of using a *Random Field Ising Model (RFIM)* instead of the conventional Ising model to study a choice making problem

are outlined. It makes use explicitly of the socioeconomic rather than the statistical physics formulation used in the previous section, following Gordon et al. [124].

Consider the system made up by N individuals making a binary choice. If agent i's choice w_i can take values 1 (codifying adoption) and 0 (non adoption), and depends on their individual preference given by their IWA H_i, and their propensity to imitate agent j J_{ij}, their utility can be expressed as:

$$U_i = \left(H_i - P + \sum_{j \in n_i} J_{ij} E_i(w_j) \right) w_i + \epsilon_i(w_i) \tag{2.65}$$

where n_i is the individual's neighbourhood (made up by the agents with whom they are prone to align their choice), $E_i(w_j)$ represents their expectation on agent j's choice, and $\epsilon_i(w_i)$ is a random shock allowing for uncertainty in the individual decision making process [4]. The price P is included in case the decision under study concerns a purchase (in which case H_i is the IWP). If this is not the case it can be simply set to 0. If the difference of random shocks for the two options follow a logistic distribution across the population:

$$\mathcal{P}(\epsilon_i(w_i = 0) - \epsilon_i(w_i = 1) \leq z) = \frac{1}{1 + e^{-\beta z}} \tag{2.66}$$

the resulting system is equivalent to considering a RFIM.

Considering that the all individuals have rational expectations $E_i(w_j) = w_j$, and that all players interact with all others with the same strength, is equivalent to studying the RFIM in mean field approximation. IWAs are considered to be independent identically distributed (iid) variables characterising the population towards the decision making process. What follows considers the zero temperature case only, i.e., the system is deterministic and all agents completely rational. In this case, the utility of equation 2.65 can be written as

$$U_i = S_i w_i \tag{2.67}$$

4 Note the use of a different notation from the previous section for the binary choice, that will be maintained throughout this thesis, indicating the options here are codified as 0 or 1 (instead of -1 and 1). This can be convenient as it very easily codifies non-adoption/adoption, making the average choice identical to the fraction of adopters. Both notations are used in this sections to facilitate comparisons to different previous works of which some of the content of this thesis is an extension.

where S_i is the *surplus* individual i will gain from adopting, and $S_i = H_i - P + J\mu$ for $\mu = \frac{1}{N}\sum_i w_i$ the fraction of adopters and expected choice. To maximise their utility, players will decide depending on the sign of their surplus, and in the thermodynamic limit

$$\mu = \mathcal{P}(H_i - P + J\mu > 0) = \mathcal{P}(H_i - H > S) \tag{2.68}$$

where H is the average opinion field or IWB, and $S = H - P + J\mu$ the average surplus.

A useful normalisation of the parameters is given by dividing all of them by the IWB's standard deviation σ: $j = J/\sigma$, $h = H/\sigma$ and $p = P/\sigma$, which allow for the expression of all relevant population's characteristics (or the exogenous price) to be measured in terms of the typical scale of the preference distribution.

As is discussed in detail in [124], based only on properties probability distributions, it can be inferred that the demand is a decreasing function with the prices. For smooth unimodal IWB distributions, there will be a region (for large enough j) of the $h - p$-j phase diagram where a low and high demand solution exist (these are in fact multiple Nash equilibria), i.e., demand curves can be multiple valued and non-monotonic. Outside this region, there is a single well defined equilibrium, and a phase transition takes place at its borders. For multimodal distributions, there will be several multiple equilibria regions, with the possibility of more than two of these to simultaneously exist.

2.11 SOFTWARE USED

Analysis and predictions carried out with SSMs (chapter 4) were done with the datetime, numpy [216], pandas [198] and statsmodels [250] libraries for Python.

All survival analysis (used throughout chapters 5 and 6) predictions were carried out using the pysurvival [99] and lifelines [69] python libraries.

Conditional inference ensembles were used for binary classification (chapter 5 section 5.3 and chapter 6 section 6.4) using the cforest implementation of the partykit R package [148].

LTV computations (chapter 6 section 6.5) using parametric models were performed using the BTYDplus package for R [227], and with TensorFlow [1] for Python when using deep neural network architectures.

Clustering (chapter 8) was performed using the pyclustering (for k-means and k-medoids) [215] and hdbscan (for the HDBSCAN algorithm) [197].

Gower's distance used in the clustering was computed using the cluster package in R. The CF was performed using the implicit python library [100]. The ERT ensembles and and XGBoost used in the same chapter were used as implemented in the R packages extraTrees [258] and xgboost [55] respectively.

Python libraries used to carry out the numerical analysis using the Newton-Raphson algorithm in chapter 9 sections 9.2.3 and 9.3.3 can be found in https://github.com/anafrio/matmetpy/rootAlgo.py.

All plots were produced using the matplotlib [152] library with python, the ggplot2 [291] package for R, or gnuplot [292].

3

BASIC PLAYER PROFILING

Always remember that you are absolutely unique. Just like everyone else.

— Margaret Mead

A common approach to the study of human behaviour in different contexts includes their grouping in a meaningful way, such that it facilitates the understanding of the behaviour under study. It can be done in terms of a single characteristic or combining multiple (with the help, for example, of the clustering methods described in chapter 2 section 2.7), and using demographic, geographic, behavioural and social variables, among others. This process is sometimes referred to as *profiling* or *segmentation*. The terminology comes from marketing studies, with mentions of the term *market segmentation* appearing in the literature as early as 1956 [261], and where it has continued to be an object of extensive study (see for example [20] for an earlier review, or [127] for a more recent one). The definition of certain groups of interest to better understand a particular problem is however obviously ubiquitous in the social sciences as an approach.

In video games, the term usually used is that of *player profiling* or *player clustering*. Note this generally refers to groups defined in terms of the in-game player behaviour or characteristics, rather than those of the actual physical player (although for some characteristics this is indistinguishable, and for others there will obviously be interesting correlations between both). In this chapter, a very basic segmentation is defined, that allows for the classification of players depending on whether they are active or not, and whether they are purchasing or not. It is in some sense more of a compilation of basic terminology than an actual profiling, but as will be seen, it is not as straight forward or exempt of subtleties as it may seem at first. It can also be the starting point, in a top bottom approach, to building a structure with different levels of segmentation complexity, by further subdividing and combining the groups described in what follows.

Other interesting attempts at player profiling can be found in [17, 62, 77, 78, 239]. In chapter 7 the use of engagement predictions to generate meaningful groupings is discussed, as was first proposed in [72]. The concepts and categorizations described in this chapter are outlined in [132].

Most of them, however, were more or less explicitly being used in previously published work, so should not be considered as original contributions of this thesis.

3.1 PURCHASING BEHAVIOUR

The video game business model has changed very much in the last years, with most titles being online and free-to-play, and most of the revenue coming from in-app purchases. Typically, in these games, less than 5% of the players will ever make a purchase (around 1,5% for AoI, see chapter 1 section 1.2 for more details). This makes one very important distinction that will be made between players be that of whether they are *paying users* (PUs, those that have made at least one purchase), or *non paying users* (non-PUs, those that have never purchased in the game).

This of course does not mean that PUs are frequently purchasing. To distinguish those that are actively spending the concept of *purchase churn* will be introduced in section 3.4.

While in regards to AoI throughout this thesis, anything referring to spending should be understood as actual purchases (in real money), the same logic can be applied to *virtual purchases* (in virtual in-game currency). Depending on the game, virtual currency can be purchased and/or earned in the game. In some titles it will be essential for in-game progression, in others it will only allow for certain customisation and quirks. In some role type games, it very also be of utmost importance to understanding game dynamics and player behaviour for many such games. There are however no virtual sales in AoI, so no data on in-game virtual spending will be analysed or modelled. Much to most of what will be mentioned throughout this thesis in regards to purchasing engagement and behaviour, however, can be directly applied to its virtual counterpart in games where such data is relevant and available.

3.1.1 *VIP players or whales*

It is also interesting in many exercises to focus on the top spenders, which receive the name of *VIP players* or *whales*. Not only are these of obvious interest to studios, they literally are the group of players more invested in the game, and thus can shed light on the behaviour of very engaged players.

In this work, a player becomes VIP after their accumulated expenditure reaches a certain amount, that will be referred to as *whale* or *VIP threshold*. This threshold will vary from game to game, and is computed using the first two months of data, so that whales provide at least 50% of the revenue in that period. This means there are approximately 6000 VIP players in the studied dataset of AoI, i.e., they make around 20% of all PUs.

Figure 3.1 shows the estimated probability distribution (using kernel density estimation [223, 235]) of the total sales. The x-axis represents the total number of sales in yens (and is plotted using logarithmic scale), with the area under each curve normalised to one. It is obvious that the distributions are markedly different (although there is naturally some overlap). As has been briefly described in chapter 2 section 2.5, and will be discussed in detail together with several methods to predict it in chapter 6 section 6.5, LTV (lifetime value) refers to the total outlay of an individual player throughout their complete history in the game. Figure 3.2 shows PUs and whales as a function of their normalised LTV. Note that even within whales, only a small fraction of them fall within the last decile. These players are of course very relevant to the game's revenue, but they are also those that should be tracked carefully, as some could be displaying problematic and/or addictive spending behaviour that should be addressed.

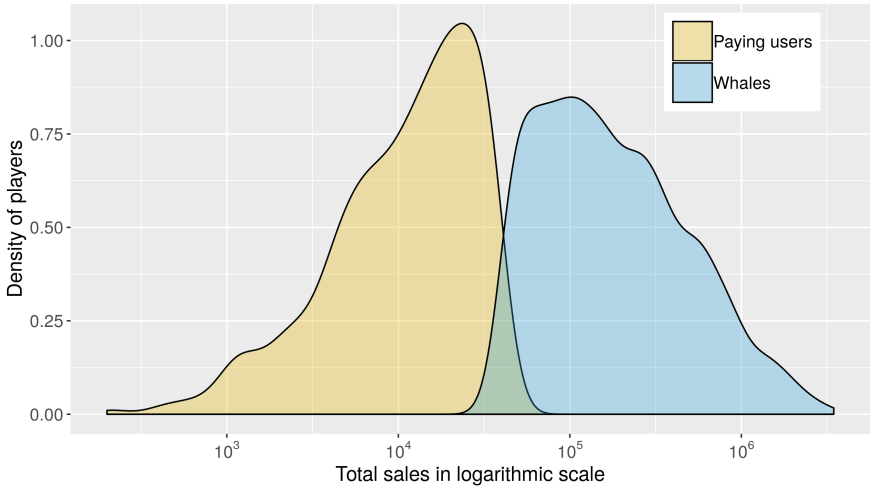


FIGURE 3.1: Probability density function from the kernel density estimation of total sales for paying users (in yellow) and whales (in blue). Elaboration using data from AoI. The image has previously appeared in [51, 224].

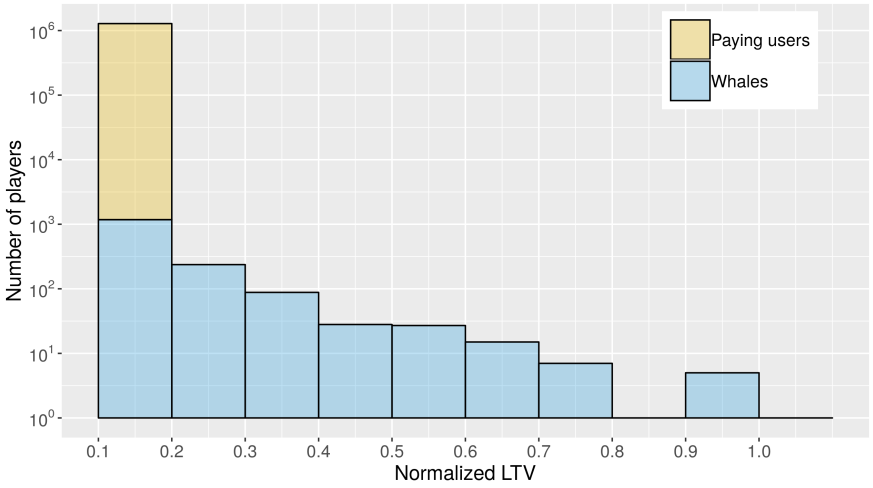


FIGURE 3.2: Histogram of the number of paying users and whales by normalised LTV (lifetime value) between 0 and 1. Paying users are shown in yellow and VIP players or whales in blue. Elaboration using data from AoI. The image has previously appeared in [51, 224].

Note both PU and VIP definitions imply that players acquiring this status will never loose it, regardless of whether their spending patterns change. That is, until they quit the game.

3.2 ACTIVE PLAYERS AND CHURN

If a *player* is defined as any individual account that has ever logged into the game, probably the most important distinction or basic segmentation that can be considered is whether the player is still *active*, or if they have already quit the game or *churned* and are *inactive* (which will also be occasionally referred to as *dead* players).

Churn is a topic of great importance when evaluating the success and health of customer service focused businesses. A high retention rate (or low churn) indicates that customer's loyalty or satisfaction with the service is high. It has been thus widely studied in very different contexts including telecommunications [153, 206], retail banking [207], and of course, gaming [28, 135, 167, 224, 225, 238]. In free-to-play games, acquiring new users is typically more costly than retaining players [96], making the study of churn and retention, and the understanding of its drivers, crucial.

While for all services where the relation's duration is set up contractually the definition of churn is completely straight forward (see for example [206]), that is not the case of gaming (and other commercial activities). Most players stop playing and effectively quit the game without deleting their account. Players are then considered to have churned after a period of inactivity. Fixed time windows have been used for some particular cases [167], tagging for example as churners all players that have not logged in during the current month. Typically, however, the churn-defining inactivity period is defined using a moving time window for each player. The latter approach, which is more costly in terms of computation but normally more adequate, is the one considered throughout this thesis, and in previous work such as that described in [28, 224, 225, 238].

The duration of the inactivity period to be considered is in itself a very interesting problem. It will vary greatly for different titles. While in, for example, very casual games with typically short player lifespans, a few days of inactivity will allow for a successful detection of real churn, in role multiplayer games the period needs to be much larger for a correct identification, typically of over a month. It will also depend on the particular use of the segmentation and if it should favour early or accurate detection. Obviously, the longer the inactivity period considered, the more accurate the classification will be. But this also means that longer periods are required to consider players churned, players that might very have effectively lost any interest in playing again long before they are marked as inactive. This inactivity period length, that will be referred to as *churn definition*, can therefore be set more or less restrictively depending on the aim of the study.

Here it is done after analysing the *percentage of false churners* (percentage of players that are considered to have churned but eventually go back to playing) and the *percentage of missed sales* (percentage of the total sales made by false churners after returning to the game), as is discussed in detail in [132]. Figure 3.3 shows these two indicators (percentage of missed sales in plot (a) and of false churners in plot (b)) for AoI, for varying lengths of the churn definition using the first two months of data, considering all PUs (red) and only VIP players (blue). The dashed lines mark the shortest possible churn definition for each of these groups for which missed sales remain below 1% and false churners below 5%, namely, 13 days for PUs and 9 days for whales.

Obviously, the percentage of false churners will normally increase if we consider longer periods than two months for its computation. Most of the studies in this thesis however, deal with the use of modelling in

production setups, where having a definition as soon as possible is of utmost importance. It was however checked in this case, what the results of extending the analysis to longer periods of up to 6 months would be, trying also different starting dates across the available dataset. The increase in percentage of false churners was not found to be significant, staying always below 10% (when considering 13 days and 9 days for whales and PUs respectively). This indicates a two month data sample for a game like AoI is representative of the player's churn behaviour, pointing at the validity of our approach. In particular, figure 3.4 (a) plots the fraction of all real churners captured for different lengths of the churn definition window using all available data. Note that it just falls short of 90% for 9 days even in this case.

Note how different groups of players can be characterised by their churn definition (given the required thresholds for the two indicators considered). Players who are typically more active and engaged in the game will have shorter churn definitions. A relative short period of inactivity, that might be typical for less active users, signals for them a genuine loss of interest in the game.

Unless stated otherwise, the churn definition used throughout this thesis for all AoI players will be of 9 days. This length is considered restrictive enough, as it still keeps the percentage of false churners below 10% and of missed sales below 1.5% when considering all PUs.

3.3 PROFILING CHURNERS

Players can be further profiled depending on their churn behaviour. It is particularly interesting, as will become clear when dealing with individual player predictions (see chapter 5 section 5.3 and chapter 6 section 6.4), to characterise false churners. In particular, the following different groups will be considered for some exercises:

- *Resurrected* players: These are a subset of the false churners. They are players that have remained inactive for a very long period of time (several times the churn definition window), but then go back to the game at some point. They can actually be considered as having really churned, only to recover their interest later on, hence the name. For AoI the inactivity period considered is of 30 days.

- *Zombies*: Players that present such a disengaged behaviour and play so sporadically that, regardless of whether they are on paper false

churners, should not be considered as rightly active. Throughout this thesis players will be considered zombies when they present less than 3 hours of accumulated playtime, no level-ups and no purchases in the previous 30 days.

- *Genuine false churners*: False churners that are not resurrected or zombies. These are players that in spite of having been inactive long enough to meet the churn definition, can be considered as never really having left the game, but rather, as having been mistakenly taken for churners.

Note that in the analysis carried out in section 3.2, the indicator that should be kept as low as possible is that of percentage of genuine false churners (as opposed to the general percentage of (all) false churners). Although the percentage of all false churners may become larger as the period considered increases, it could be related to an increase only in resurrected players. Further more, these are typically at the end of their life cycle anyway. Although some resurrected players may really become completely re-engaged an say, live a *second life* in the game, most recover interest only momentarily before churning definitely (or becoming zombies) soon after. They therefore do not contribute significantly to the increase of the false churner rate in the long run.

3.4 GENUINE PAYING USERS AND PURCHASE CHURN

As was mentioned briefly in section 3.1, it is sometimes interesting to distinguish players that are currently purchasing from the rest of PUs. These players will be referred to as *genuine paying users*. As was the case with login activity, the definition of who is an active spender is in itself interesting.

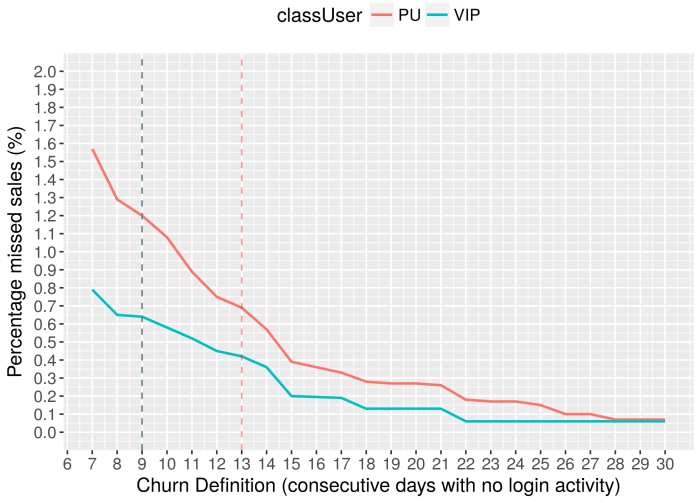
A user becomes a PU by making a purchase. Analogous to (login) churn discussed in section 3.2, a period without spending is defined such that it renders PUs non-PUs again, process that will be referred to as *purchase churn*. Similarly to the case of churn (or login churn to be precise), purchase churn should be defined according to the game and object of the classification in mind. In this work, and analogously to churn, it is set analysing the *percentage of false purchase churners* (percentage of purchase churners that purchase again) and *percentage of missed sales to false purchase churners* (percentage of the total sales coming from false purchase churners once they start purchasing again), as discussed in [132]. The latter is in this

case arguably the most important indicator (as the focus is on purchasing behaviour), and is shown in Figure 3.5. Percentage of missed sales is plotted for varying lengths of purchase churn definition window, for VIP players in blue and for all PUs is red. Dashed lines indicate the shortest possible windows yielding missed sales below 1%, namely 60 days for whales and 70 days for all PUs.

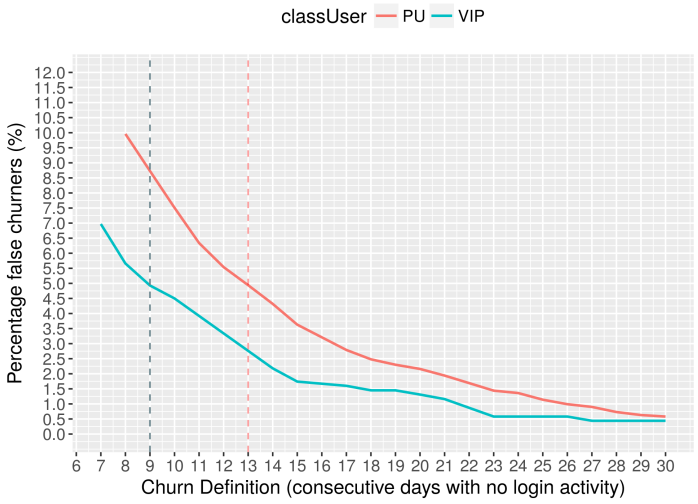
The inactivity period required here is much larger than in the case of login churn. It follows that a sample corresponding to a much larger period of time (around 6 times longer) would be needed to determine it. It is however impractical to wait that long if the definition is intended for operational use. The purchase churn definition can be therefore set from the beginning based on data previously collected for similar titles, or as rough multiple of the churn definition determined with the first few months of data. It can then be revisited later on if it is deemed convenient. Unless stated otherwise, throughout this work the purchase churn definition used will be of 50 days. While sometimes for convenience genuine PUs will be referred to simply as PUs, it will always be stated clearly if the PU definition being used is the general one (all players that have ever made a purchase) or the more restrictive one (players that have made a purchase in the last purchase churn definition -50 in the case of AoI- days).

Analogously to the resurrected players defined in section 3.3, *purchase resurrected players* can be defined. In this work however, the window of no spending considered is exactly that of purchase churn definition (50 days), thus making all false purchase churners purchase resurrected players once they start purchasing again.

Figure 3.4 (b) shows the fraction of real purchase churners (PUs that never purchase again) for increasing lengths of purchase churn definition using all historic data and considering all PUs. A purchase churn definition window of 50 days accounts for around 85% successful identification.

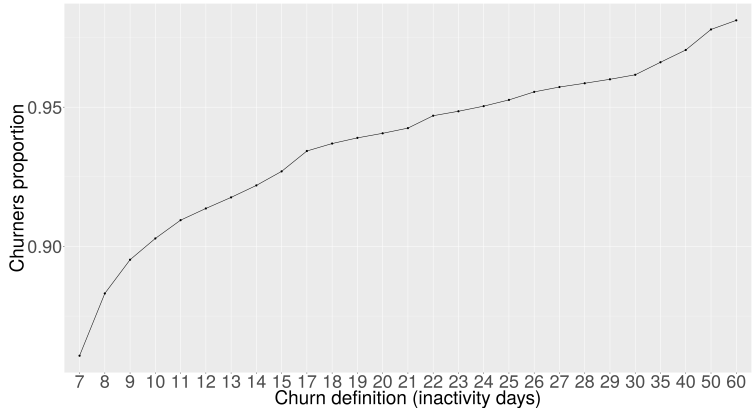


(a)

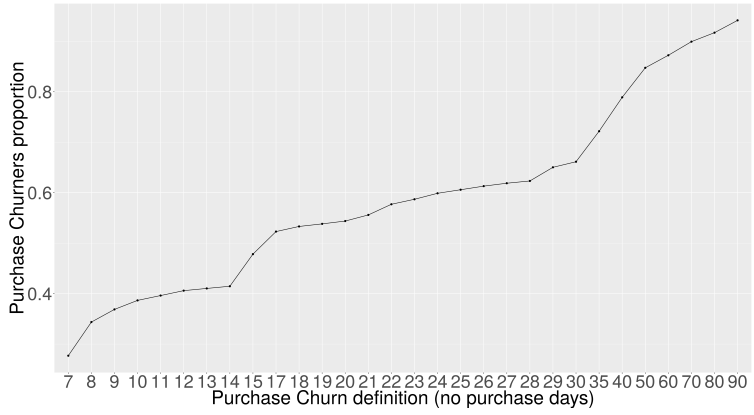


(b)

FIGURE 3.3: Percentages of missed sales (a) and false churners (b) during the first two months of data for VIP players (blue) and all paying users (PUs, red). Dashed lines indicate churn definitions that make these percentages fall below 1% and 5% respectively, yielding 9 days (for VIP players) and 13 days (for PUs). Elaboration using data from AoI. The image has previously appeared in [132].



(a)



(b)

FIGURE 3.4: Fraction of all real login (a) and purchase (b) real churners for different churn (a) and purchase churn (b) definitions using the complete AoI dataset. I.e, fraction of all users that never logged into the game again after a given number of consecutive days without login (a); and fraction of PUs who never purchased again after a given number of consecutive day with no spending (b). Elaboration using data from AoI.

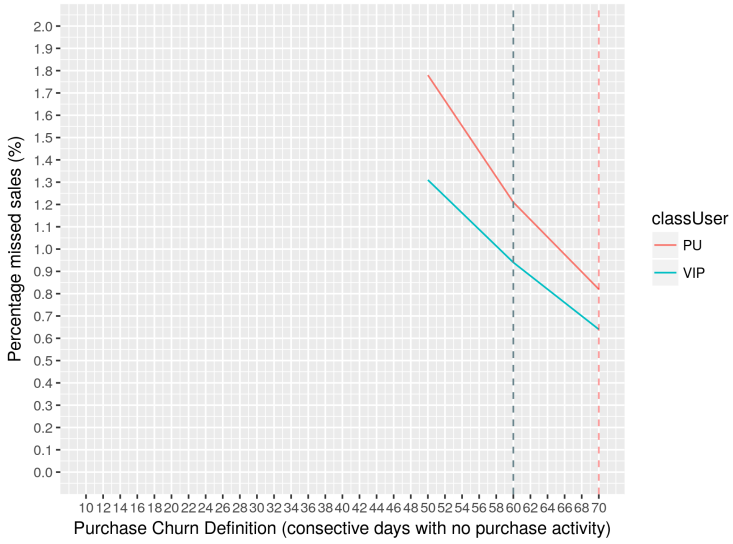


FIGURE 3.5: Percentage of missed sales for VIP players (blue) and all paying users (PUs, red). Dashed lines indicate churn definitions that make these percentages fall below 1%, yielding 60 days (for VIP players) and 70 days (for PUs). Elaboration using data from AoI.

4

TIME SERIES, CHURN AND CONVERSION

If you insist on learning something from every statistical analysis, you will tend to learn something stupid.

— Cassie Kozyrkov

Once the basic player categories defined in the previous chapter are established (active non-PUs, active PUs and inactive players), it is interesting to consider how their populations, and transitions between each group, evolve. The object of this chapter is to apply State Space Models (SSMs, see chapter 2 section 2.1) to understand and predict the evolution of the probabilities of churn (separately for PUs and non-PUs), non-PU to PU conversion (which is sometimes referred to in the literature simply as *player conversion*), and purchase churn. No distinction is made between different kinds of churners or spenders. Note that, importantly, players can transition from being PUs again to non-PUs, so the more restrictive PU definition (where a purchase in the previous 50 days is required) is in use. This means that every mention to PUs throughout this chapter refers in fact to what were described as genuine PUs in chapter 3 section 3.4.

In particular, an ARIMA (chapter 2 section 2.1.2) and an unobserved component or structural time series (chapter 2 section 2.1.3) model approach are considered. In both cases exogenous explanatory variables are also taken into account. The combination of an stochastic evolution model with a deterministic regression to explanatory variables[1] is an attempt to also understand the main drivers of the evolution of these series, as well as to have a measure of their impact. All results presented in this chapter are original contributions of this thesis, and were first published in [73].

Besides producing accurate daily forecasts of the transition probabilities, and estimations of the impact of in-game and external events in these transitions, following the methodology proposed in this chapter also allows for the detection of critical missing information, as will be described in some detail. It can therefore be used operationally, not only to improve resource allocation and game planning, but also as an automatic monitoring system,

1 The terms exogenous or explanatory variables, regressors and covariates will be used indistinctly throughout this chapter to refer to the dependent variables in this linear regression.

to detect for example, buggy releases or server failures. It can also be used to rate the success of any particular event or campaign as compared to the rest of events or campaigns of a similar type, as will be discussed. Furthermore, moving to more complex segmentation landscapes and applying, following this approach with give a measure of how different in-game and external events affect different kinds of players (in regards to skills, experience, type or frequency of play, purchasing behaviour...) differently. This could be of invaluable help when defining game planning strategies suited specifically to the targeted segment in mind.

Similar approaches could be used outside the realm of video games, as the methodology is suited to study any system where there are users that can be classified into different groups, and where understanding and/or predicting transitions between different groups is of interest. The exact same segmentation used here can be used for online platforms where purchases are possible. It could also be applied, for example, to the study on how events drive differently the interest (or lack thereof) of members of online communities that only consume content versus those that also generate it, and understand how the former may become part of the latter (and vice versa).

The use of this approach does not need to be restricted to online groups or interactions, although the automatic collection of data in these setups make them particularly good candidates. It could be also used, for example, to model the participation in extracurricular activities in a school.

Time series approaches have been used extensively to tackle all sorts of predictions regarding social and economic processes, from demand [226, 252] or prices [2, 61, 204], to crime [53] or epidemics [302, 303]. Examples can also be found outside the realm of human related time series, with goals as varied as predicting weather [273], water quality [92], or population growth [30, 84, 276]. There is however not much time series related literature in regards to video games. Online traffic generated by online first person shooter games was studied from a time series point of view in [65]. The total daily sales and playtime in AoI and Grand Sphere (another title developed by Silicon Studios) were predicted using different time series techniques in [131], where the effect of in-game events was also taken into account. Players were clustered according to their time series response to in-game events using unsupervised learning in [239].

In section 4.1 the time series to be studied will be introduced. Section 4.2 is devoted to the explanatory variables that will be used in the modelling, while section 4.3 describes in detail the methodology that was followed.

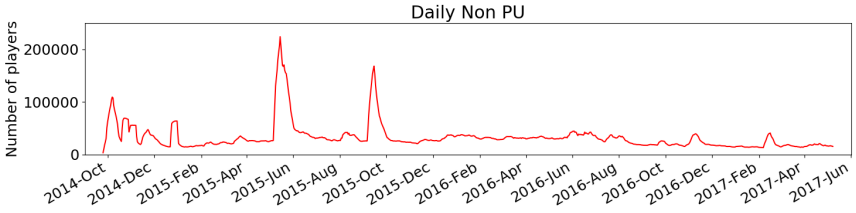
The results are presented and discussed in section 4.4. The chapter ends with a summary and conclusions in section 4.5.

4.1 POPULATIONS AND TRANSITIONS

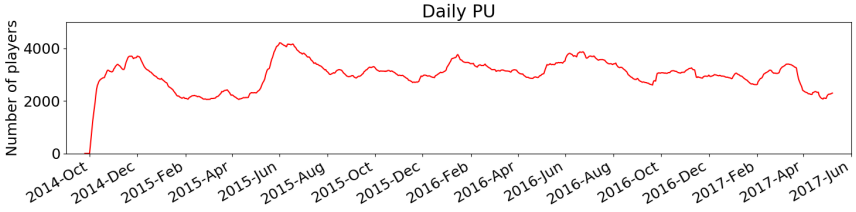
Figure 4.1 depicts the daily evolution of the populations of each of the basic segments considered: non-PU (a), (genuine) PU (b) and churned players (c). Note that these are not necessarily made up by players logging in on that day (in fact, for the churned players one this is an impossibility). For any given day, the sum of the three time series yields the total number of accounts that have logged into the game at least once in its entire history. Being this so, while the non-PU and PU time series will experiment fluctuations (and only go to zero when the game disappears), the churned players one will almost inevitably always grow.

Taking therefore into account the definitions for each segment given in chapter 3, the daily churned or inactive players population time series (figure 4.1 (c)) gives for each day, the number of players that have not logged into the game in the previous 9 days. The number of users each day that have played in the last 9 days, and have also made a purchase in the previous 50 days make up the daily (genuine) PU population (figure 4.1 (b)). Finally, the daily non-PU population is made up by players who logged into the game in the previous 9 days, but have not made a purchase in the last 50 days (figure 4.1 (a)).

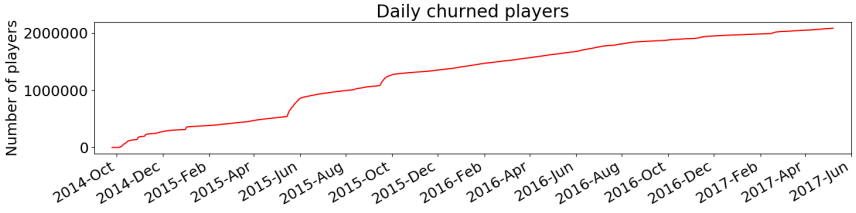
In figure 4.2 the time series plots making up the matrix of transitioning number of players can be found. All transition time series plots are compiled in this figure to provide the whole landscape of transitions at one glance. Each column is then also plotted in a separate figure for clarity, in figures 4.3 (transitions to non-PU), 4.4 (transitions to PU) and 4.5 (transitions to churned). The first row is concerned with players that were non-PUs the previous day, the middle row with players that were PUs and the bottom row with inactive players. The top row plots therefore depict the number of non-PUs that remain non-PUs (a), become PUs (b), or churn (c). That is, plot (a) (non-PUs to non-PUs) of the figure therefore shows, for each day, players that have logged in in the previous 9 days and have not made a purchase in the previous 50. Plot (b) (non-PUs to PUs) shows the number of players that purchase on that day but had not for the previous 50 days, and plot (c) (non-PUs to churned) is that of players that, without haven't made a purchase in the previous 50 days, logged in for the last time 10 days ago. Similarly, plots in the middle row correspond to players that purchased for



(a)



(b)



(c)

FIGURE 4.1: Daily population of non PU (a), genuine PU (b) and dead player time series (c). Own elaboration using data from AoI.

the last time 51 days ago and have played in the previous 9 days (PU to non-PU, plot (d)); players that have made a purchase in the previous 50 days and connected at least once in the previous 9 (PU to PU, plot (e)); and players that, having made a purchase in the previous 51 days, connected 10 days ago for the last time (PU to dead, plot (f)). Finally, the last row of plots shows the number of players that, after remaining inactive for at least 10 days, logged in on that day but made no purchase (dead to non-PU, plot (g)); those that, after remaining inactive for at least 10 days, not only connected again but did at least one purchase (dead to PU, plot (h)); and those that have not connected at least in the 10 previous days and remain inactive (dead to dead, plot (i)).

With the information about daily populations and transitions the conversion rates can be easily computed by dividing the daily transitions between the population of the group of origin on the previous day. These are plotted in figures 4.6 (conversion rates to non-PU), 4.7 (conversion rates to PU), and 4.8 (churn rates). These can be understood as the daily probability of a given player to transition from one group to another (or remain in the same segment). This means the top row plots show the evolution of the daily probability of a non-PU: not purchasing or churning (thus remaining non-PU, figure 4.6 plot (a)); purchasing (thus becoming PU, figure 4.7 plot (a)); and of having been inactive for 10 consecutive days (thus churning, figure 4.8 plot (c)). Analogously, the middle row plots show the probability of a PU having not purchased in the last 51 days while still having played in the previous 9 (thus becoming non-PU, figure 4.6 plot (b)), of having purchased in the previous 50 days (thus remaining PU, figure 4.7 plot (b)) and of not having logged in in the last 10 days (thus churning, figure 4.8 plot (b)). Finally, the bottom row shows plots of the evolution of the daily probability of an already inactive player connecting again without making any purchase (figure 4.6 plot (c)), playing again and purchasing (figure 4.7 plot (c)) and remaining inactive (figure 4.8 plot (c)). Note the addition of the left and middle bottom series (inactive to PU or non-PU) is nothing but the false churner probability. Although of interest in some contexts where studios pick the strategy of trying to reengage long lost players, in this work the *back to life* transitions were not modelled. They involve a small amount of players and are anyway of less interest than the rest.

Figure 4.9 shows the daily time series of new non-PUs (a) and new PUs (b). The former is the number of players logging in for the first time each day, and not making any purchase on that first day of play. The latter the number of players connecting to the game who had never played before,

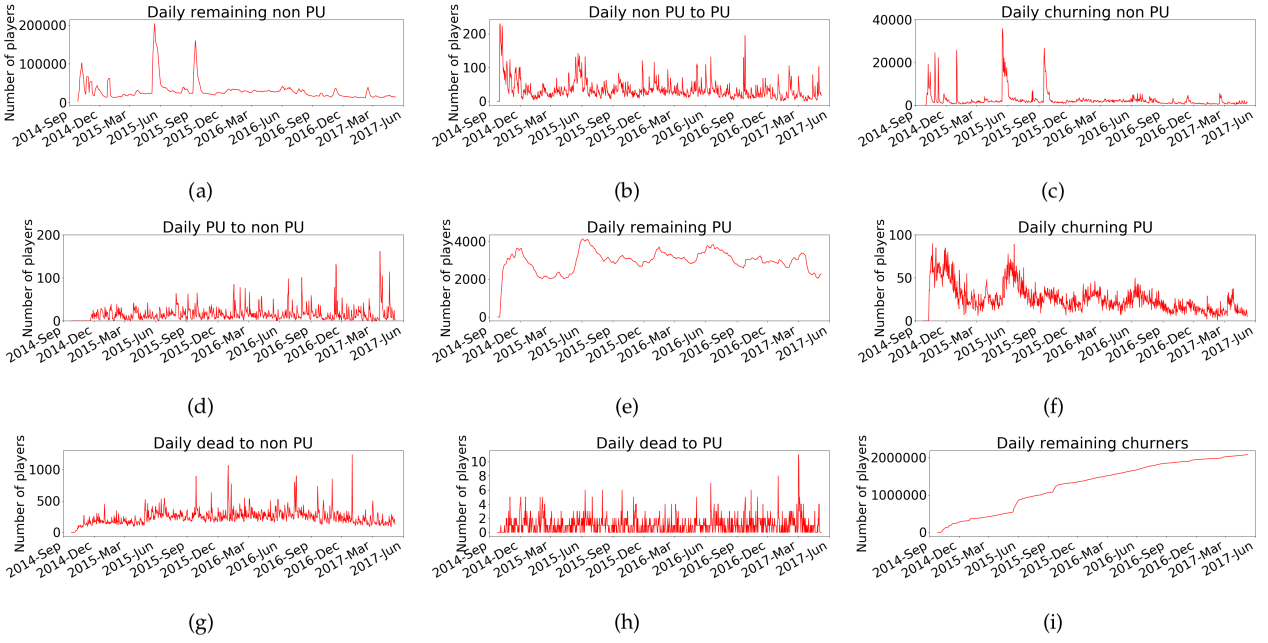
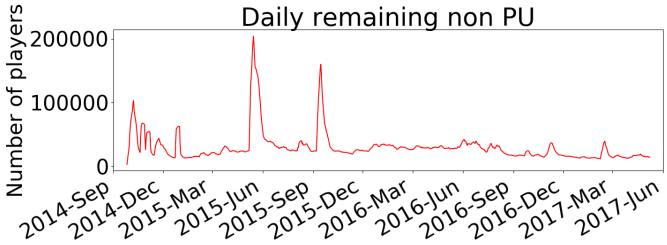
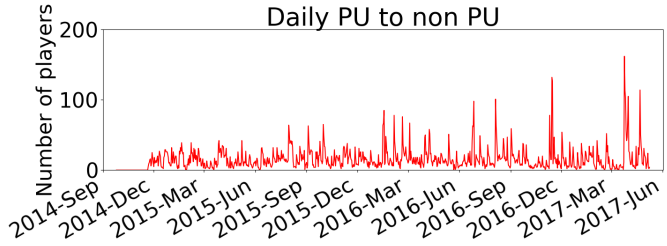


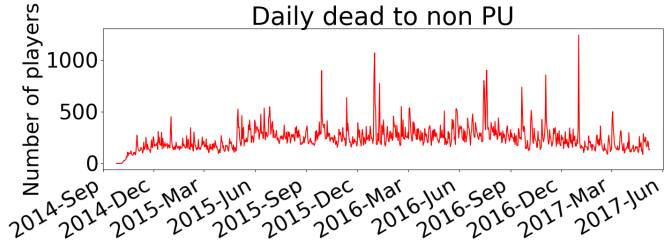
FIGURE 4.2: Matrix of daily transitions, i.e., number of daily players in the transitions: (a) non-PU→non-PU, (b) non-PU→PU, (c) non-PU→dead, (d) PU→non-PU, (e) PU→PU, (f) PU→dead, (g) dead→non-PU, (h) dead→PU, and (i) dead→dead. In all cases PU are in fact genuine PU. Own elaboration using data from AoI. The images have previously appeared in [73] (in black and white).



(a)

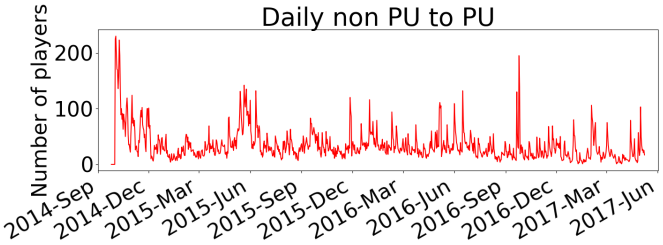


(b)

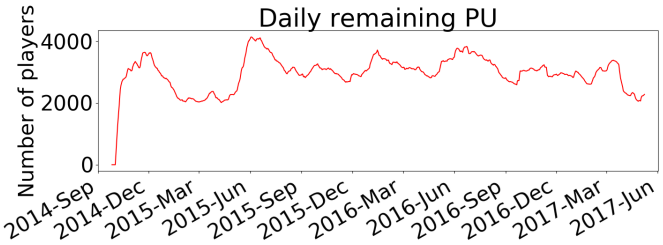


(c)

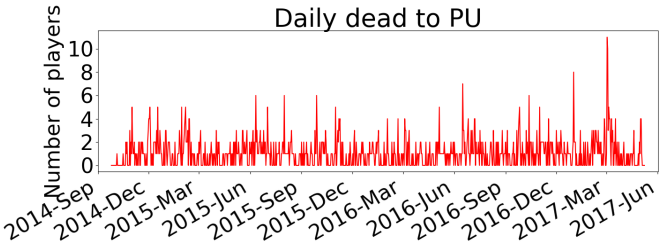
FIGURE 4.3: Daily transitions to non-PU, i.e., number of daily players in the transitions: (a) non-PU→non-PU, (b) PU→non-PU, and (c) dead→non-PU. PU here means genuine PU. Own elaboration using data from AoI. The images have previously appeared in [73] (in black and white).



(a)

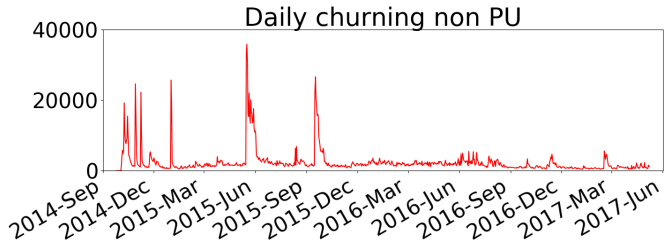


(b)

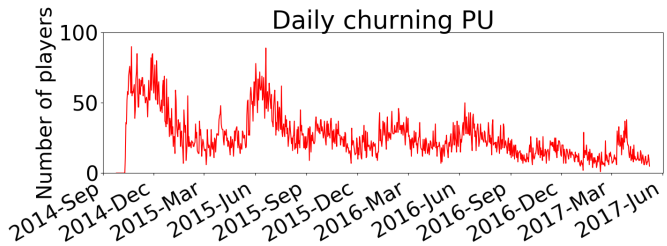


(c)

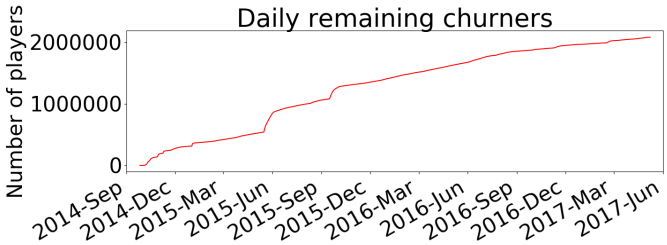
FIGURE 4.4: Daily transitions to PU, i.e., number of daily players in the transitions: (a) non-PU→PU, (b) PU→PU, and (c) dead→PU. PU here means genuine PU. Own elaboration using data from AoI. The images have previously appeared in [73] (in black and white).



(a)

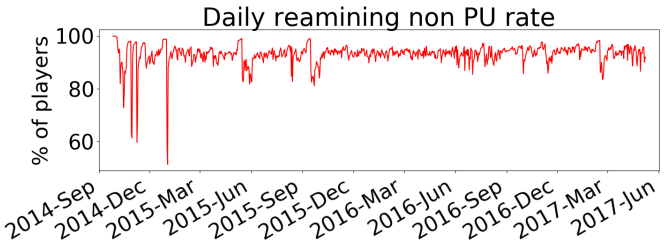


(b)

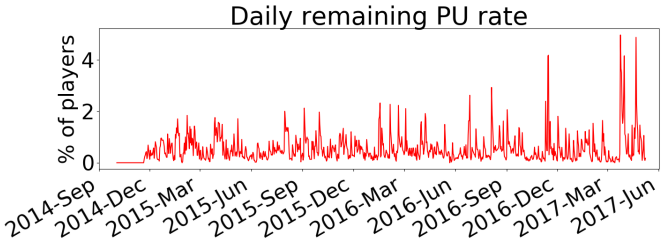


(c)

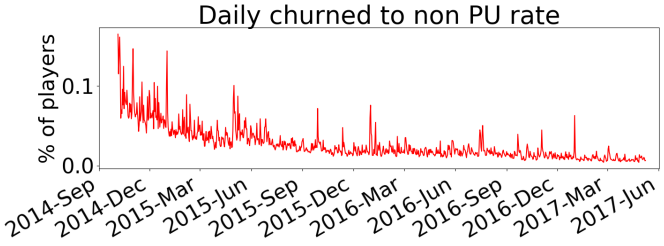
FIGURE 4.5: Daily transitions to dead (churned), i.e., number of daily players in the transitions: (a) non-PU→dead, (b) PU→dead, and (c) dead→dead. PU here means genuine PU. Own elaboration using data from AoI. The images have previously appeared in [73] (in black and white).



(a)

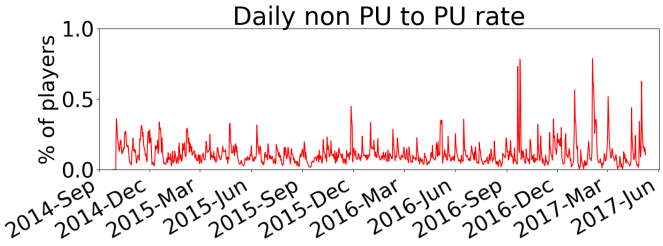


(b)

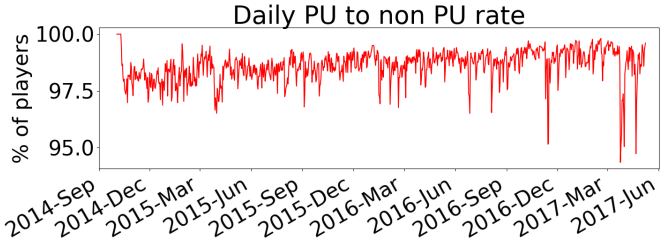


(c)

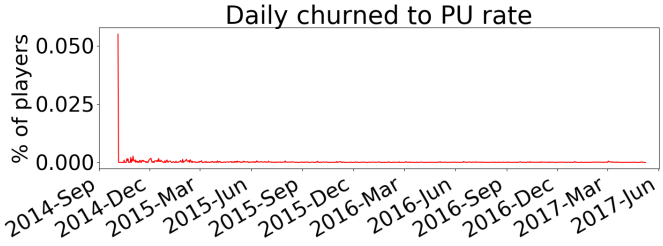
FIGURE 4.6: Daily transition rates to non-PU, i.e., number of daily players in the transitions: (a) non-PU→non-PU, (b) PU→non-PU, and (c) dead→non-PU. PU here means genuine PU. Own elaboration using data from AoI.



(a)

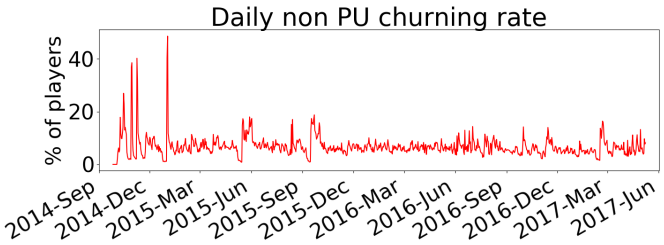


(b)

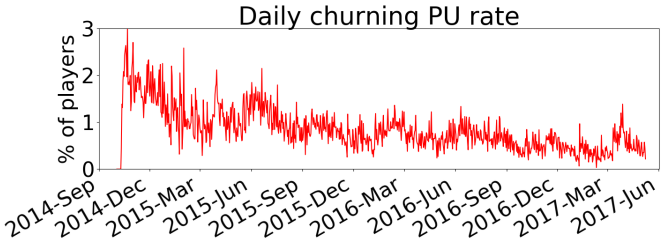


(c)

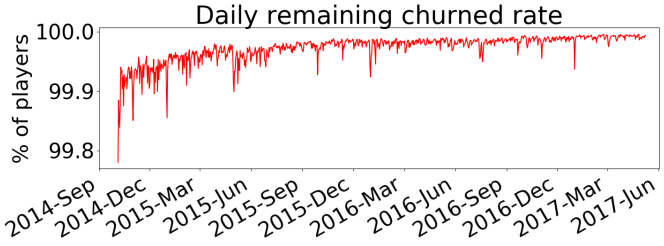
FIGURE 4.7: Daily transition rates to PU, i.e., number of daily players in the transitions: (a) non-PU→PU, (b) PU→PU, and (c) dead→PU. PU here means genuine PU. Own elaboration using data from AoI.



(a)



(b)



(c)

FIGURE 4.8: Daily transition rates to dead (churned), i.e., number of daily players in the transitions: (a) non-PU→dead, (b) PU→dead, and (c) dead→dead. PU here means genuine PU. Own elaboration using data from AoI.

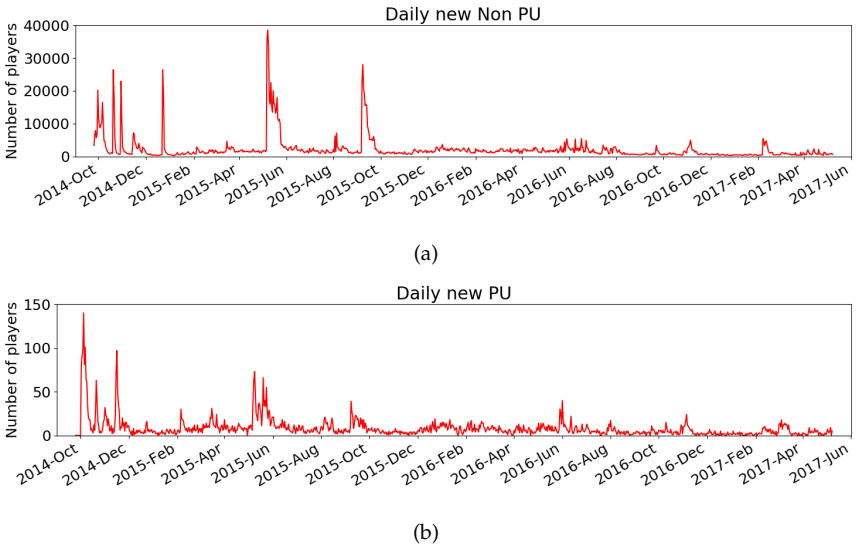


FIGURE 4.9: Daily new non PU (top) and daily new PU (bottom) time series. Own elaboration using data from AoI.

and making at least one purchase within that day. Note the non-PU series is two orders of magnitude larger than the PU one. The sum of both was also modelled in this exercise. As will be discussed, not only is it interesting in itself, as new user acquisition is expensive and critical to game health, but it is essential in the marketing campaign detection (and indirectly to the promotion campaign identification, as these should have no noticeable effect in this series), which was one of the main goals of this study, as will soon be clear. As was to be expected, both the new daily PU and non-PU series show high correlation: people are drawn to try the game, and then some of them decide to make a purchase on that same day (while most do not). The spikes and periods of much higher than usual values should correspond with the most intense and successful marketing campaigns.

4.2 EXPLANATORY VARIABLES

Calendar effects, holidays and in-game planning information is used to build covariates or explanatory regressors. As will be described in more detail in section 4.3, the modelling procedure is also used to unveil marketing

and promotion campaigns that took place, as well as other unknown events with a significant impact in the transition probabilities. The modelling thus, besides providing daily predictions for the transition rates, can be used to classify in-game events and other campaigns, depending on whether they impact or not, and with which sign and intensity, the different conversion rates.

Regarding marketing (aimed at new user acquisition) and promotion (aimed at generating player conversion and increasing spending among PUs) campaigns, there was no prior information available except for the fact that they had existed and been frequent, with significant impact.

With the in-game and external information readily available the following covariates were built:

1. *Day of week*: Intends to capture the effect of each day of the week. A variable per day of the week is used, which is 1 for days corresponding to that day of the week and 0 otherwise.

2. *Other calendar effects*: Intend to capture the effect of some very particular days of the month or year. First day of month, last day of month, first day of year, last day of year were built. Similarly as for day of the week variables, the first day of the month regressor is 1 if the day is first of the month and zero otherwise, and analogously for the rest of calendar effects.

3. *Holidays*: Two variables were built to capture separately the effect of Japanese school holidays and national holidays. These variables are also 1 when the day belongs to the corresponding category and 0 otherwise. This implies the model will estimate the same effect for all national holidays, that will be different of that of school holidays, but where all the latter will also have the same impact.

4. *In-game events*: The events are grouped according to their event type and event scale (described in chapter 1 section 1.2). For each of these type-scale categories two variables are built: one to capture the effect of having an event of that type-scale going on (1 when there is such an event going on, 0 otherwise), and another to capture the additional effect events typically have when they begin (1 if it is the first day of such an event, 0 otherwise).

5. *Number of in-game events*: In an attempt to somehow capture non linearities arising from the interaction of having more than an event going on at the same type, two additional covariates are built: one

which is the number of events going on on any given day, the other the number of different events starting on any given day.

On top of these, as will be described in section 4.3.3, additional variables are built in the process to try an reflect the effects of information which is missing. These are referred to as *interventions*, and are exogenous variables defined *ad hoc* in an attempt to uncover missing information, while making the residuals more normal and/or reducing their variance.

In the case of the new user acquisition campaigns and promotions, as has already been mentioned, although no dates or descriptions were available, it was known that they have existed often, and that many were deemed to be very successful. Some even have impacts visible in the series with the naked eye. It would be therefore a bad modelling practice not to try to account for them. The detection, and even assessment of the relative success of these campaigns, was actually one of the main motivations of this exercise. These will fall in one of three categories:

1. *Marketing interventions*: These should have a large positive effect in the new user model, as they are basically new user acquisition campaigns. They will typically also have a positive effect in the churn probabilities (delayed by the churn definition, i.e. 9 days), particularly in non-PU churn probability, as the campaign will get people interested in trying the game that will not however find it interesting enough to continue playing (and more so if they are not engaged enough in their initial try to spend money in the game). They can also possibly have a (limited) positive effect in the conversion to PU rate, if the campaign reaches people that are already playing and somehow reengages them, or if they are accompanied by promotions to try and get the new players spending. In this case they could have too a positive effect in the purchase churn probability after 50 days. Comparing the impacts in the new user series (how good the campaign in making people try the game were) and in the churn series (that indicates how good they were at targeting the right audience), gives a way of evaluating the success of these campaigns.

2. *Promotion interventions*: These should have a large positive impact in the conversion to PU series, as they should capture the effect of promotion campaigns in the game, targeted at people already playing. They will thus also typically have an effect in purchase churn with a 50 day delay. If the promotion does not work out well, they could even have a limited effect on churn (imagine for example

players being spammed with notifications that disengage them). They should not have any impact in the new users series, as these are promotions inside the game and only visible to players. Whenever there was a simultaneous combination of marketing and promotion campaign, this will be captured by a marketing type intervention (that are allowed to have effect in conversion too as was described above). The difference in impact in the conversion to PU and purchase churn series can be used to detect campaigns that were particularly good at generating long term conversion and/or encouraging players that were already PUs to purchase.

3. *Unknown interventions*: Large effects in any of the series that do not fit the promotion or marketing possibilities. These would include, for example, server failures or buggy releases (with positive impact in churn), changes in game dynamics (that reengage or disengage users), new content releases...

4.3 METHODOLOGY

One of the main goals of this exercise was to acquire qualitative knowledge about the system dynamics and processes at play. For this, human expertise to understand and interpret the results is needed. The guiding idea then, is not that of defining a fully automatic way of carrying out this process. Having however a well defined methodology, and a fixed procedure, will limit and guide the human intervention needed, and will guarantee that different humans will arrive at very close modelling results. It will also make easier the transition to more complex segmentation landscapes, should there be interest in doing so.

Particularly critical is this needed human expertise, as will soon be obvious, when defining interventions. Once the model selection stage is completed, and all missing information is uncovered, the resulting system could be then left running in production. It could be designed to generate alarms when unusual effects (probably due to missing information) appear, and to automatically produce forecasts (taking in-game planning into account). Such a system would still require periodical maintenance by someone knowledgeable with the models and the data, but human intervention needed would be at that point much more limited. The initial stage described in this paper is, although to some extent automatic, time consuming, and requires a great deal of expert human intervention. A lot is to be learned in this process nonetheless.

For each of the five time series, two different SSM models were defined: an ARIMA and an UC, both with the covariates described in 4.2. The following five steps were then followed: (1) model selection; (2) selection of significant exogenous variables; (3) intervention definition; (4) model selection revisited; and (5) forecasting and verification. Each of these will be described in detail in the following subsections.

4.3.1 *Model selection*

The first step is to decide which specific ARIMA and UC configurations will be used for each time series as a starting point (this choice will be revisited after selecting explanatory variables, see section 4.3.4). To begin with, plots of the series and their differences (regular, weekly and monthly) are examined, together with their correlograms (autocorrelation Function or ACF and partial autocorrelation function or PACF). This includes examining how all values are distributed (mean, variance, outliers...). All available historic data is considered. In the new user series case, in addition to the original, the log-transformed time series is also considered, and actually selected as it has a more stable variance. This is equivalent to using a multiplicative model instead of an additive one. Using inspection of these series and ADF stationarity tests (see section 2.2.1 in chapter 2) of the time series, it is decided to model in the ARIMA case all series with a regular difference. Although there is an obvious weekly seasonality in most of the time series, taking a weekly difference yields higher variance and anticorrelation. Original series (plus log-transformed in the new series case), regular difference, ACF and PACF for all the modelled series are shown in figures 4.10 (new users), 4.11 (non-PU to PU conversion rate), 4.12 (PU churn), 4.13 (non-PU churn) and 4.14 (purchase churn).

The next decision, which is made independently for each time series examined, is fixing the starting date for the period used in the training. This is set by inspection, in a to some extent arbitrary way, by the human expert analysing the series, in order to leave out the initial highly volatile period every game experiences right after launch. Unless the object under study is in itself behaviour or game dynamics immediately after launch, this period is always better left out, as it is a unique phenomenon that will not be repeated and will *confuse* (introduce noise in) the models.

In the case under study, as the period available is relatively long, including this period does not have a significant impact in the modelling except for the intervention definition. As the residuals will typically have very

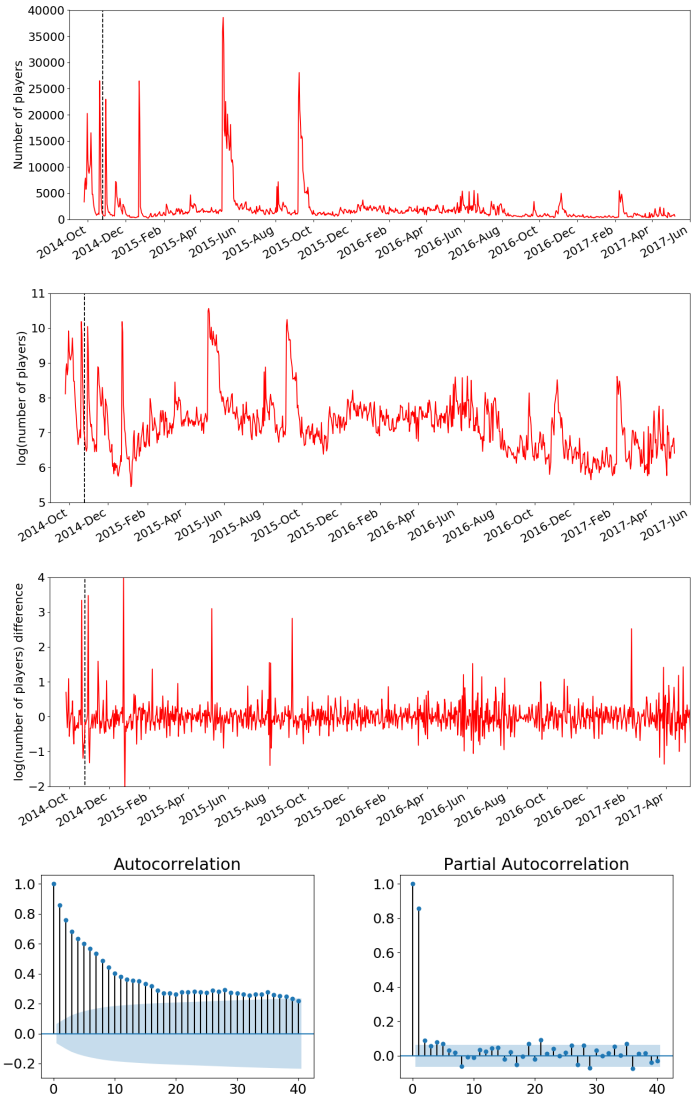


FIGURE 4.10: New users original series (top), log-transformed (second row), difference of the log-transformed (third row), ACF (bottom left) and PACF (bottom right). Start of the training period is marked with a dashed line and corresponds to October 10 2014. Non-significant correlation area (95% confidence) is shaded in ACF and PACF. Own elaboration using data from AoI. The images have previously appeared in [73] (in black and white).

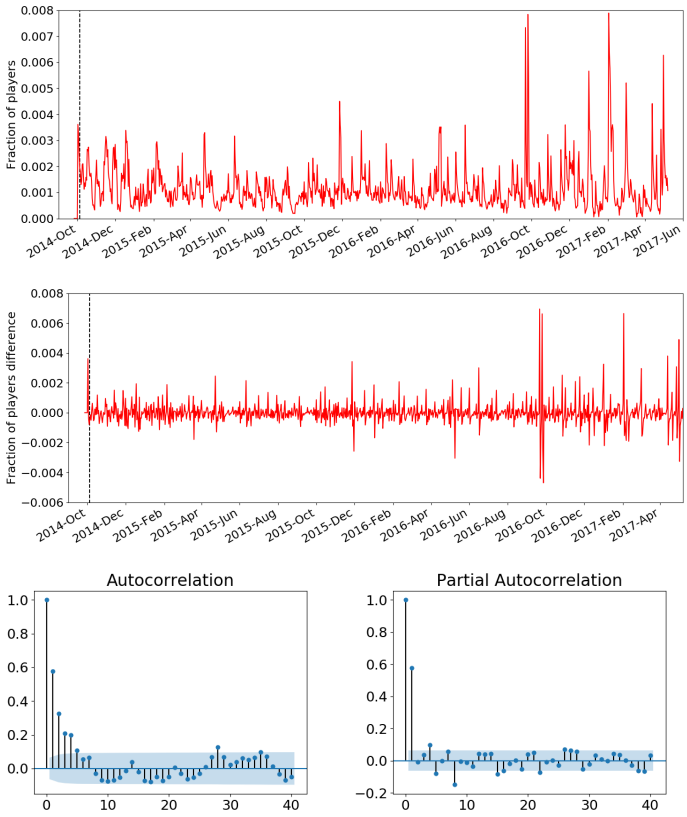


FIGURE 4.11: Non-PU to Pu conversion original series (top), its regular difference (middle), ACF (bottom left) and PACF (bottom right). Start of the training period is marked with a dashed line and corresponds to October 5 2014. Non-significant correlation area (95% confidence) is shaded in ACF and PACF. Own elaboration using data from AoI. The images have previously appeared in [73] (in black and white).

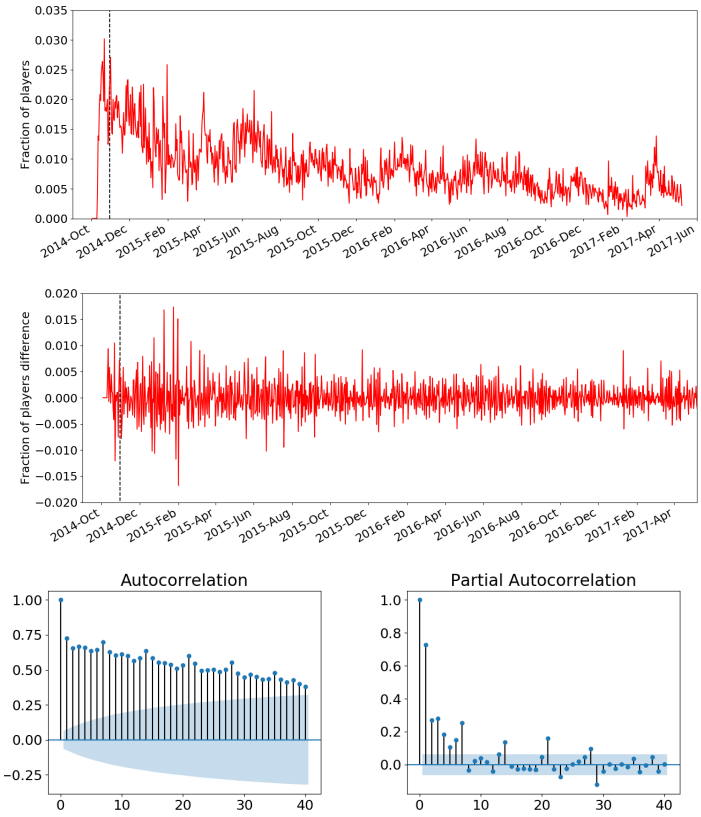


FIGURE 4.12: Churning PU original series (top), its regular difference (middle), ACF (bottom left) and PACF (bottom right). Start of the training period is marked with a dashed line and corresponds to October 31 2014. Non-significant correlation area (95% confidence) is shaded in ACF and PACF. Own elaboration using data from AoI. The images have previously appeared in [73] (in black and white).

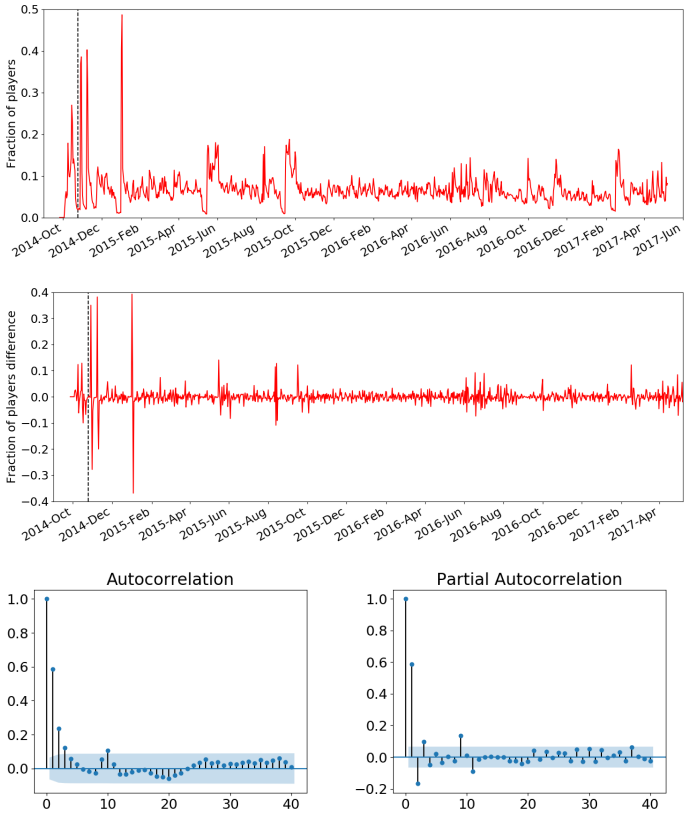


FIGURE 4.13: Churning non PU original series (top), its regular difference (middle), ACF (bottom left) and PACF (bottom right). Start of the training period is marked with a dashed line and corresponds to October 31 2014. Non-significant correlation area (95% confidence) is shaded in ACF and PACF. Own elaboration using data from AoI. The images have previously appeared in [73] (in black and white).

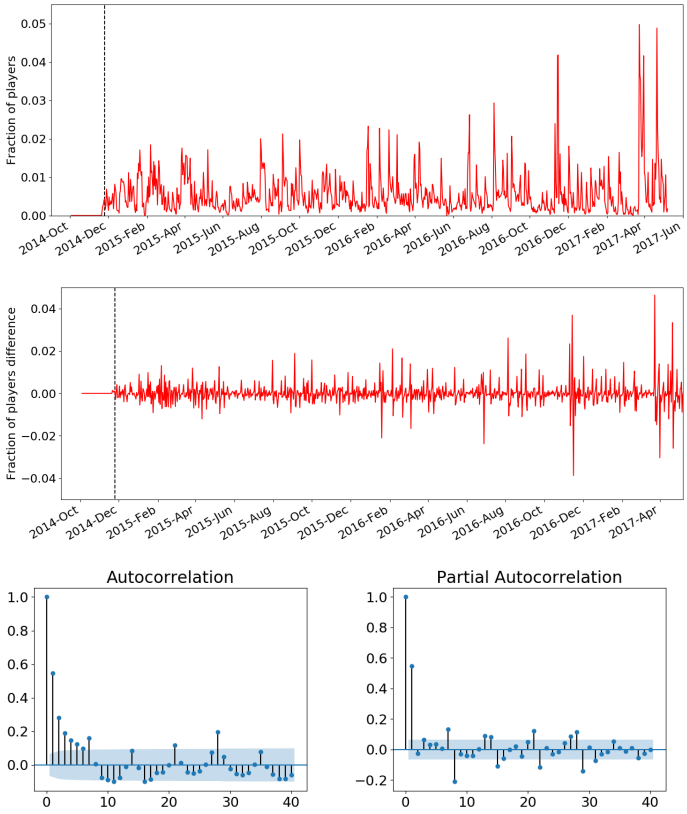


FIGURE 4.14: Purchase churn original series (top), its regular difference (middle), ACF (bottom left) and PACF (bottom right). Start of the training period is marked with a dashed line and corresponds to November 5 2014. Non-significant correlation area (95% confidence) is shaded in ACF and PACF. Own elaboration using data from AoI. The images have previously appeared in [73] (in black and white).

large deviations in this period, it will mean extra work as more (meaningless) interventions need to be defined in the process (see section 4.3.3 to understand the intervention definition process).

Setting different starting dates for the different series is not only heuristically justified. The series themselves start at different dates. In some cases by definition, as for the churn and purchase churn series (that have a delay of the churn and purchase churn, i.e 9 and 50 days, respectively). Others due to the launch circumstances: in this case for example, in-app purchases were not available for some time immediately after the launch. The selected starting dates for the different time series are: October 10 for the new user series, October 5 for the PU conversion rate, October 31 for both PU and non-PU churn and November 5 for purchase churn. These are shown as dashed lines in the time series plots in figures in figures 4.10 to 4.14.

Now, for the ARIMA the degrees of the AR and MA polynomials, both regular and seasonal (weekly in this case), still need to be defined (up to order 5 will be considered in all cases). For the UC, the use or not of weekly and monthly seasonalities, of a cyclic term, and the type of level-trend is also to be determined. As was described in section 2.1.3 (chapter 2), the types of level-trend that will be considered are: no trend, fixed intercept (deterministic constant), local level (random walk), fixed slope (deterministic trend), local level with deterministic trend (random walk with drift), local linear trend and smooth trend (integrated random walk).

Model specification is typically done relying on human expertise: carefully inspecting the series, trying out a few plausible model definitions, and comparing these, is the usual process, which relies on the knowledge of how different ARMA and UC realisations look like. Another approach consists on exploring the parameter space defined and using some criterion to automatically decide upon the results. This work uses a combined approach, which intends to make up for the possibility of lack of very strong knowledge of ARIMA and UC processes, and to make the process to some extent automatic, while still allowing the expert human to assess the results and make a choice within a limited range of options.

A selection of five candidates is done by *brute force*, exploring all possible combinations of the subset of the model space described above, and selecting the five best according to the AIC (see section 2.2.4 in chapter 2). For these five preferred candidates, the BIC (see section 2.2.5 in chapter 2) and HQIC (see section 2.2.6 in chapter 2) are examined, together with residual variance, independence (as tested with Ljung-Box, see section 2.2.2 in chapter 2) and normality (as tested with Jarque-Bera, see section 2.2.3

in chapter 2). Parameter significance is also taken into account (see section 2.2.7 in chapter 2), with only parameters with p-values under 0.1 considered as significant. While the brute exploration might be too extensive (very rarely is it justified, for example, to use ARMA polynomials of order of more than 2; and the type of level-trends explored could be reduced, for example, in its stochasticity), as the different tries can be run in parallel and are not computationally expensive, it is still deemed a reasonable approach.

The selected models for each series will be described in section 4.4, but it is interesting to note here some general findings. No model with seasonal ARIMA was selected in any case, while weekly seasonality was favoured in all the UC cases. The day of the week covariates described in section 4.2 were therefore only used in the ARIMA case. The absence of a seasonal ARMA is particularly interesting given that (i) most series do have a clear weekly seasonality (as shown by significant correlations for lags 7 and multiples in the correlograms, in particular for PU churn and purchase churn, see figures 4.12 and 4.14); (ii) the UC model does always prefer to use this periodicity; and (iii) in the ARIMA case, as will be seen, day of week inputs were estimated as significant. Monthly seasonality was rejected in the UC case for all series. Local level type was always considered the best level-trend choice. While other options with more degrees of freedom naturally succeeded in further decreasing residual variance, these always yielded significantly worse information scores.

Note that, if we consider the ARIMA or UC model to capture the seasonality and inertia of the time series, the remaining residuals can be used to learn about the natural underlying variability and stochasticity. Even more so after the impact of external variables or covariates -which are modelled as deterministic effects- is filtered out as described in next section (section 4.3.2).

4.3.2 *Exogenous variable selection*

The different explanatory exogenous variables available for the modelling were described in section 4.2. These include the interventions, whose definition process is described in some detail in section 4.3.3. The process described below to select the covariates for each time series will be repeated twice. In the *first round*, for each time series, after having a model including variables with parameters estimated to be significant, interventions are defined. The *second round* is used to try again all the previous regressors plus all the interventions defined (for all conversion series) in the previous

round. This will fix the final selection of covariates among all the available (predefined and interventions).

All exogenous variables that have a Z-score with associated p-value less or equal than 0.1 are considered significant, and kept as explanatory regressors for that time series. In the first round, they are tried grouped according to their type in the following order: day of the week, calendar effects, holidays, in-game events and number of in-game events. They are grouped and tried in bundles of ten for groups with a larger than that number of inputs. This is done to avoid (or at least limit) convergence problems when trying to estimate with too many variables at the same time. In the case of churn and purchase churn series, the explanatory variables are used with a lag equivalent to the relevant churn definition. With the interventions the procedure is very similar. After completing the process, additional estimations are run trying to separately include one by one all the covariates that have not been estimated significant, as there is a chance that they are with the final configuration.

In the case of additive models, the non-transformed series is modelled with regressors. This is the case for all of this exercise's series with the exception of the new user series. The parameters estimated for each of the exogenous variables in additive daily models can be understood as the variation in the modelled series due to each unit of change in the covariate series. When the series are log-transformed, as is the case of the new users time series, the resulting model with regressors is multiplicative, in which the parameters corresponding to explanatory time series represent elasticities, i.e., they are the proportionality constant relating fractional variations in the modelled series an the covariate.

4.3.3 *Interventions*

The procedure to define interventions for a given time series is as follows. First, the predefined exogenous variable first round selection described in the previous section is carried out. It is then assumed the largest deviations left in the residuals probably correspond to missing information. The day with the largest deviation in absolute value is selected and the series (original, estimated, sum of regressor effects and residuals) examined in detail (as has already been mentioned, this is the part of the process for which human expertise is more critical). The exact shape of the intervention to be used is then determined. This could be, for example, an additional degree of freedom for the day of the deviation, if it seems like a one

day effect. Typically, campaigns would have run through several days. This means that if the deviation under study is positive, followed by an additional negative deviation in the residuals some days later, and we are considering one of the ARIMA models (which in all cases in this work have a regular difference), it will be assumed these points mark the beginning and the end of the campaign. The campaign intervention will then be defined with that shape, i.e., 1 between the first and last day and 0 elsewhere (plus possibly an additional effect for the first day). The model is then re-estimated, and the goodness of the new fit as compared to the previous version assessed, using a similar approach to the initial model selection phase described in section 4.3.1: parameter significance, information criteria and variance, normality and independence of residuals are examined. It is then decided if the intervention is kept or not. If it is kept, the residuals of the new model are now considered, and their largest deviation now tackled in the same way.

The types of interventions that can be defined, depending on which time series they impact and with which sign, were described in section 4.2, and can be grouped into marketing interventions corresponding to new user acquisition campaigns, promotion interventions and unknown interventions (all the rest).

All defined interventions were tried with the models of all time series, regardless of which time series they had been originally detected for. Most promotion interventions, for example, are expected to be defined when modelling conversion to PU. Even if these should have no effect in the new user series, the new user model is also estimated with them. This will allow us to discriminate between different types of interventions, have a more consistent model across transitions, and reduce the risk of missing the effect of some of the discovered explanations in other transitions where their impact is less obvious.

To facilitate their detection and interpretation, a fixed order in the way the time series were modelled was followed during the first round: first, the new user series in order to uncover the marketing campaigns; secondly conversion to PU to detect promotions; third, non-PU churn, looking for additional new user acquisition campaigns; followed by non-PU churn; with purchase churn modelled the very last (and where additional promotions could be detected). Once defined they are treated in the exact same way as the rest of exogenous variable groups.

In the second round, the process described in section 4.3.2 is repeated, now considering the three groups of interventions additional exogenous

variable groups in equal terms to the rest. Now the order of the groups is altered (to reflect the hierarchy of expected impacts): for the new user and login churn series marketing interventions are tried before in-game events; and for conversion to PU and purchase churn, promotion interventions are tried before marketing ones. Unknown interventions are tried in last place for all series.

As will be described in section 4.4, following this methodology with the ARIMA models yields a plausible scenario of new user acquisition campaigns and promotions. This was however not the case when dealing with UC models. This makes sense in that their structure, and particularly their larger number of noise terms, allows for a bigger flexibility and capability for absorbing large shocks in the original series. It was in the end decided to run the intervention definition phase only with the ARIMA models, and use this same intervention landscape with the UC models.

This is obviously the more time-consuming and expertise-dependent part of the model definition phase. This process could be carried out in a fully automatic manner, adding only single day variables to make residuals more normal, and repeating until the model quality is degraded instead of improved. This would still make the models more theoretically sound and consistent, even in the absence of human expertise or time for a close inspection. In essence, the underlying reasoning is the same: very large deviations, or deviations from Gaussian behaviour in the noise, are due to unaccounted for effects that should be filtered out. In this way, unusual behaviour would also be filtered, but not accounted for. In contrast, the procedure described above accomplishes this and looks for plausible explanations for the deviations. This process is still necessary to prevent underfitting and to make the models used formally valid, but must be done carefully in order to avoid overfitting. While the model will not be able to predict similar effects in the future, it will be much better equipped to predict accurately in the absence of these.

4.3.4 *Model selection revisited*

After there is a fixed selection of exogenous variables for each model (including interventions), the same procedure described in section 4.3.1 was repeated now with covariates, in order to ensure that the best possible ARIMA or UC definition was being used. This resulted in practically no changes in the model definition, with only a couple of corrections in the ARMA orders for two of the series. In particular, the main findings of the

initial model selection (all ARIMA models with a regular difference and no seasonal ARMA, all UC models with weekly an no monthly periodicity and with local level) remained unaltered.

4.3.5 *Forecasting and verification*

Until this step, all available information (historic data) has been considered (with the period prior to the starting date of each series excluded in the model selection, parameter estimation and intervention definition). In this last step, training and forecasting steps are carried out, leaving out the end of the series to test the predictive accuracy such models would have had. For more details on how the forecasts are performed in each of the components (deterministic regression filter and ARIMA or UC), check chapter 2 section 2.1.

Daily forecasts for the following month were run, with data until the last day of the previous month used in the training, and this for all of 2016 and what was available of 2017. The intention was to replicate a possible production setup, in which predictions are run to allocate resources for the following month. Note however, that the model definition and variable selection were done using all data available, which would not have been the case in a real life operational situation (although the forecasts produced still relied only in the data that would have been available by then).

As all interventions are defined locally, and only once the period for which they are defined is available for training, these are never included in the forecasting. This means very large errors in the predictions were expected, which would be the real life situation if forecasts were being run without taking into account promotions and marketing campaigns. In a reasonable operational scenario, the information about future campaigns would be available, and included in the predictions using a similar approach as has been used for in-game events. Of course, unexpected impacts such as server failures would still produce large deviations from the expected behaviour, which is precisely why such a system could be very useful in production, as it would detect events that are having an unaccounted for impact in the conversion rates.

Monthly MAE and RMSE (see chapter 2 section 2.9) were used to compare model accuracy between ARIMA and UC, and between different months for each model (there is no baseline to assess overall model accuracy).

4.4 TRANSITION PROBABILITY MODELLING RESULTS

Figure 4.10 shows the daily number of users logging into the game for the first time (top), its log transformation (second plot from the top), and the regular difference of the latter (third plot from the top). In all three plots October 10, 2014 is marked with a dashed line, indicating the beginning of the period used for parameter estimation. The ARIMA selected in this case was (2, 1, 1), and the UC model had cycle periodicity included (besides the weekly seasonality and local level common to all the series). In both cases the log-transformed series was selected for this exercise, i.e., a multiplicative model in the exogenous variables. However, all references to errors in what follows will be regarding the non-transformed series. The last row of 4.10 shows its ACF and PACF (excluding the period before the starting date), where the shaded area represents non-significant correlation values with 95% confidence.

Conversion to PU is shown in figure 4.11, where the top plot depicts the daily conversion rate, the middle plot its regular difference, with ACF and PACF at the bottom. The dashed line marks here again the beginning of the training period (October 5, 2015), which is too the period used in the computation of the ACF and PACF. The ARIMA model selected in this case was (0, 1, 3), and the UC model used no additional cycle component.

Similarly, 4.12 and 4.13 represent PU and non-PU churn respectively (daily probability series, regular difference and correlograms), both with starting date for the training October 31, 2014, marked here too with a dashed line. Both UC models did not use cycle component, and the ARIMA selection yielded (0, 1, 2) for PU and (1, 1, 2) for non-PU.

Finally, similar plots for purchase churn are shown in 4.14, with starting date November 25, 2014, ARIMA (0, 1, 3) and no long term cycle used in the UC model. The models selected for all time series are summarised in table 4.1.

As has already been described, the intervention definition was carried out using the ARIMA models only. However, after this was done, the second round of exogenous variable selection yielded the exact same covariate selection for the UC and ARIMA models, for all of the time series. Further more, the parameters estimated were also very similar, differing typically by less than 10% and in only a handful of cases by more than 20%. An illustrative non-comprehensive selection of the values estimated for a few

		ARIMA		UC		
Time Series	Start date	Regular	Seasonal	Trend type	Seasonal	Cycle
New users (log)	10-10-2014	(2,1,1)	✗	local level	weekly	✓
Conversion to PU	5-10-2014	(0,1,3)	✗	local level	weekly	✗
PU churn	31-10-2014	(0,1,2)	✗	local level	weekly	✗
Non-PU churn	31-10-2014	(1,1,2)	✗	local level	weekly	✗
Purchase churn	25-11-2014	(0,1,3)	✗	local level	weekly	✗

TABLE 4.1: Summary of the models selected for each time series. The start date is the first date from which all historic data available is used for the model parameter estimation. The regualr and sesonal ARIMA polynomials are listed for the ARIMA model, and the type of trend, the seasonality, and whether a cycle term was included for the unobserved components model.

parameters in the different series is shown in tables 4.2 (ARIMA models) and 4.3 (local level models).

Taking into account the different nature of additive and multiplicative models, interpretation of parameters is as follows. Consider, for example, the effect of national holidays. In the new user series (multiplicative), both models are estimating an increase of approximately 5% (4.89% for ARIMA, 5.15% for UC) in the number of new players on national holiday days. Non-PU churn (additive), also becomes more likely on these days, with the probability of churning for these players being around 0.003 (0.0033 for ARIMA, 0.0028 for UC) higher (absolute as opposed to relative increase). There is also an increment of around 0.0002 in the conversion probability, and of 0.0007 (UC) to 0.0008 (ARIMA) in the purchase churn rate on national holidays (both additive models), but no measurable effect in PU churn (this covariate is not deemed to be significant for this series after the modelling procedure described in section 4.3). School holidays, on the contrary, were not estimated significant for any of the series, which could be indicating a low presence of school age players relative to working players in the game.

The positive effect of national holidays in most transition probabilities reflects the fact that users have more time to play (for the first time and also afterwards), and thus of spending money, but also of getting bored of the game. Very engaged players, who make a significant fraction of all PUs, are far from being at risk of losing interest in the game by playing a few hours more, thus the lack of impact of holidays in PU churn.

Parameter	New users	Conversion to PU	non-PU churn	PU churn	Purchase churn
National holidays	4.89×10^{-2}	1.61×10^{-4}	3.30×10^{-3}	-	7.67×10^{-4}
Battle event (start)	-	2.16×10^{-4}	-	-	-2.77×10^{-4}
Gacha 4	-	1.03×10^{-3}	-	-	1.61×10^{-2}
Raid event (start)	-	-1.07×10^{-3}	-	3.80×10^{-3}	3.23×10^{-3}
Unknown 2017/02/09	-	-1.33×10^{-3}	1.51×10^{-2}	1.37×10^{-4}	1.88×10^{-2}
Marketing 2015/02/05-07	5.71×10^{-3}	9.97×10^{-4}	3.45×10^{-2}	4.38×10^{-3}	1.23×10^{-3}
Marketing 2015/03/16-25	6.06×10^{-3}	-	3.70×10^{-2}	5.14×10^{-3}	-
Marketing 2015/05/25-31	6.07×10^{-3}	-	2.74×10^{-3}	-	-
Marketing 2016/09/21-22	6.25×10^{-3}	1.90×10^{-4}	2.40×10^{-2}	-	-
Marketing 2017/02/07-09	2.06×10^{-3}	-	-1.26×10^{-3}	-	-
Promotion 2015/03/19	-	1.53×10^{-3}	-	-	4.60×10^{-3}
Promotion 2015/04/23-24	-	1.74×10^{-3}	-	-	2.30×10^{-3}
Promotion 2016/09/21-23	-	3.38×10^{-3}	-	-	-

TABLE 4.2: ARIMA estimates for a selection of parameters for the different series. Parameter signs and values are an indication of the type and strength (respectively) of the effect the modelled event is estimated to have had. For the new users series (log-transformed) they can be understood as elasticities, for the other time series, additive effects to the transition probabilities. A - indicates the covariate was not found to have a significant effect on that time series.

Regarding the weekly structure estimated (explicitly through the use of inputs in the ARIMA case, and as a term of the UC modelling), it is similar for all series and is, as was the case of bank holidays, related to how the amount of players logging in and their average playtime varies along the week. Users play much less at the beginning of the week, they start to play more around Wednesday, and this increases as the week progresses, peaking on the weekends, when players typically have more free time.

From the calendar effects that were tried out, only the beginning of the month and the end of the year were estimated to have a significant impact in some of the series. Both were estimated as significant in all series except non-PU churn. The beginning of the month has a positive impact on PU conversion probably due to the availability of new salaries for spending, and this too, generates a positive effect in both purchase churn (with 50 days delay), and in PU churn (driven by the increment on PU login in that day). It also has a small positive effect in the new user series, probably related to the same phenomenon of individuals exploring new games in

Parameter	New users	Conversion to PU	non-PU churn	PU churn	Purchase churn
National holidays	5.17×10^{-2}	1.47×10^{-4}	2.83×10^{-3}	-	7.10×10^{-4}
Battle event (start)	-	2.13×10^{-4}	-	-	-2.96×10^{-4}
Gacha 4	-	1.01×10^{-3}	-	-	1.59×10^{-2}
Raid event (start)	-	-1.06×10^{-3}	-	3.88×10^{-3}	2.50×10^{-3}
Unknown 2017/02/09	-	-1.25×10^{-3}	1.57×10^{-2}	1.31×10^{-4}	1.79×10^{-2}
Marketing 2015/02/05-07	5.78×10^{-3}	9.81×10^{-4}	3.56×10^{-2}	4.17×10^{-3}	2.73×10^{-3}
Marketing 2015/03/16-25	6.45×10^{-3}	-	3.95×10^{-2}	5.17×10^{-3}	-
Marketing 2015/05/25-31	5.49×10^{-3}	-	1.34×10^{-3}	-	-
Marketing 2016/09/21-22	6.12×10^{-3}	1.95×10^{-4}	2.29×10^{-2}	-	-
Marketing 2017/02/07-09	2.72×10^{-3}	-	-1.88×10^{-3}	-	-
Promotion 2015/03/19	-	1.46×10^{-3}	-	-	3.44×10^{-3}
Promotion 2015/04/23-24	-	1.74×10^{-3}	-	-	2.11×10^{-3}
Promotion 2016/09/21-23	-	3.46×10^{-3}	-	-	-

TABLE 4.3: Local level estimates for a selection of parameters for the different series. Parameter signs and values are an indication of the type and strength (respectively) of the effect the modelled event is estimated to have had. For the new users series (log-transformed) they can be understood as elasticities, for the other time series, additive effects to the transition probabilities. A - indicates the covariate was not found to have a significant effect on that time series.

which to spend their newly earned money. It has however no apparent effect on non-PU behaviour, reinforcing the believe its effect is mainly related to purchasing rather than playing behaviour. The last day of the year sees an increment in all probabilities except in non-PU churn, where it has no effect, and in the new user series, that decrease on that day. That less people try the game on a day so packed with other activities and festivities makes sense. The increase in the other series is likely related to special events in the game that enough players (in this case PUs) find tempting enough.

The hypothesis that events are more a driver of purchases than of logins is further reinforced by the estimation of in-game event effects. These have, as was expected, no effect in new users, but also no impact in non-PU churn, and very limited (only a couple of event types with very low value) in PU churn. Events can have positive or negative impact both in PU conversion and in purchase churn, so there is an interesting landscape of events either encouraging or discouraging spending with respect to a no event scenario, as understood by the models. Most event types have positive effect on both series (see for example *Gacha* in tables 4.2 and 4.3). As they encourage spending, conversion to PU is increased, but some of this effect is lost

once the event is over, driving too purchase churn (with 50 days delay). There are however event types with more interesting effects. *Raid event* type, for example, has a negative effect in conversion to PU while increasing purchase churn and even slightly PU churn: they seem like events that both discourage spending and are disliked by PUs. On the opposite side of the spectrum, *Battle event* type encourages conversion to PU and also reduces purchase churn, appearing to really drive spending in both PU or non-PU (or at least they generate long lasting conversion).

Regarding the detection on missing information on marketing and promotion campaigns, as has already been noted, those detected using the ARIMA setup were used for both SSM approaches. The additional noise terms in the UC model enable it to better capture sudden changes in level in the series without interventions, so the attempt to uncover these with this approach ended up with many less campaigns detected, and degraded notably the quality of the forecasts.

The marketing and promotion campaign scenario uncovered with the ARIMA modelling was plausible and consistent, both in terms of the series impacted and the sign of the effect (as will be discussed below), as of the in-game planning they suggest. According to the interventions used, for the first months of life of the game there would have been a profusion of campaigns of both types, which is typically done by studios right after the launch of a new a game.

After that, there would have been a much higher intensity of marketing campaigns during summer months than the rest of the year, which also makes sense. As has already been discussed, there tends to be an increase in all measurable play indicators when players have more free time. New user acquisition campaigns are typically expensive, so more of them are planned during periods when more people are likely to be on holidays, increasing the probability of both them actually noticing the campaign, and on acting on it. Outside the summer months, these campaigns would have taken place approximately every second month.

In the case of the picture painted by the promotion interventions, these campaigns were intense at the beginning, and then, after 2015 started, became shorter and sparser. Towards the end of 2016, and specially throughout 2017, they would have however experienced an increase in length, frequency and intensity. This also makes sense in that it is at this time when other indicators (in-game sales, DAUs, total playtime...) suggest the game was transitioning from a growing phase to a contraction one, as they were decreasing during the second half of 2016. This intensified campaign planning

would have had the goal of preventing this. They also appear to have been successful, as they do appear to bring the amount of purchases back to the level they had before their decline began.

All marketing interventions have a large positive impact in the new user series. To assess which campaigns were more successful, one would have to look, not only at the size of this effect, but also at the impact they had in the other series. Most campaigns would also have a clear effect in the churn series, particularly in non-PU churn, as many of the people tempted to try the game will not find it interesting enough to continue playing. Smaller or undetectable impact in churn would indicate campaigns which were better at targeting the right audience, i.e., at bringing players into the game that remained in the game. Take for example the four first marketing campaigns listed in tables 4.2 and 4.3. They all have an effect of increasing the amount of new users by around 6%. All of them impact too non-PU churn, and of them two make PU churn too increase. The May 2015 campaign has no effect in PU churn and the lowest of the four in non-PU churn, so can probably be deemed as the more successful of the four campaigns. On the opposite side of the spectrum, the March 2015 one appears to have been the worst, as it has the highest impact in non-PU churn and also increases PU churn.

Some of the marketing campaigns seem to also been linked to something going on inside the game, as a limited amount of these interventions were estimated to have a positive impact in conversion to PU and/or a negative effect in churn. These could be, for example, new user acquisition campaigns related to new content or items in the game. It could also be related to promotions running parallel to the marketing campaign in an attempt to make new users make their first purchase. It could also signal some campaigns were so well designed, that they too motivated people that had played already, and were reached by it, to play and spend. This is the case, for example, of the February 2015 and the September 2016 campaigns, that have already been considered in the previous paragraph. The former even has an impact in purchase churn (with the corresponding delay). The final marketing intervention included for February 2017, illustrates what would have been a particularly good campaign of this type: while it increases new users less significantly than the other four campaigns discussed, it actually has a negative impact in churn, suggesting it was linked to something inside the game that engaged new and old players alike.

Marketing interventions were also tried with one and two days of delay in the conversion to PU series, in an attempt to capture the possible effect

of new players deciding to make their first purchases a couple of days in. This effect appears to not be significant, indicating that most users either become PU straight away, never, or in later stages of the game.

Turning now to promotion interventions, all of them have effect in conversion to PU, and most of them in purchase churn (with 50 days delay). None were estimated as significant for any other series (as was expected). Those rare ones with no noticeable effect in purchase churn, are those that managed to make long lasting conversion and/or significantly encourage spending among PUs. Some of them were probably also linked with new attractive content available for purchase. From the three examples presented in tables 4.2 and 4.3, the one for September 2016 would belong to this category. Of the remaining two, the April 2015 one would have been more successful, as has a lower impact in purchase churn, and a similar one in conversion to PU.

Very few interventions not falling into the marketing or promotion categories were defined. Some of them appear to point at incidents clearly disengaging players. These could be, for example, changes in content disliked by many, buggy releases, or server failures (or other technical issues damaging the play experience). Others are less difficult to understand, and were needed mainly for the purchase churn series. This series is the one with a much larger part of the dynamics remaining unexplained after using the rest of covariates. This could indicate that what drives players to stop spending has more to do with the outside of the game and personal circumstances than with in-game planning than the rest of conversions. Or simply that it is more unpredictable for different reasons.

As was explained in subsection 4.3.5, one month long daily forecasts were ran using data for training until the last day of the previous month, starting for January 2016. No intervention covariates were used for the prediction, as they were assumed to be only uncovered once the discrepancies with the model (and hence the large residuals) would have been observed. The mean values and standard deviations of MAE and RMSE for both ARIMA and local level are shown in tables 4.4 and 4.5 respectively, while figures 4.15 (for MAE) and 4.16 (for RMSE) plot the monthly mean values for both models, with ARIMA represented by a solid line and the UC local level by a dashed-dotted line. The goal of all these are to compare model performance between ARIMA and UC, and between different months. There is no available baseline to which to compare the performance of these models and judge their general goodness, as they are the first attempt in this context to provide daily forecasts. When used in production, they

would become these baseline models to which to compare any proposed improvements in the time series modelling of these transitions.

Time Series	ARIMA Mean	ARIMA SD	Local Level Mean	Local Level SD
New users	440.72	270.73	483.20	282.71
Conversion to PU	4.6×10^{-4}	2.5×10^{-4}	5.1×10^{-4}	2.5×10^{-4}
PU churn	1.7×10^{-3}	5.8×10^{-4}	1.8×10^{-3}	5.8×10^{-4}
Non-PU churn	1.4×10^{-2}	7.0×10^{-3}	5.2×10^{-2}	6.8×10^{-3}
Purchase churn	3.3×10^{-3}	1.9×10^{-3}	3.3×10^{-3}	1.6×10^{-3}

TABLE 4.4: Monthly forecast MAE: mean and standard deviation (SD) for the ARIMA and local level models. Note new users measures number of users, while the rest probabilities.

Time Series	ARIMA Mean	ARIMA SD	Local Level Mean	Local Level SD
New users	634.64	461.44	677.91	463.86
Conversion to PU	7.1×10^{-4}	5.0×10^{-4}	8.2×10^{-4}	5.2×10^{-4}
PU churn	2.1×10^{-3}	7.0×10^{-4}	2.2×10^{-3}	6.8×10^{-4}
Non-PU churn	1.9×10^{-2}	9.4×10^{-3}	5.6×10^{-2}	8.9×10^{-3}
Purchase churn	5.1×10^{-3}	3.7×10^{-3}	4.9×10^{-3}	3.0×10^{-3}

TABLE 4.5: Monthly forecast RMSE: mean and standard deviation (SD) for the ARIMA and local level models. Note new users measures number of users, while the rest probabilities.

When comparing accuracy for different months, very large errors are associated to months for which interventions were defined with many high impact and/or long campaigns, and with effect thus not taken into account until the forecasted period becomes part of the training set. As was expected, the performance for different months is highly dependent in the amount and importance of missing information (that will later be uncovered and brought into the models through interventions).

The mean absolute error in users is between 400 and 500 users for both ARIMA and UC, for a series were there are typically several thousand new daily users. The variability of the MAE is similar in both models and large (as is the variability of the original time series), ranging from less than a 100 for January 2017 to over 1000 in February 2017. As stated above, large deviations are only found for months that appear to have had very intense new user acquisition marketing activity. It is therefore reasonable to assume the largest errors would be greatly reduced if adequate information campaign planing was included in the forecasts.

The MAE for the rest of the series can be interpreted directly in terms of probability. Size of errors scale as the typical magnitude of the series modelled, with the same order of magnitude for conversion, PU churn and purchase churn, and a larger one for non-PU churn. Errors are around one order of magnitude smaller than the real typical daily values of the original rate series, indicating the forecasts could be indeed accurate enough to prove useful. Even more so if the marketing and promotion campaigns were included from the beginning, as large deviations are consistently associated to this, as has already been stated.

The ARIMA models generally outperform the local level ones, but only in the case of non-PU in any significant manner. In this case, the local level model has systematically larger errors by around 0.05, in a series with daily values that only rarely exceed 0.1 churn rate. This implies that the UC model would be of very little practical use in this case. When looking for an explanation on why this happens only in this particular case, one notices that for this time series, the UC model selected has a much higher signal-noise ratio than the rest. The signal-noise ratio is the ratio between the local level and the residual or irregular variances, and is related to the number of effective lags that are being taken into account in the modelling: the higher its value is, the less lags are considered [139]. This in turn makes it more difficult for the model to capture the correct level of the series, and could account for the larger relative errors in this case.

Discussions on the similarities and differences of ARIMA and UC models can be found, for example, in [6] and [139]. They find both models perform very similarly, and so the much simpler form of the structural time series model would be preferred, as they are easier to interpret, and also need less human expertise in their correct definition. All examples they give however, are weekly or monthly. For daily time series, it is reasonable to expect a larger number of outliers, as there is less (to no) smoothing happening due to averaging or addition could explain the existence of more cases where the discrepancy between both models is large (such as for non-PU churn in this case), particularly for larger signal-noise ratios. It would be interesting to study how performance varies when considering different frequencies (daily, weekly, monthly and even yearly).

4.5 SUMMARY AND CONCLUSIONS

Time series SSMs provide a way of studying the probability of transition between different groups or segments of players. This gives insights on how

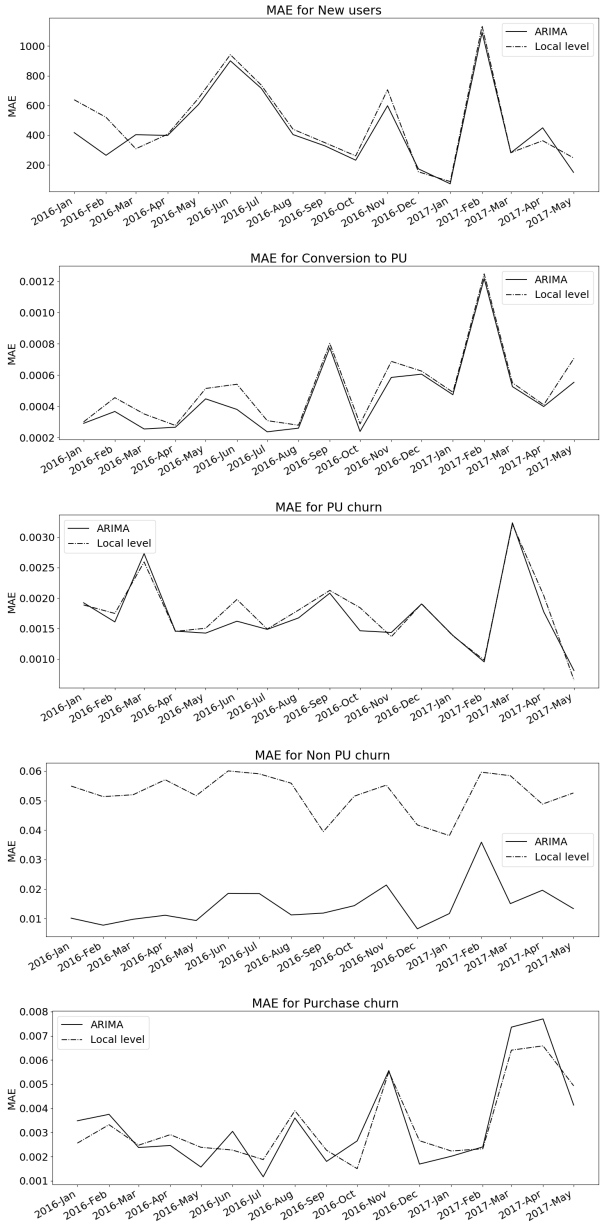


FIGURE 4.15: MAE for all successive monthly forecasts for new users (top), conversion to PU (second row), PU churn (third row), non PU churn (fourth row) and purchase churn (bottom). Own elaboration using data from AoI.

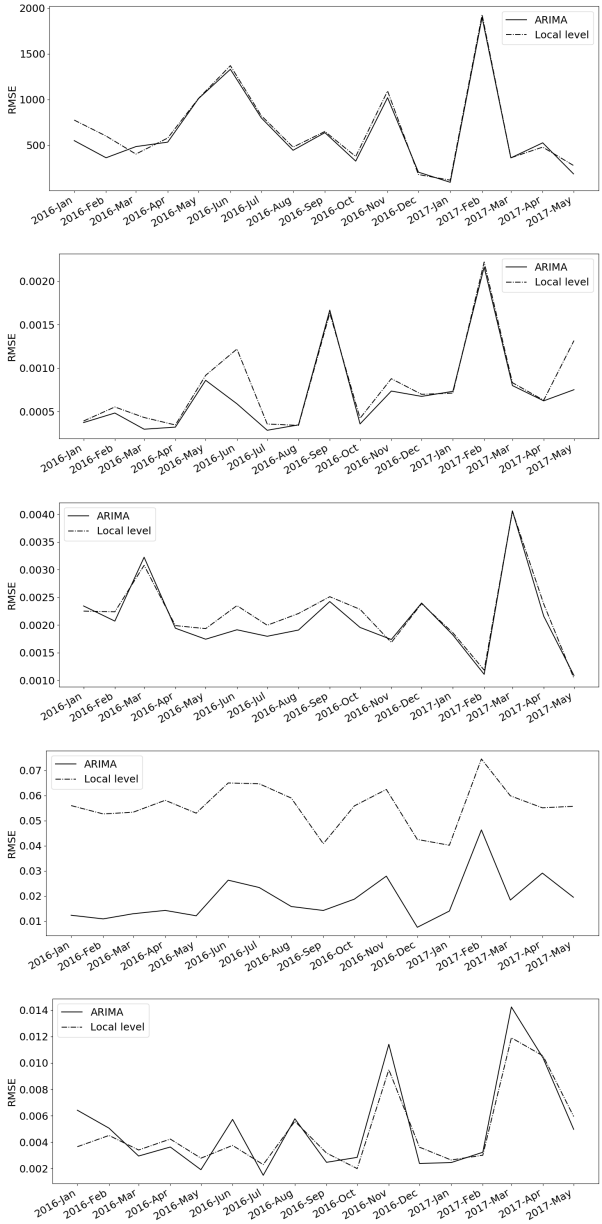


FIGURE 4.16: RMSE for all successive monthly forecasts for new users (top), conversion to PU (second row), PU churn (third row), non PU churn (fourth row) and purchase churn (bottom). Own elaboration using data from AoI. The images have previously appeared in [73].

different in-game planning strategies affect different kinds of players, and allows to classify, for example, certain in-game events and campaigns in terms of their relative success. They can also be used to generate predictions, taking into account not only in-game planning but also external events such as calendar effects. Last but not least, it is a powerful tool in detecting missing relevant information.

In this exercise, besides the time series of new users, PU and non-PU churn, conversion and purchase churn rates were modelled with two different SSM approaches: ARIMA and structural time series. The ARIMA model is statistically more complex, and requires substantially more human expertise in its correct identification. It was also however the only of the two alternatives that successfully yielded a plausible marketing and promotion campaign intervention landscape (information on these was not available, but they were known to have existed frequently and with significant impact). In everything else, from covariate selection, to parameter estimation, to resulting forecast quality, both models were remarkably very close. Forecast performance was slightly better for the ARIMA, except for non-PU churn were it was substantially better. However, if the focus was in throwing light on how the different events or campaigns affect the different segments, and on general game dynamics (rather than in detecting missing information or producing accurate daily forecasts), UC model would be selected for its simplicity and less intensive human intervention needs.

While the initial model definition phase is time consuming and highly dependent on human expertise, a methodology has been defined that should make this stage easier to complete, and would guarantee that different experts would arrive at similar or identical final model definitions and regressor selections. Model definition should then be revisited periodically for maintenance, but otherwise having such a system running in production would be practically cost free both in terms of human intervention and computational resources needed. Besides the forecasts running periodically every week or month for planning and resource allocation, the models could be actually re-estimated every day, which would allow for early detection of anything unusual going on, as well as an on-going assessment on how each particular event or campaign is doing as compared to the previous ones of the same type. This process could be fully automatic.

As was briefly discussed in the introduction of this chapter, this same approach can applied to more complex segmentations, and it would then throw light on how different events, campaigns, external events, etc. affect churn and conversion of different types of players differently. Additionally,

the conversion rates, predictions, and estimated effect of the explanatory variables can be used to build features for models concerned with individual player behaviour. Such models are used in the following chapters (chapters 5 and 6). The use of the conversion rates and their models, together with individual player predictions, in generating a well rounded understanding of churn and conversion will be further discussed in the conclusions of this thesis (chapter 10).

5

PREDICTING PLAYER ENGAGEMENT

Every individual matters. Every individual has a role to play. Every individual makes a difference.
— Jane Goodall

Player engagement is an abstract and complex concept, and its definition is thus elusive. This chapter focuses in understanding and predicting engagement, using the risk or probability of churning as a proxy to its inverse -*disengagement*. As opposed to the previous chapter (chapter 4), here the object of study are individual players. Predictions of the immediate risk of churn can be used to characterise each player. Additionally, predictions on how long each player is expected to remain active in the game, in terms of time since first login, accumulated playtime, and in-game progression, can further enrich the picture. These also help capture play intensity and ability of the player, which are, too, related to user engagement. Note that these are not at all AoI specific, which makes this approach easy to generalise to basically any sort of game.

Early detection of churn risk is crucial in all service commercial setups. Video games are no exception, as was already discussed in chapter 3 section 3.2. Having a good understanding on what is driving disengagement, and which players are in greater risk and when, can be used to target individual players in an attempt to re-engage them. This opens the door to personalization in game planing (and even game development). Certain content (items, levels, characters, actions...) for example, could be reserved for a certain degree of disengagement or point in the game in terms of churn risk. Furthermore, following methodology as the one that will be discussed in chapter 8, specific content particularly well suited to each player's preferences can be selected. Generally speaking, the more modelling results available at individual player level, the greater the opportunities to tailor the game to each player's taste. This is a win-win approach: it makes the game more enticing to the player, while increasing their lifetime within the game, thus normally positively impacting the studio's revenue.

This chapter focuses solely in play or login engagement, which is here considered to be the variable which in the end determines whether a player is active or inactive (categories as described in chapter 3). *Purchasing*

engagement will be defined analogously as that which resolves when a player is a genuine PU and when they are not. Everything related to purchasing behaviour (i.e., predicting which players will become PUs, when this will happen, and how much money they will spend) will be covered in the next chapter (chapter 6).

In section 5.1, the prediction of churn is treated as a binary classification problem, while section 5.2 introduces the use of survival analysis to that effect. The methodology used was first discussed in [28, 52, 167, 225] (before the studies leading to this thesis), and the results presented were first published in [224] (which summaries the former and presents new improved results). Section 5.3 analyses how taking into account the churner profiling described in chapter 3 section 3.3 can be used to enhance the understanding of this problem. The content of the latter section is another original contribution of this thesis (as second author) and was first published in [52]. Finally, the chapter ends with a summary and some concluding remarks in 5.4.

5.1 IMMINENT RISK OF CHURNING

The most simple approach to individual player churn prediction is to treat it as a binary classification problem. That is, to use decision trees or ensembles, support vector machines, neural networks, probit model... to assign each player to one of two categories: high or low churn risk. Video game churn had already been tackled with this type of algorithms in [135, 163, 166, 236, 238, 256]. Having each player categorised as in high or low risk of churning, is a first simple approach to understanding engagement. It gives a snapshot of the current situation, dividing users into engaged players (low churn risk) and those poorly engaged (in high risk of churning).

Certain classification algorithms, such as for example decision trees or forests (decision tree ensembles), will actually output a probability of churn when training for this classification problem. This further enriches the understanding of the situation, as it allows for a more complex (than binary) image in terms of this probability. It also enables to set the threshold deciding whether to assign a player to a high or a low risk churn group. This way, depending on the intended use of the predictions, and thus whether false positives or false negatives are of greater concern, a different value can be decided upon, arriving at a more suitable binary classification.

In particular, conditional inference ensembles (see chapter 2 section 2.3.2) are a good choice to provide probability predictions in a binary approach,

as they correct the tendency to overfitting and bias towards variables with more splits present in the more traditional random forest model.

While this is not the preferred approach to this problem in this thesis, the use of classification ensembles has been presented here, as section 5.3 will also evaluate the behaviour of conditional inference forest generated predictions, when deciding to incorporate or not the different types of churners in the training. See table 5.1 for the AUC scores (see chapter 2 section 2.3.2) of the different alternatives tested, and section 5.3 for a discussion.

5.2 PREDICTING REMAINING TIME, PLAYTIME AND LEVEL TO CHURN

Survival models can introduce more complexity in the understanding of engagement. They are particularly well suited to deal with censored data, being churn datasets censored in their very nature (at least until the game ends). They provide, not a prediction on the imminent risk of churning, but rather how long each player is going to remain active. Different variables can be used to measure this *time to churn*:

- *Lifetime*: Number of days since the user's first login into the game.
- *In-game progression*: Level achieved or experience points accumulated by the player.
- *Accumulated playtime*: Total number of hours played by the user.

This means every player is then characterised by multiple predictions: on how many days after starting to play, how much playtime, and at which level, they are going to completely loose interest in continue playing. The study of their disengagement in multiple variables allows for the ability, experience and play intensity of the player to be taken into account, and thus a more complete picture emerges. This methodology was proposed in [28, 52, 167, 224, 225], where conditional inference survival ensembles (see chapter 2 section 2.4.3) were shown to outperform other survival models such as the Cox regression (see chapter 2 section 2.4.1) or random survival forests (see chapter 2 section 2.4.2) in this particular task.

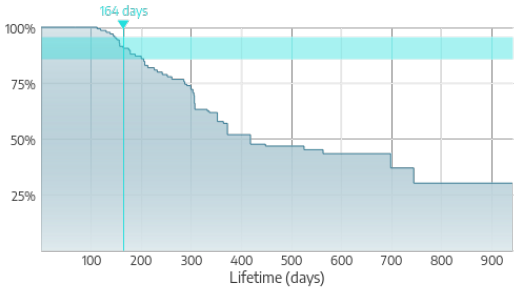
Conditional inference survival ensembles output a survival probability curve per player. In this case actually three: one in lifetime, one in playtime and one in in-game progression. Rather than a point prediction on when the event of interest (churn in this case) is going to take place, each player is in fact characterised by the probability of still being active at each time step

(with *time* being days after first log in, hours of playtime, or level). When a point prediction is needed, the median will be typically used. This means players are predicted to quit once their survival probability falls below 50%. Again, as in the conditional inference classification ensemble described in the previous section (section 5.1), this threshold can be tweaked depending on the intended use of the predictions. As compared to these, that provided a single probability of imminent churn per player, the survival ensembles yield, not only a probability of churn at the present moment (and that in terms of lifetime, playtime and level), but also the probability per player at any other (past and future) time. It obviously gives a more well rounded perspective, reason why is the method of choice in this chapter to study churn.

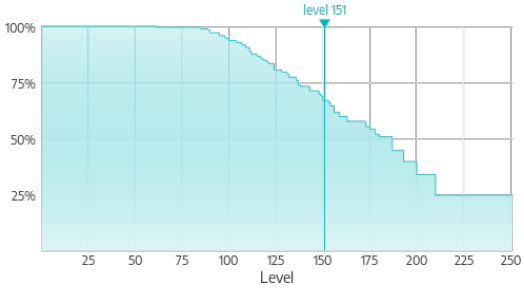
Figure 5.1 shows an example, for a player in particular, of this collection of three survival curves that characterises each user. Their probability of still being active in the game is plotted against lifetime (a), level (b) and playtime (c). The player is still active, with the current values on each variable marked with vertical lines in the plot. Note that all survival probabilities for the present of that player are well above 50%. Interestingly, however, while lifetime and playtime they are close to 100%, it is closer to 70% in terms of level.

Kaplan-Meier estimates are used as split criteria for the conditional inference survival ensembles (see chapter 2 section 2.4.3). They can be used to show graphically the actual survival information contained in the dataset, and hence are very useful to understand how disengagement has behaved until the present moment, and to compare how it typically evolves for different players. Figure 5.2 shows the Kaplan-Meier estimates of the survival probability as a function of days since first login (a) and level (b) for for non-PUs (purple), PUs (pink) and VIP players (green). As they progress in the game, the probability of players continuing to be engaged decreases. Note the acute differences in the engagement of different types of players: while well over 25% of the VIP players continue to play two years after they first logged into the game, for non-PUs, the same percentage of remaining active players is reached only a few days after first login.

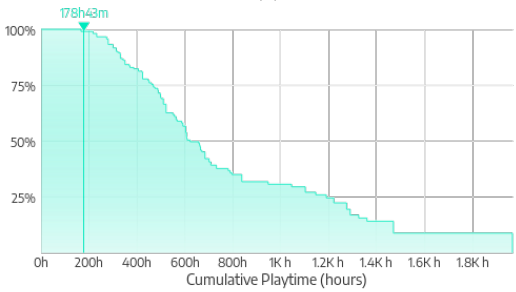
Non VIP PUs can be found in the middle ground, although qualitatively their behaviour resembles more that of non-PUs. This is particularly obvious when looking at the survival probability curves in level. In the period considered, although they are much more numerous, basically no non VIP player makes it past a certain level (the particular level being less than 100 for non-PUs and around 275 for PUs). On the other hand, all VIP players



(a)



(b)



(c)

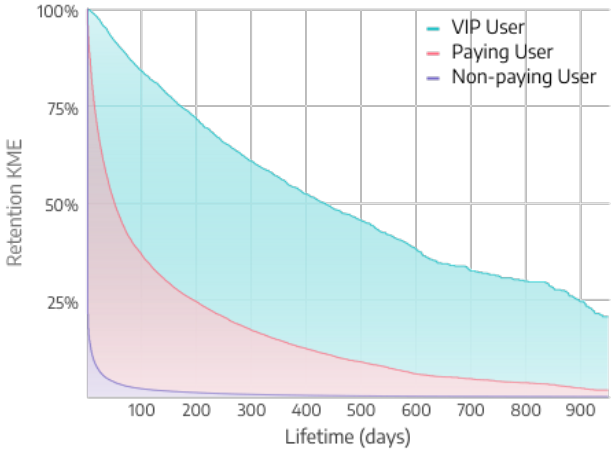
FIGURE 5.1: Predicted survival curves in days since first login (a), game level (b) and accumulated playtime (c) for a particular player. They logged in for the first time 164 days ago, have played over 178 hours since and reached level 151. They are expected to play roughly 230 days more, reach level 190 after having played a total of more than 600 hours before quitting the game. Elaboration using data from AoI. The images have previously appeared in [224].

that make it past level 300 continue to be active until the end of the period considered (reaching well over level 400).

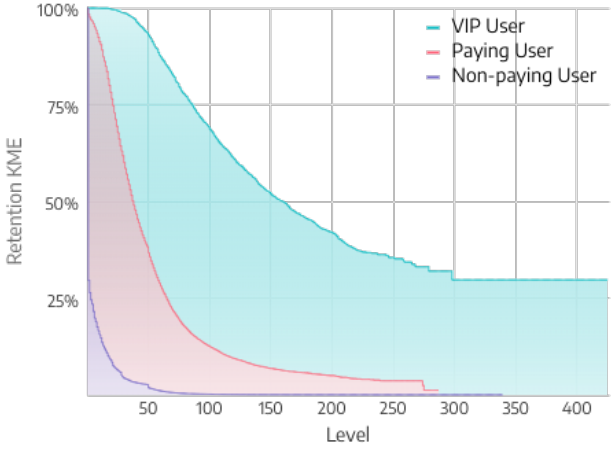
Another interesting particularity of the VIP curve is that it remains almost flat for the first 25 levels, meaning absolutely all VIP players make it to this level (while nearly half of the non VIP PUs and less that 5% non-PUs do). Given that the VIP survival curve in lifetime does not show the same behaviour, and assuming in general those leaving the game early on will also be the same not progressing a lot, this can only be explained by either whales with short lifetimes still having a high accumulated playtime that would allow to still make significant progress, or to the actual spending being of great help in getting through the early levels fast. Figure 5.3 shows the Kaplan-Meier estimates of the survival probability as a function of days since first login (plot (a)), level (plot (b)) and playtime (plot (c)) for VIP players. The lack of a flat beginning for the playtime survival curve points to the spending facilitated progression as that of greater influence. This makes even more sense when it is taken into account that players are VIP after their accumulated expenditure reaches a certain threshold. This means that VIP players leaving with short lifetimes/low accumulated playtime spend much more per day/hour played than the more typical loyal whales. More details on the relations of this variables (and their predictions) for VIP players are given in chapter 7.

Going back to figure 5.2, survival curves in lifetime are hyperbolic shaped, with the VIP one having however a much more lowly varying slope, approximately constant for long periods. At the end of the lifetime values available in our sample, after all curves seem to be approaching their asymptotic behaviour, there is however a sudden increase in the slopes of both types of PUs (no non-PUs are left) around day 850. This suggests even very loyal players loose interest after this point.

As was already mentioned, Figure 5.3 shows the Kaplan-Meier estimates only for VIP players, including now playtime (c). Note that while the playtime survival probability decreases much faster than the lifetime probability (and even more so than the level one, which has already been noted to remain approximately constant at the beginning), similarly to what happens with in-game progression, it saturates at around 25% after around 5000 hours of accumulated playtime. At the end of the period considered, of the 25% of the whales still active, all of them had played over 5000 hours and reached at least level 300. No similar statement con be done regarding lifetime. Shaded area in this figure represents 95% confidence intervals.



(a)



(b)

FIGURE 5.2: Kaplan–Meier estimates of the survival probability as a function of time since first login (a) and game level (b) for non-PUs (purple), PUs (pink) and VIP players (green). Elaboration using data from AoI.The images have previously appeared in [224].

Figure 5.4 exemplifies the recursive partition with which a conditional inference survival tree proceeds. At each node, the players are split in terms

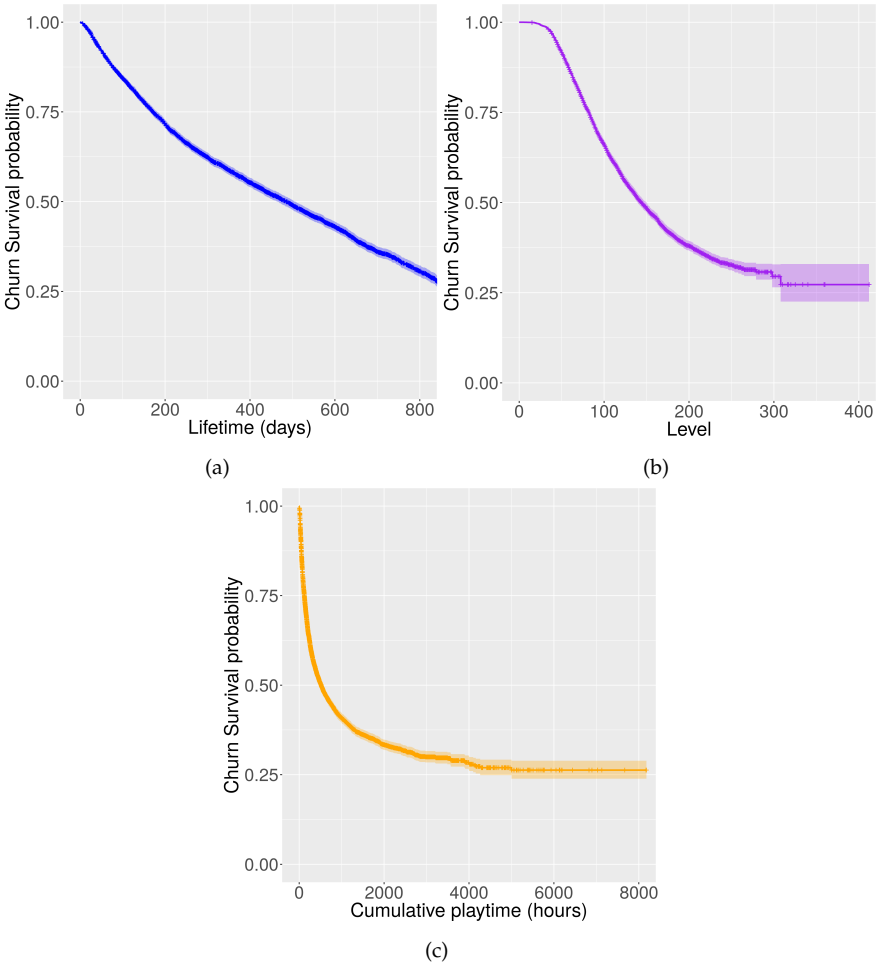


FIGURE 5.3: Kaplan–Meier estimates of the survival probability for VIP players as a function of time since first login (a), game level (b) and playtime (c). Elaboration using data from AoI.

of the selected feature using Kaplan-Meier estimates. These are plotted for the four terminal nodes shown in this figure, exhibiting markedly distinct behaviours.

Validation plots for a conditional inference survival ensemble run for VIP players are shown in figure 5.5 for lifetime predictions, figure 5.6 for level predictions and figure 5.7 for playtime predictions. They show predicted vs

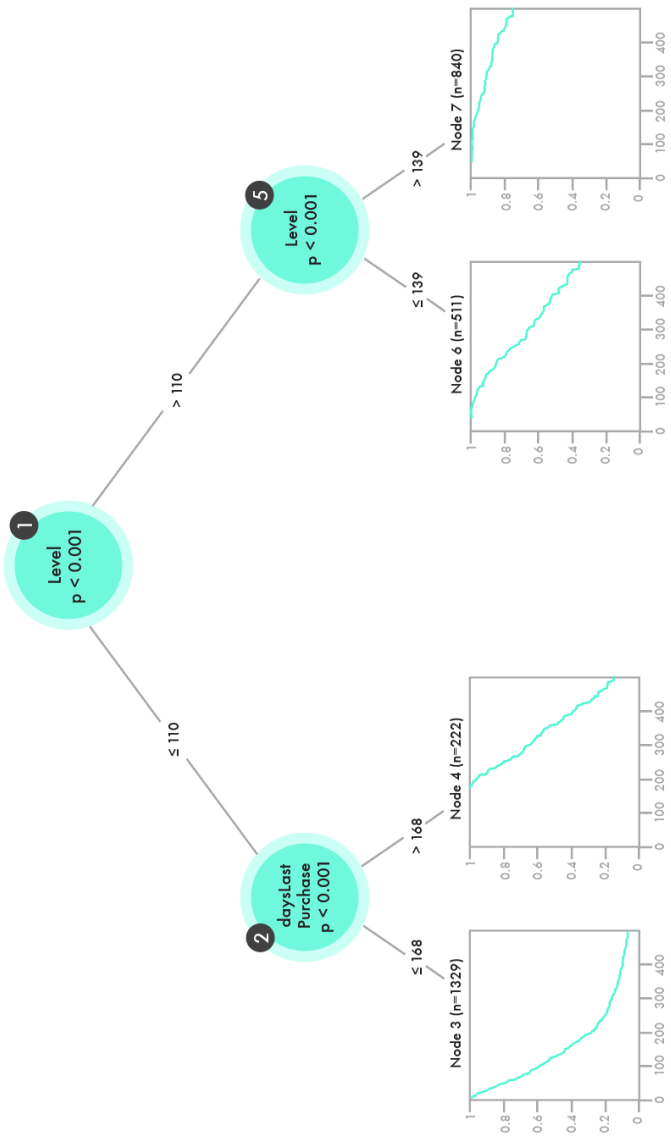


FIGURE 5.4: Example of a possible conditional inference tree. The four terminal nodes are shown together with their corresponding Kaplan-Meier survival estimates for each group of n players. Elaboration using data from AoI. The images have previously appeared in [224].

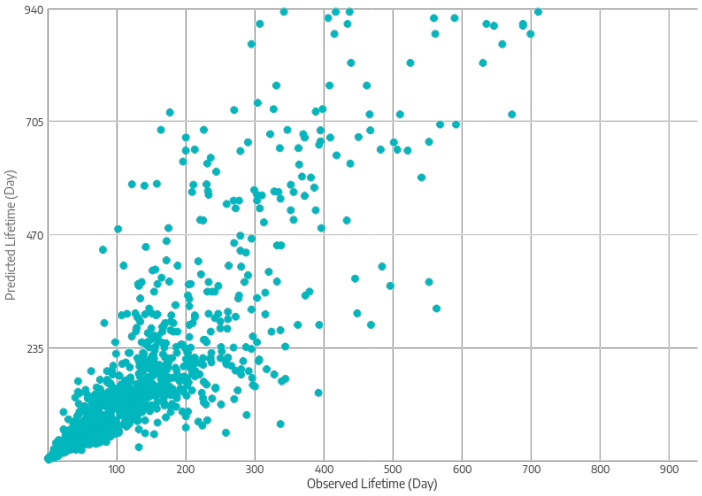
observed values ((a) plots) and mean-difference plots ((b) plots). They show overall a remarkably good performance of the models, with a clear bias in all three variable to predict higher than observed values though.

5.3 IMPACT OF ZOMBIES AND RESURRECTED

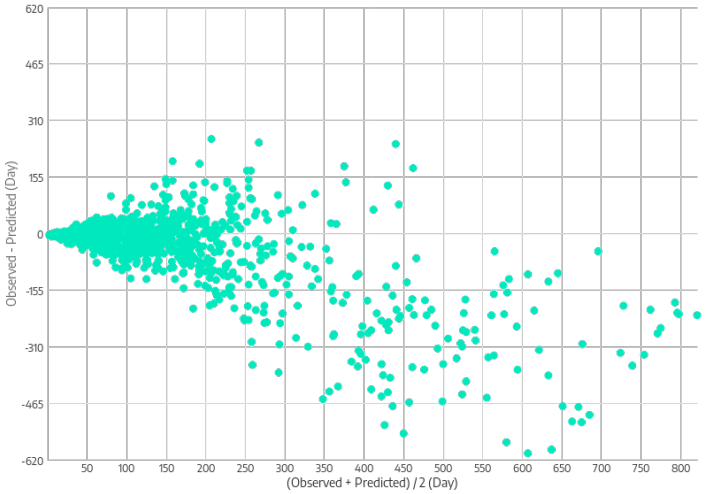
Figure 5.8 shows the Kaplan-Meier estimates of the survival probability as a function of days since first login, level and playtime for zombies, resurrected, purchase churners and normal (all the rest) VIP players. It is basically figure 5.3 stratified by type of churner. The markedly different survival patterns for different churn profiles supports the idea that it could be useful in the modelling to take the nature of the churners into account. Simply using this as an additional feature did not however improve model performance. This chapter examines the impact in the predictive accuracy (of both the binary classification and survival approaches) of including or excluding zombies, resurrected and purchase resurrected from VIP churn prediction models. The goal is to elucidate if information from those types of churners prevents the models from learning optimally the behaviour of a *typical* VIP player by introducing noise. This work was first presented in [132], which also assesses the impact of considering or not these players in purchase churn predictions, which will be discussed in chapter 6 section 6.4 in this thesis.

In particular, figure 5.8 shows that zombies have the lowest survival probabilities in the three variables, except for lifetimes shorter than approximately 150 days, for which normal VIP players are in a greater risk of churning.Purchase resurrected players, on the other hand, have systematically much higher survival probabilities in all ranges and variables to all other groups. The survival behaviour of resurrected players in level and playtime is qualitatively very similar and with quantitatively only slightly higher survival probabilities than that of normal players. With respect to lifetime, however, their behaviour is remarkable. For the first approximately 500 days of play, they present higher survival probabilities than normal players (markedly so for the first year or so). After that, while the survival curve for normal players flattens, that of the resurrected continues to drop at an approximately constant rate, inverting the trend.

Performance of the conditional inference forests for the binary classification problem (described in section 5.1), and for the conditional inference survival ensembles for survival in lifetime, playtime and in-game progression (described in section 5.2), when including and excluding the different

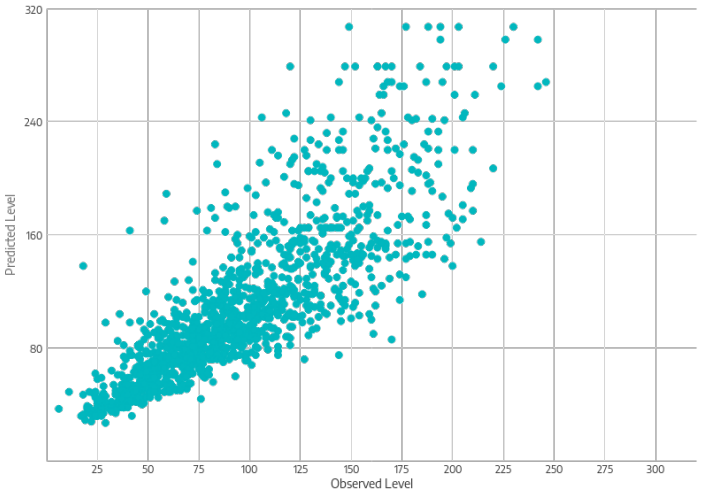


(a)

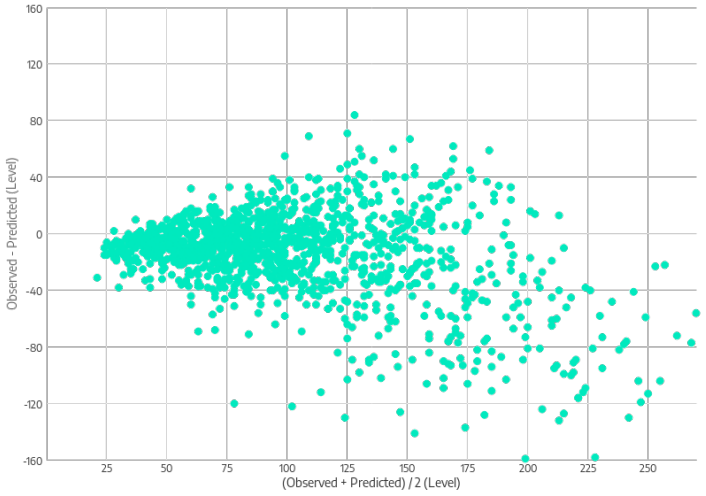


(b)

FIGURE 5.5: Validation plots for the conditional inference survival model lifetime predictions for PUs. Plots show predicted vs observed values (plot (a)) and mean-difference plots (plot (b)). Elaboration using data from AoI. The images have previously appeared in [224].

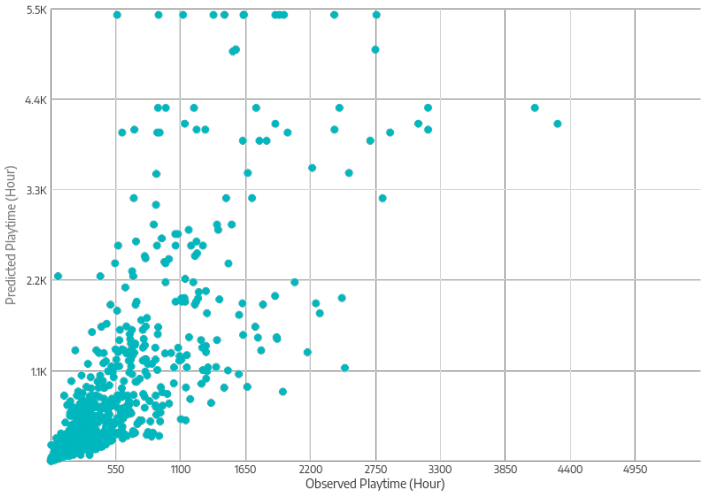


(a)

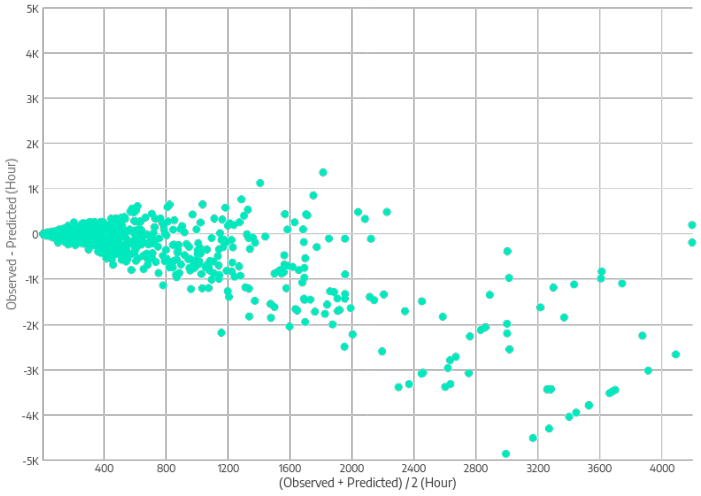


(b)

FIGURE 5.6: Validation plots for the conditional inference survival model level predictions for PUs. Plots show predicted vs observed values (plots (a)) and mean-difference plot (plot (b)). Elaboration using data from AoI. The images have previously appeared in [224].



(a)



(b)

FIGURE 5.7: Validation plots for the conditional inference survival model playtime predictions for PUs. Plots show predicted vs observed values (plot (a)) and mean-difference plot (plot (b)). Elaboration using data from AoI. The images have previously appeared in [224].

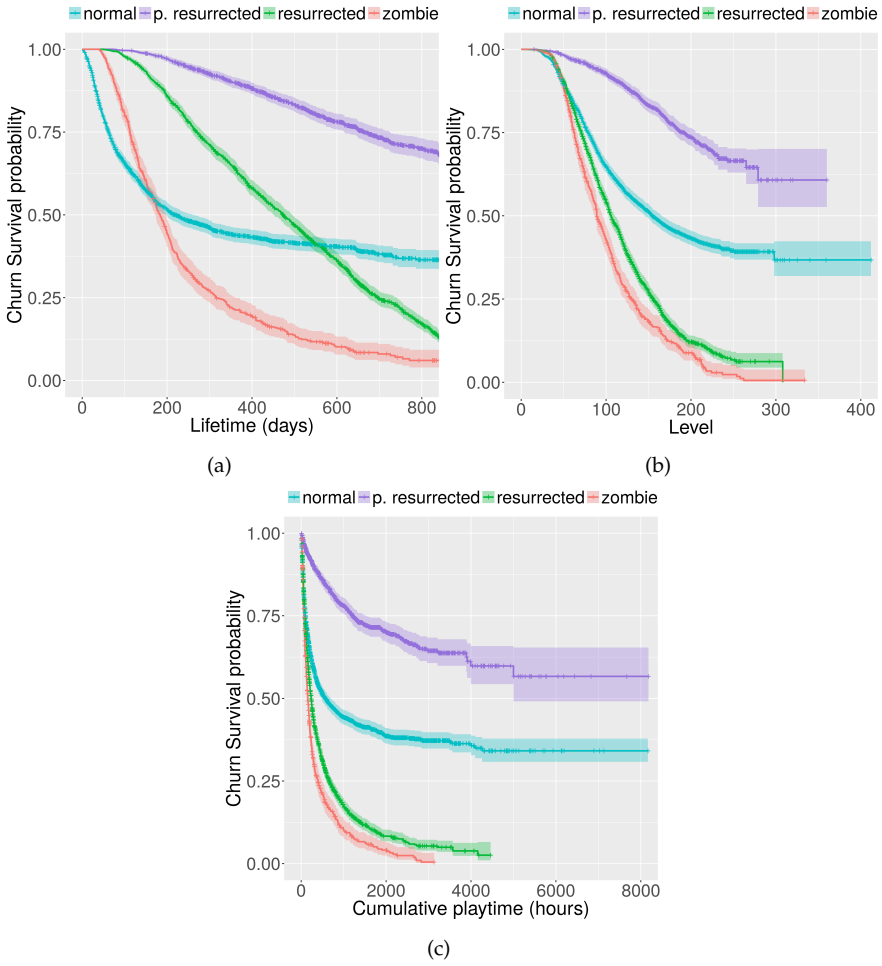


FIGURE 5.8: Kaplan–Meier estimates of the survival probability as a function of time since first login (a), game level (b) and cumulative playtime (c) for VIP players. Curves are stratified by churner type: *normal*, *zombie*, *resurrected* and *purchase resurrected* players. Shaded areas represent 95% confidence intervals. Elaboration using data from AoI. The images have previously appeared in [132].

combinations of types of churners, was investigated. Ensembles of size 1000 were used in all cases. The experiments run used data between 2014-10-02 and 2017-05-01, with data until 2018-03-01 used for training, and the rest

for validation. The experiments were run for VIP players only. By the end of the period considered, 21% of the players had been tagged as churners and 5% as purchase churners. Nearly 30% had been resurrected throughout their lifetime, and 23% had purchase resurrected. Around 10% had been zombies at some point. While the large amount of resurrected could hint at a poor churn definition, these tend to leave the game not long after resurrecting, not contributing very much to the fraction of false churners in the long run. Feature selection was done independently for the different models and variables, using always features constructed out of indicators present for many different titles (player playtime, purchases, actions...). The possibility of adding the type of churner as a feature (for example, a zombie feature which is 1 if player is zombie and 0 if they are not) was also explored but were found to bias the models towards their behaviour, degrading the predictions for normal churners.

Validation was performed using AUC for the binary case and error curves and IBS for survival (see chapter 2 section 2.9 for more details). The set of players used for validation was the same in all cases, and excluded zombies, resurrected and purchase resurrected, as the goal is to assess the accuracy of predictions for normal players. Results are summarised in table 5.1, with figures 5.9, 5.10 and 5.11 comparing prediction error curves on the different variables (for lifetime, level and playtime respectively).

Note that the exclusion of different types of churners or not has close to absolutely no impact on binary results, but a large one on the survival predictions accuracy. Survival models rely on the whole lifetime of players to learn about survival probabilities, and are thus much more sensitive to noise introduced by the erratic churn behaviours of certain players. This makes it safe to make the same training decisions for binary and survival for consistency.

Focusing on survival, it is clear from figures 5.9, 5.10 and 5.11 (particularly from their (a) plots, which compare the effect on the three variables of removing each of the groups under study individually) that different groups have different impacts depending on how large the value of the variable under consideration is. For very short lifetime, level and playtime, there is no significant impact from removing any of the groups. For short lifetime, level and playtime, removing zombies is what has a larger positive impact in error reduction, followed by removing resurrected that shows a similar pattern with less impact. While the improvement becomes starker for larger lifetime in days after first login, it is reversed for larger values of level and playtime (in these regions all training sets have similar performances

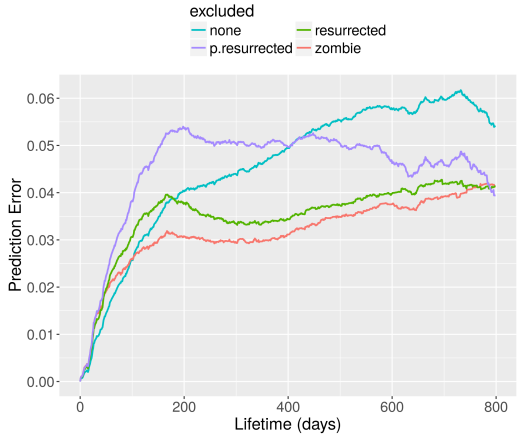
CHURN	Binary models (AUC)	Survival models (IBS)		
excluding from training		lifetime	level	playtime
none	0.95	0.072	0.069	0.060
zombie	0.93	0.034	0.047	**0.035**
resurrected	0.90	0.043	0.048	0.041
p. resurrected	0.95	0.104	0.084	0.060
zombie, resurrected	0.94	**0.029**	**0.041**	**0.035**
zombie, p. resurrected	0.93	0.057	0.068	0.049
resurrected, p. resurrected	0.92	0.071	0.068	0.057
zombie, resurrected, p.resurrected	0.94	0.053	0.059	0.050

TABLE 5.1: Area under the curve (AUC) for binary model and the integrated Brier score (IBS) for survival model (in terms of lifetime, level and cumulative playtime) for the different situations with regard to the training sample: including all users (*none*) vs. excluding zombie, resurrected or purchase resurrected players (or combinations of them). The best results for each model and variable are highlighted in bold.

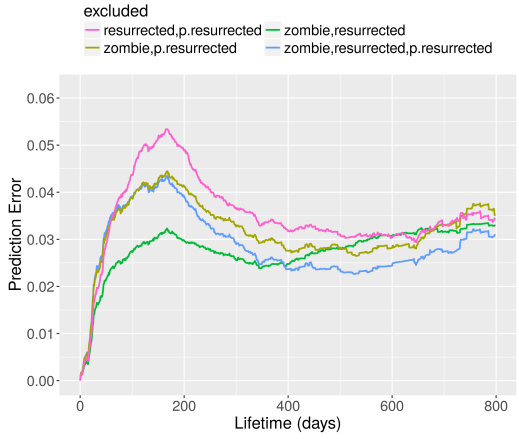
though). Purchase churners seem to have the opposite effect: their removal enhances performance for large values of lifetime, level and playtime, but severely impacts prediction quality for small values of the three variables. While removing zombies and/or resurrected has an overall positive impact, training without purchase resurrected has an overall negative impact, as can be checked in table 5.1. This makes sense, as purchase churners might show at some point an erratic behaviour with regards to purchasing, going for long periods without spending, but can still have a very typical login activity from which survival models' learning benefits. The preferred training set would be then to remove zombies and resurrected but keep purchase resurrected.

5.4 SUMMARY AND CONCLUSIONS

A first simple proxy to study player engagement is modelling player churn probability. This can be done using binary classification models, but survival ensembles provide a more interesting perspective, as they can output for each player how their survival probability evolves with lifetime, in-game



(a)



(b)

FIGURE 5.9: Prediction error curves for PU churn as a function of lifetime. The different lines represent model runs excluding zombies (red), resurrected (green) or purchase resurrected (purple) players (plot (a)) and combinations thereof (plot (b)) from the training sample. Combinations represented in plot (b) are: (i) resurrected and purchase resurrected (pink), (ii) zombies and purchase resurrected (brown), (iii) zombies and resurrected (green), and zombies, resurrected and purchase resurrected (blue). Elaboration using data from AoI. The images have previously appeared in [132].

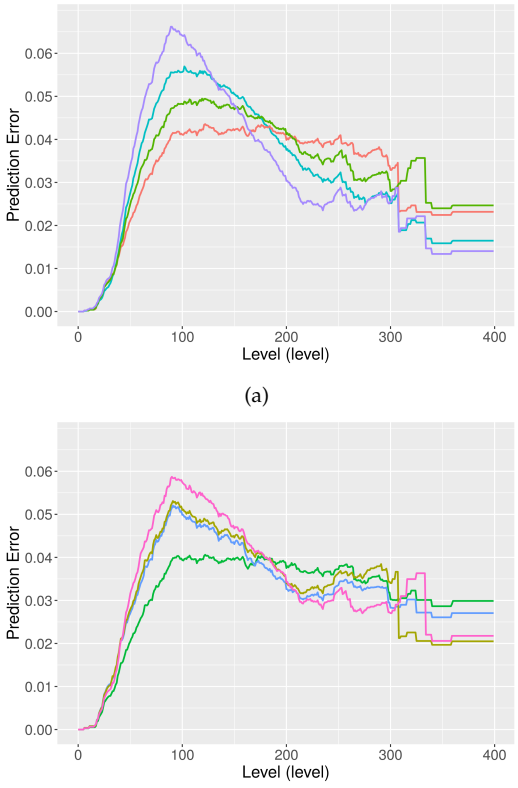
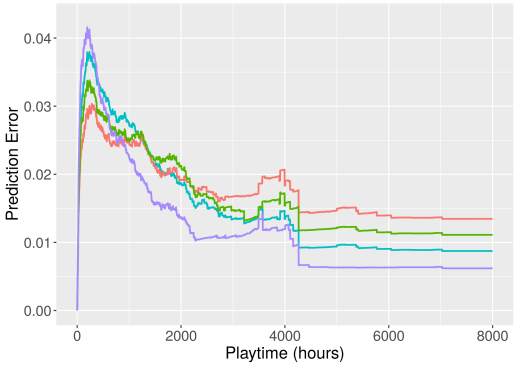


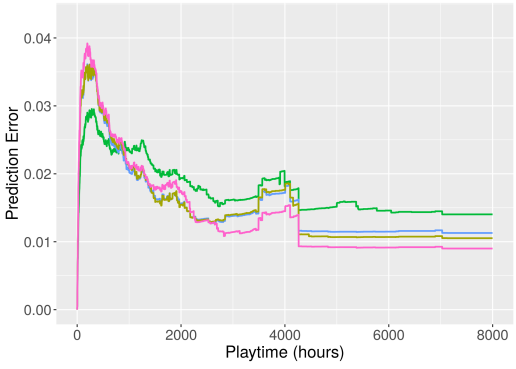
FIGURE 5.10: Prediction error curves for PU churn as a function of level. The different lines represent model runs excluding zombies (red), resurrected (green) or purchase resurrected (purple) players (plot (a)) and combinations thereof (plot (b)) from the training sample. Combinations represented in plot (b) are: (i) resurrected and purchase resurrected (pink), (ii) zombies and purchase resurrected (brown), (iii) zombies and resurrected (green), and zombies, resurrected and purchase resurrected (blue). Elaboration using data from AoI. The images have previously appeared in [132].

progression and accumulated playtime. In particular, conditional inference survival ensembles outperform other methods in this particular task.

Survival average behaviour (given by Kaplan-Meier estimates) of different types of players can be very different. In particular, disengagement takes



(a)



(b)

FIGURE 5.11: Prediction error curves for PU churn as a function of playtime. The different lines represent model runs excluding zombies (red), resurrected (green) or purchase resurrected (purple) players (plot (a)) and combinations thereof (plot (b)) from the training sample. Combinations represented in plot (b) are: (i) resurrected and purchase resurrected (pink), (ii) zombies and purchase resurrected (brown), (iii) zombies and resurrected (green), and zombies, resurrected and purchase resurrected (blue). Own elaboration using data from AoI. The images have previously appeared in [132].

place very differently for non-PU, VIP and non VIP PU players. Focusing on VIP players only, different types of churners show also markedly different survival patterns. It has been shown that excluding from the training set certain types of churn like profiles, in particular active players with very

little activity and those returning to the game after very long periods of inactivity, improves the performance of the survival ensembles in predicting churn for normal players, while not impacting the binary classification approach.

An obvious application of having these sort of predictions running in production is an early detection of players in risk of churning. This allows studios to target these users specifically, in order to reengage them, or even to try and redirect them to other titles that might be interesting to them if the modelling is indicating their life-cycle within the game has been exhausted. If a player, for example, is disengaging fast in playtime but still has a very high predicted final level, this could be signaling the game is easy for them and they will complete it soon (and hence no sense in attempting re-engagment, they will soon be *done* with the game).

Additional ML models can be used to target them in a personalised manner, for example, making content particularly suited to their likes available (see chapter 8 for a description of an item recommendation system, for example). Note also that the probabilistic nature of the output of these models allows for a discrimination, even for the same predicted time to churn, of users who are disengaging slowly and will maintain relatively high probabilities of still being active for a long period of time after the predicted time of churn (which would probably be easier to reengage at least in the short term), from those who are very rapidly disengaging and will soon after have a negligible probability of coming back to the game (and with whom any attempt of re-engagement will most likely not be successful).

The multidimensional nature of the approach (providing in predictions of time to event, but measuring *time* in three different ways) also brings added value. It could, for example, allow for an early detection of problematic content (level) in the game (too easy, too difficult, too boring...), if many players are predicted to have a significant drop in survival probability at that point. Survival predictions in different variables and their combination can prove useful in profiling players and in generally understanding game dynamics, as will be explored in chapter 7.

6

PREDICTING PURCHASING ENGAGEMENT

Das Leben muss nicht einfach sein, vorausgesetzt, es ist nicht leer.

— Lise Meitner

Purchasing engagement, defined analogously to login engagement, as that which resolves when a player is a genuine PU and when they are not, is obviously of great interest to studios. Not only should they pursue its study in order to better monetise their games, but also crucially to detect, track and disengage players that could be displaying problematic addictive behaviours. The appeal of studying player purchasing habits in games is obvious from a general human behaviour perspective. As is repeatedly stressed throughout this thesis, all methodologies presented to that effect can be directly translated to virtual in-game currency. In terms of economic behavioural studies, this becomes particularly interesting when concerning role type games mimicking human societies. Here money typically has to be earned performing certain tasks, and the players then need to decide carefully where it will be better spent.

This chapter deals with everything having to do with measuring and predicting individual player purchasing engagement. A lot of it follows along the lines of the previous one (chapter 5). All methods that were applied there to predict churn probability (considered as the main marker of engagement loss), can be also used to better understand both individual player conversion and purchase churn (both discussed at very basic segmentation or game level in chapter 4).

Another quantity of interest when considering individual purchasing behavioural analysis is obviously that of number of purchases and total expenditure. Predictions of these quantities can be used for early detection of whales among other things. The expected outlay during the duration of the relation is of general interest in non-contractual setups, and is typically referred to in the literature as *Lifetime Value (LTV)* (or customer lifetime value or lifetime customer value), as introduced in chapter 2 section 2.5.

Section 6.1 discusses the use of survival models to predict after how many days since first login, accumulated playtime and in which level will players make their first purchase, work that was first published in [133].

The use of binary classification methods and survival models to predict time to purchase churn is introduced in sections 6.2 and 6.3 respectively, and the impact of using or not different churn profiles in the training set analysed in section 6.4. This approach was first discussed in [132]. Different methods for predicting player LTV are then compared in section 6.5, in work which first appeared in [51, 224]. The chapter closes with a summary and conclusions in section 6.6. All the content of this chapter is an original contribution of this thesis (as second or third author).

6.1 PREDICTING PLAYER CONVERSION

The first step when discussing purchasing behaviour is potential paying user detection. This problem can be approached by using survival models to predict time, playtime and level to first in-game purchase. Previous attempts at predicting conversion in video games include [256], where the problem was treated as a binary classification problem using support vector machines and decision forests, or [76], where different classification methods for PU early detection and regression algorithms for LTV prediction are explored. Conversion has been studied in other fields and contexts, such as e-commerce [66, 161], medicine [293] or career switching [286].

In this section the results presented in [133] are described in detail for AoI. In the aforementioned paper, the analysis was also performed for an additional game (*Grand Sphere*) with comparable results. Three different survival analysis approaches were considered for predicting time to conversion: Cox regression, random survival forests and conditional inference survival ensembles (see chapter 2 sections 2.4.1, 2.4.2 and 2.4.3 respectively). As the game progresses, a non-PU can either churn or become a PU. The random survival forest with churn as an alternative competing risk was therefore also evaluated. The alternative event that would prevent observing the transition of a non-PU to PU would thus be churn, as illustrated in figure 6.1. Following the generic approach of this thesis in what concerns survival models, predictions on days after first login, level attained and accumulated playtime to the event of interest (conversion to PU in this case) were considered. This is implicitly taking into account how the ability and play intensity of players factor in.

The idea is to evaluate if time to conversion to PU can be accurately predicted in an operational setup. Data between January 2015 and February 2017 was used, considering only players that logged in at least two days. Of all new users logging each day for the first time, most are *one-time*

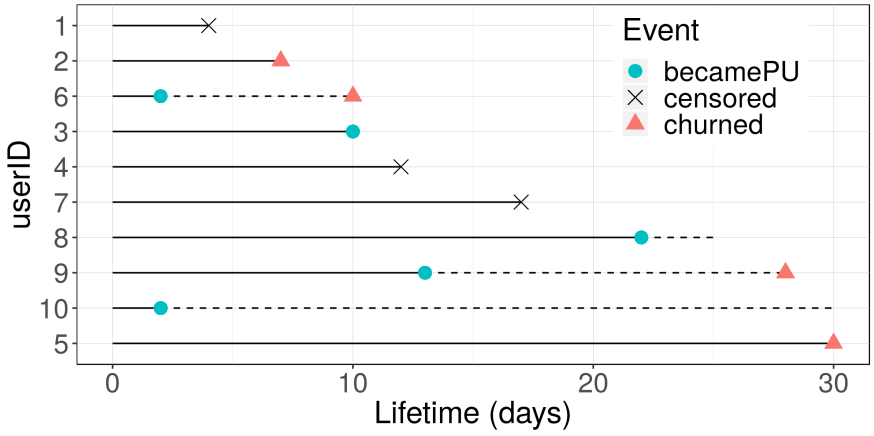


FIGURE 6.1: Plot showing schematically how churn and becoming PU can be considered competing risks. Ten players are tracked for 30 days of lifetime. Players may become PUs (circles) or churn (triangles) at some point. It is also possible that none of this events is observed within the observation period (crosses). They could however happen later on, so these are in fact illustrating the censored character of the dataset. The image has previously appeared in [133].

comers, i.e. players that log in only one day, never to come again (this is clear in figure 5.2). Eliminating new comers from the sample reduces class imbalance, as most of them will never become PUs. Non one-time comers will be considered anyway as soon as they log again, and a prediction of time to conversion issued for them. And for those new comers that become PUs in the first day, the predictions would have given very little (if any) added value.

This translates to around 30 000 players, of which 5,32% became PUs at some point. To run the experiments, random samples were taken, with 30% of the players assigned to the training set and 70% to the test sample. The training set is purposefully much smaller than the test set, as one of the aims of this exercise is to find out whether these models can provide accurate predictions in an operational setup. This would involve huge datasets, and thus the approach would be to train with only a small subsample of the total population available.

Note in this problem, survival curves give the probability for a given individual to remain non-PU at a certain point in time. It is therefore more intuitive in this case to think in terms of its inverse, the cumulative incidence

function, which gives the probability that the event of interest—becoming a PU—*does happen*. Figure 6.2 and Figure 6.3 show these cumulative incidence functions (inverse of the Kaplan-Meier estimates) in terms of days after first login (a), level (b) and accumulated playtime (c), for all players of the game and for PUs only respectively.

These figures highlight a different behaviour in terms of lifetime as compared to playtime and level. Even when considering the whole population, most of the players that reached the highest levels or played for very large cumulative playtime did indeed become PUs. When looking at days since first login, however, players can remain in the game for very long periods of time without necessarily becoming PUs. This signals that, while lifetime is a key measure when analysing player behaviour, playtime and in-game progression are typically better indicators of engagement. While less than 25% of players have become PUs after two and half years using the game (longest period available in the sample considered), 50% have made at least one purchase after 500 hours of play or by the time they have reached level 150. Players with up to 8000 hours of accumulated playtime and of nearly level 350 can be found in the sample, and of those, over 80% are PUs.

When looking at the cumulative incidence function when considering only PUs, it is obvious that conversion happens much more quickly in lifetime and playtime than in level, probably indicating the spread in skills of PUs can still make them reach certain levels at very different points in time, even when playing with similar frequency and duration. After less than a month after their first login, and less than 10 hours of playtime, more than 75% of the players that will indeed make the conversion have already made at least one purchase. Note these are both very short periods as compared to scales of observed lifetimes and playtimes available in the sample considered (of up to 900 days and 8000 hours). Meanwhile, it takes players to get to level 50 to reach similar levels of 75% conversion when considering in-game progression, which is already over 15% of the maximum attained by any player in the sample.

Note that in this case, unlike when using survival models to predict churn (chapter 5 section 2.4) and purchase churn (section 6.3), there is no ambiguity with respect to the event definition: conversion to PU happens when the user makes their first purchase.

Features were selected independently for each of the response variables. As for all survival models discussed in this thesis, they were constructed applying different statistical operations (minimum, maximum, median, average...) to time series based on quantities generic across many games, such

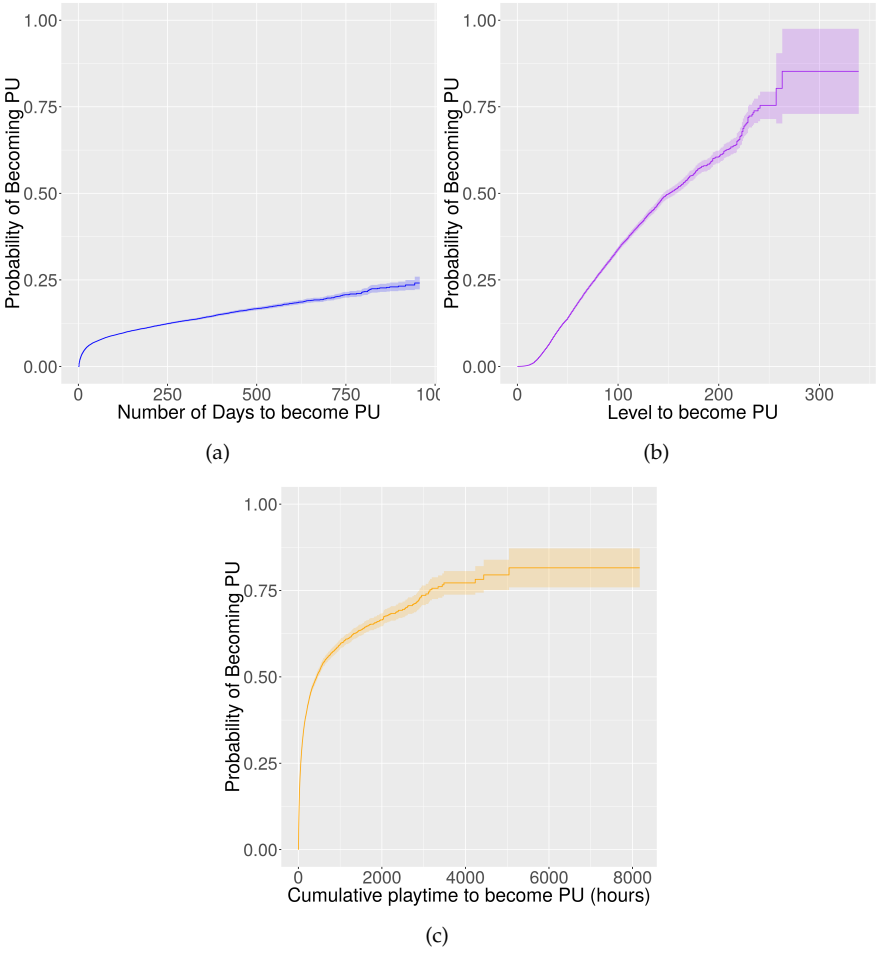


FIGURE 6.2: Probability of being a PU as a function of lifetime (a), in-game progression (b) and accumulated playtime (c) for all players except one-time comers (as given by the inverse of the Kaplan – Meier estimates). The shaded area represents the 95% confidence interval. Elaboration using data from AoI. The images have previously appeared in [133].

as level-ups, playtime, actions, sessions, social interactions... All ensemble methods used 900 trees as base learners.

To assess the accuracy of the predictions, scatter plots of predicted vs. observed variables (figure 6.6 for conditional inference survival ensembles,

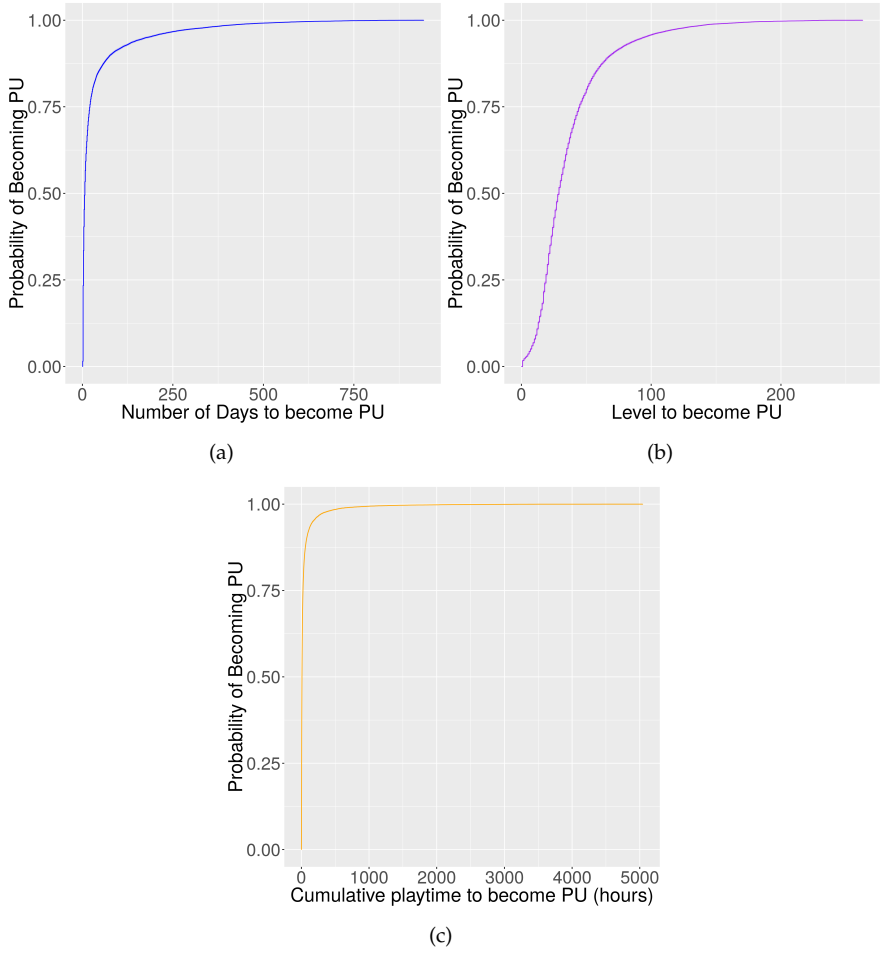


FIGURE 6.3: Probability of beeping a PU as a function of lifetime (a), in-game progression (b) and accumulated playtime (c) for PUs (as given by the inverse of the Kaplan – Meier estimates). The shaded area represents the 95% confidence interval. Own elaboration using data from AoI. The images have previously appeared in [133].

figure 6.5 for random survival forests, and 6.4 for Cox regression) and of their log transformed (figure 6.9 for conditional inference survival ensembles, figure 6.8 for random survival forests, and 6.7 for Cox regression) are studied. Using logarithms has two advantages: it prevents over-penalisation

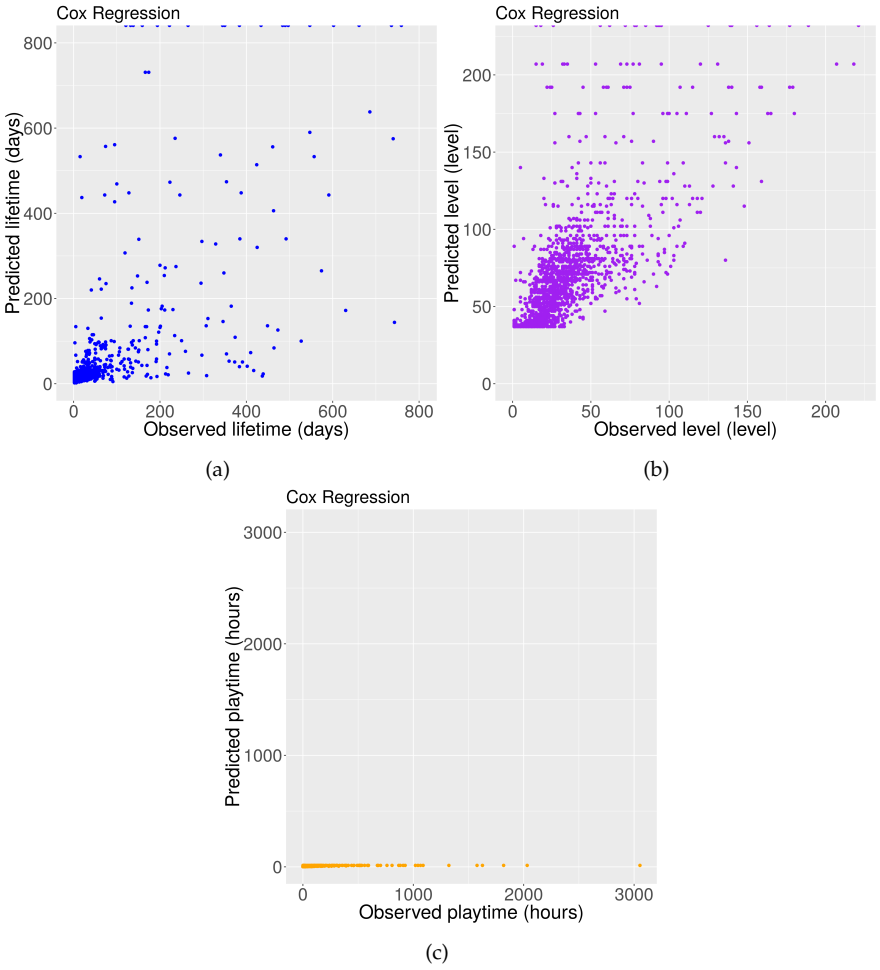


FIGURE 6.4: Validation plots for the conversion predictions of the Cox regression. Plots show predicted vs observed values for conversion times in lifetime (plot (a)), game level (plot (b)) and playtime (plot (c)). Predictions correspond to the median survival values. Elaboration using data from AoI. The images have previously appeared in [133].

of large errors for large values, while providing a close-up look at smaller values of the observed and predicted quantities. Root mean squared logarithmic errors (RMSLE) are listed in table 6.1, and percentage of false negatives (players that were not predicted to become PU but did) and

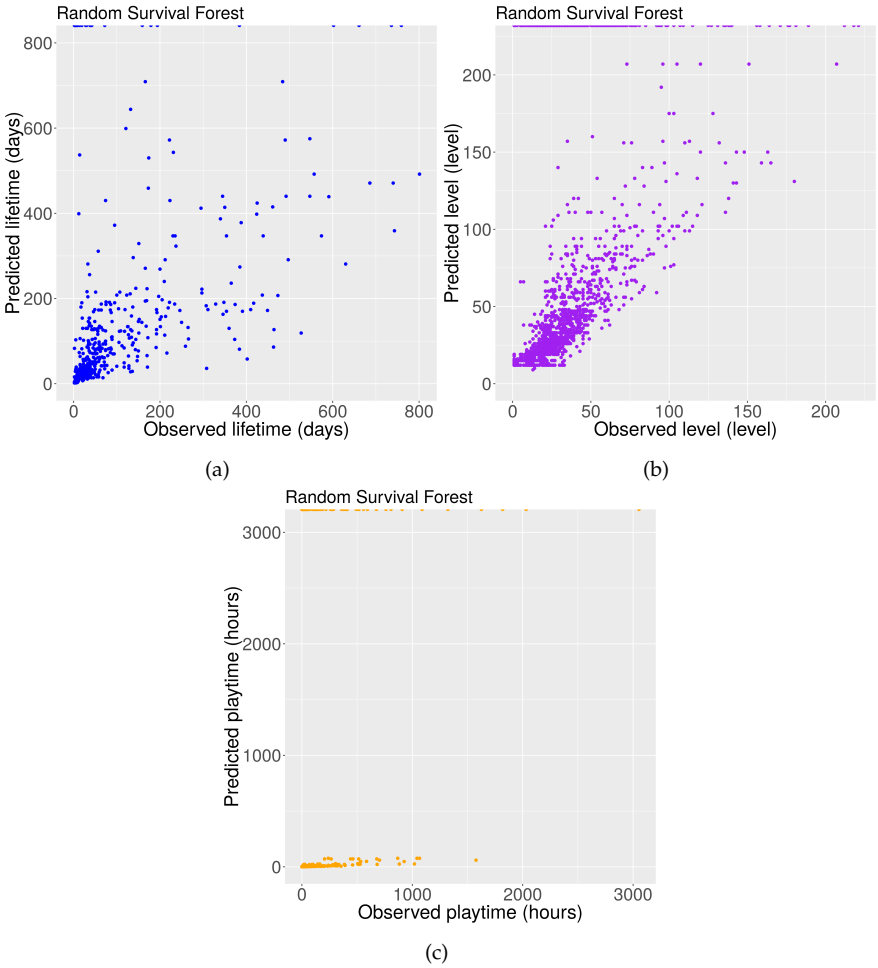


FIGURE 6.5: Validation plots for the conversion predictions of the random survival forest. Plots show predicted vs observed values for conversion times in lifetime (plot (a)), game level (plot (b)) and playtime (plot (c)). Predictions correspond to the median survival values. Elaboration using data from AoI. The images have previously appeared in [133].

positives (players who were predicted to become PU and did not) in tables 6.2 and 6.3 respectively.

The low percentages of false positives (table 6.3) and negatives (table 6.2) yielded by all models and variables highlight that they can all be

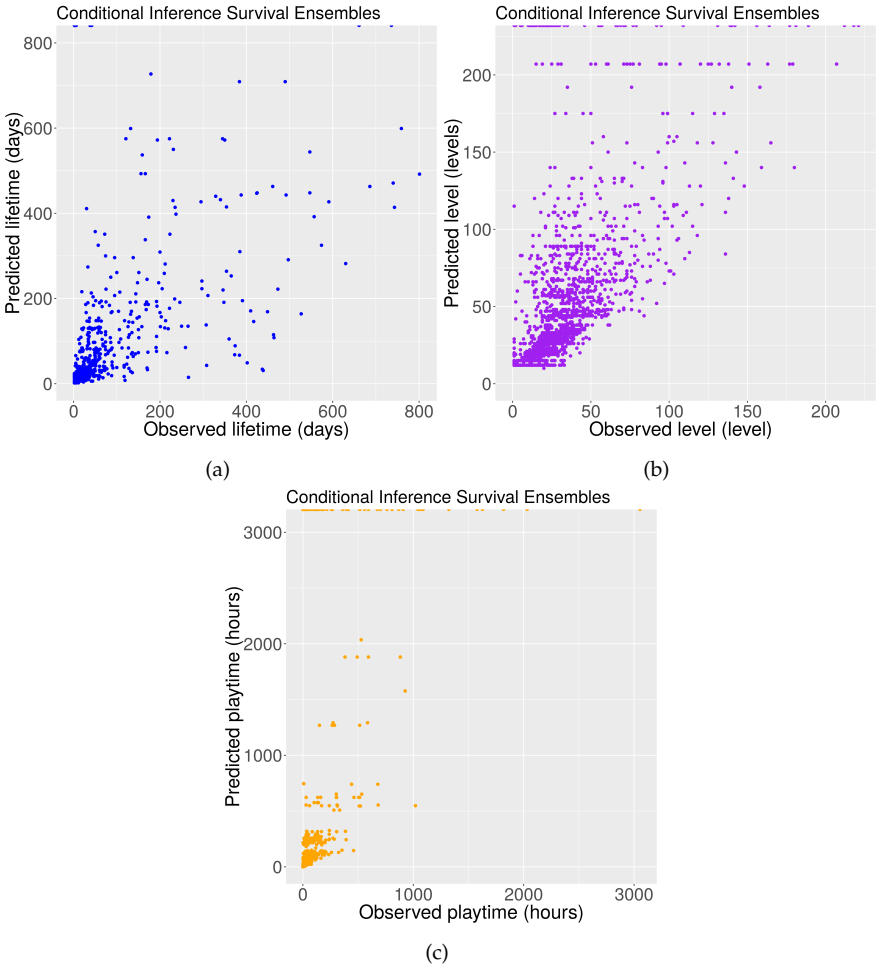


FIGURE 6.6: Validation plots for the conversion predictions of the conditional inference survival model. Plots show predicted vs observed values for conversion times in lifetime (plot (a)), game level (plot (b)) and playtime (plot (c)). Predictions correspond to the median survival values. Elaboration using data from AoI. The images have previously appeared in [133].

adequately used to provide a good binary classification of players that have PU potential and those that do not. Generally speaking, all methods also give reasonable predictions for when the conversion will take place

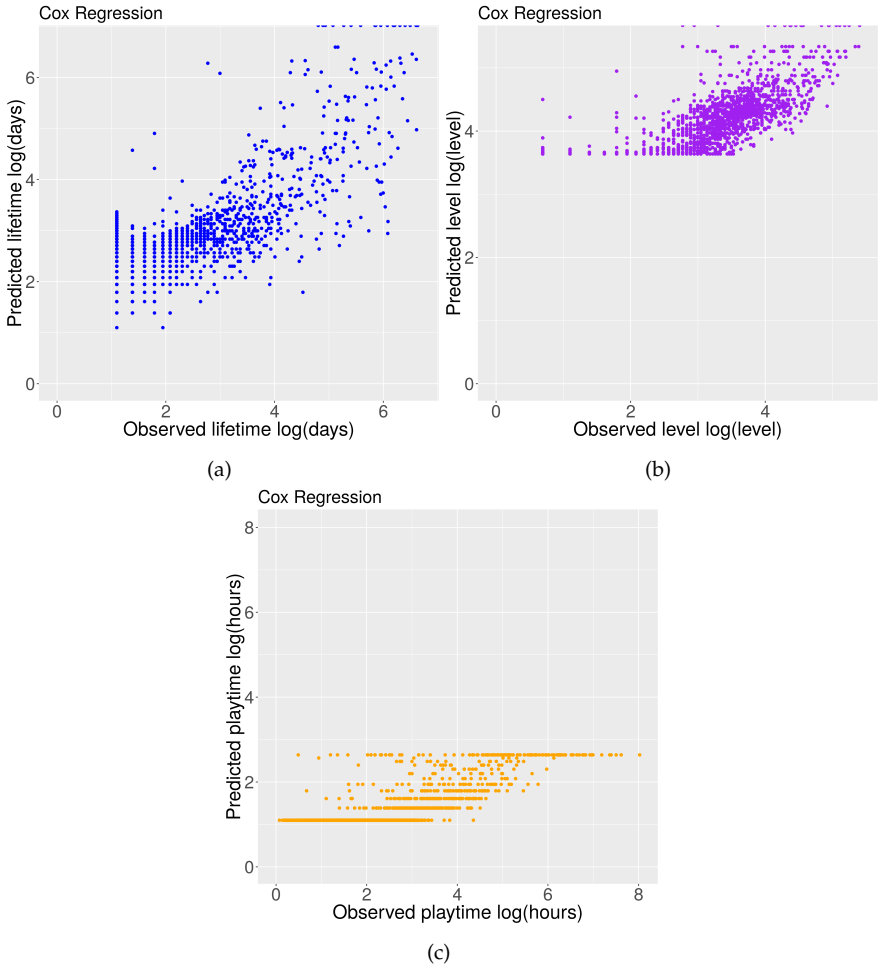


FIGURE 6.7: Log-log scatter plots of predicted vs observed values for conversion times in lifetime (plot (a)), game level (plots (b)) and playtime (plot (c)) using a Cox regression. Predictions correspond to the median survival values. The logarithm transformation provides a close-up look at the spread of the data points (cf. Figure 6.4). Elaboration using data from AoI. The images have previously appeared in [133].

(table 6.1). Cox regression, as expected, yields worse predictions across all variables than the ensemble methods.

Model	Lifetime	Level	Playtime
Conditional inference survival ensembles	0.54	0.69	0.47
Random survival forest	0.45	0.50	0.71
Random survival forest (competing risks)	0.50	0.63	0.85
Cox regression	1.08	1.00	0.79

TABLE 6.1: Root mean square logarithmic error (RMSLE) for time to conversion predictions of all survival models and variables considered.

Model	Lifetime	Level	Playtime
Conditional inference survival ensembles	0.27%	0.84%	0.60%
Random survival forest	0.18%	1.08%	1.01%
Random survival forest (competing risks)	0.61%	3.21%	0.58%
Cox regression	12.22%	1.69%	2.34%

TABLE 6.2: False negatives for all survival models and variables considered for PU detection.

Model	Lifetime	Level	Playtime
Conditional inference survival ensembles	3.68%	4.02%	4.02%
Random survival forest	3.70%	3.32%	3.42%
Random survival forest (competing risks)	3.41%	1.17%	3.27%
Cox regression	3.75%	4.19%	2.30%

TABLE 6.3: False positives for all survival models and variables considered for PU detection.

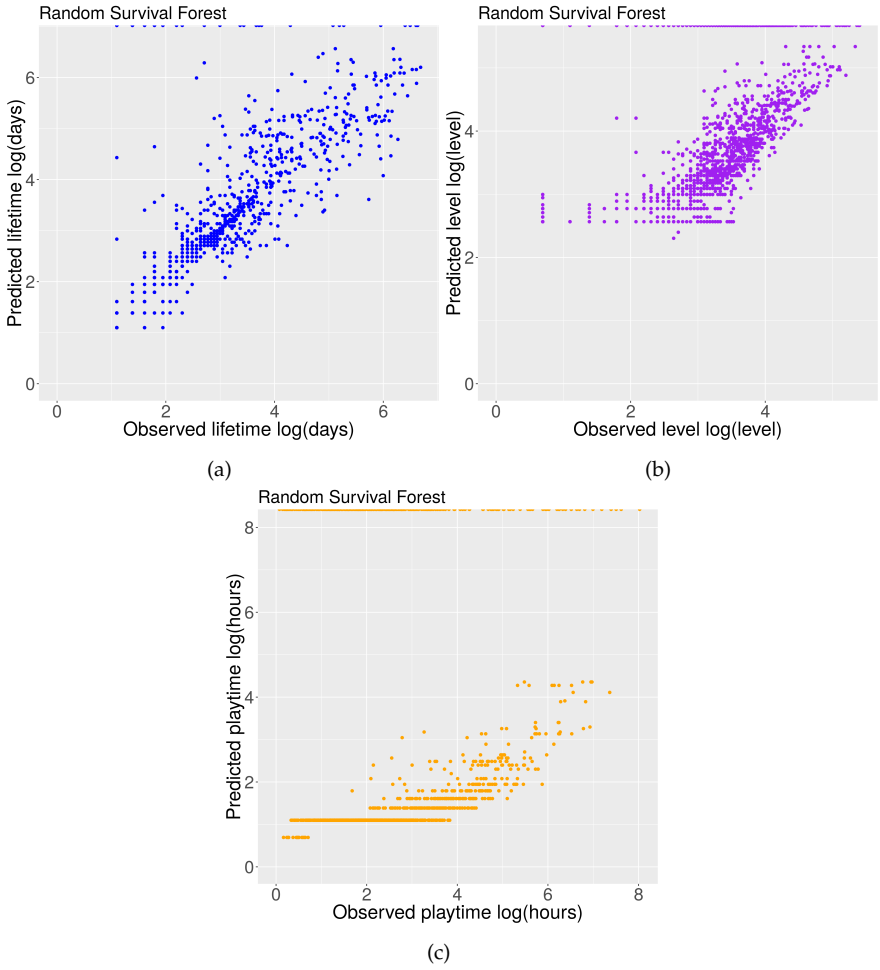


FIGURE 6.8: Log-log scatter plots of predicted vs observed values for conversion times in lifetime (plot (a)), game level (plot (b)) and playtime (plot (c)) using a random survival forest. Predictions correspond to the median survival values. The logarithm transformation provides a close-up look at the spread of the data points (cf. Figure 6.5). Elaboration using data from AoI. The images have previously appeared in [133].

Both ensemble methods have comparable performance. Note that the inclusion of competing risks in the case of the random survival forest does not improve the classification. As expected, it does reduce a little

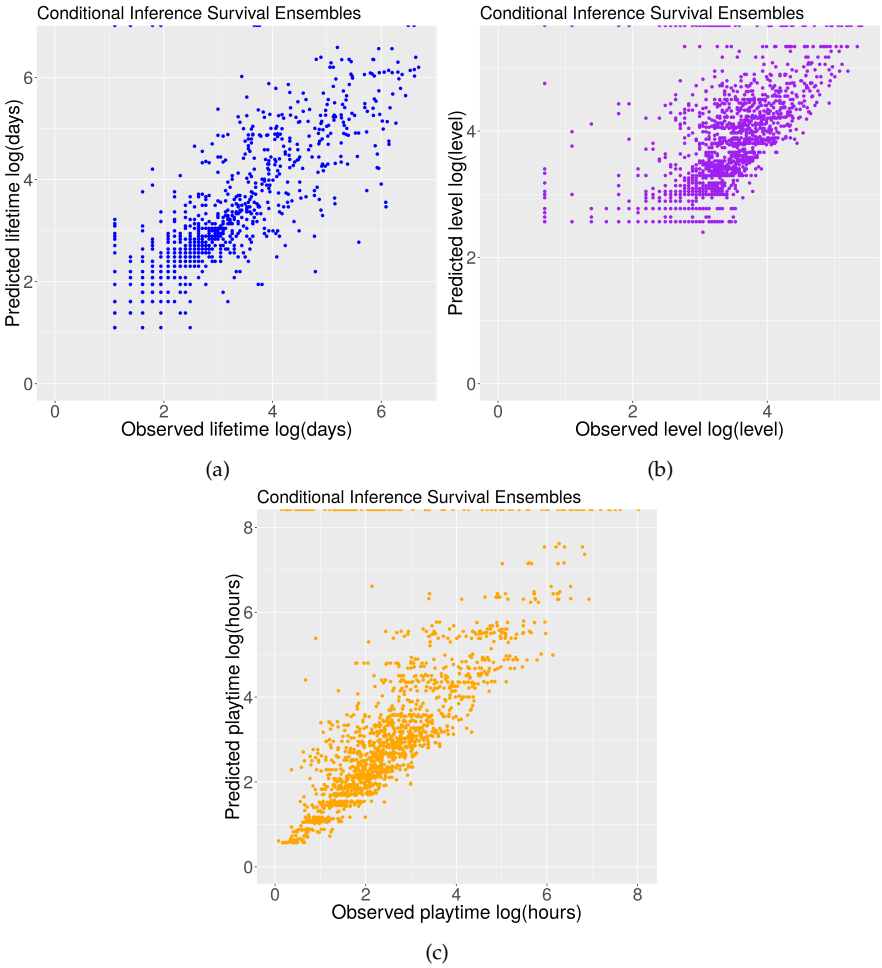


FIGURE 6.9: Log-log scatter plots of predicted vs observed values for conversion times in lifetime (plot (a)), game level (plot (b)) and playtime (plot (c)) using a conditional inference survival model. Predictions correspond to the median survival values. The logarithm transformation provides a close-up look at the spread of the data points (cf. Figure 6.6). Elaboration using data from AoI. The images have previously appeared in [133].

the rate of false negatives (players that churn before becoming PU are more likely to not be classified as potential PUs). Its detrimental effect in

the false positive rate is however larger (except in playtime). As for the predictions on when the conversion will occur, including competing risks degrades the predictions in the three variables. This suggests that churn and conversion to PU are in fact not competing risks: players are rarely considering simultaneously whether they should quit the game or start spending money on it. Players with a high probability of becoming PUs are normally not considering abandoning the game, and players in high risk of churning are normally not considering spending money in the game.

The scatter plots in figures 6.4-6.9 provide insights on for which scales of the different variables considered are the predictions of the different models performing worse, and highlight any obvious biases present.

Figures 6.4 and 6.7 are the scatter plot and log-log scatter plot (respectively) of predicted vs observed values using Cox regression. Performance seems reasonable particularly for lifetime, with the (non-transformed) plot (figure 6.4 plot (a)) displaying the typical behaviour of points densely packed around the observed=predicted diagonal, with higher spread for larger values of both quantities (as it becomes increasingly difficult to predict with very high accuracy for longer lifetimes). The close-up look provided by the log-log plot (figure 6.7 plot (a)) shows a bias towards longer than observed predicted lifetimes for short lived players (relatively high spread clearly onto the upper semi-quadrant), but still acceptable behaviour. When studying the performance of in-game progression (figures 6.4 and 6.7 plot (b)), Cox regression shows its inability to predict levels below 35, and some bias towards higher predicted than observed values, but still a reasonable performance across many scales. The major problem displayed by this relatively simple parametric approach appears when considering playtime (figures 6.4 and 6.7 plot (c)): the model is only able to predict a very short range of expected playtimes, and even for players with observed values in that range, it has relatively bad performance with a clear bias towards underestimation.

Moving to the plots evaluating the RSF performance (figures 6.5 and 6.8), the same problems as with Cox regression are observed, but in a much less acute manner. Accuracy in lifetime predictions is improved, and the bias observed for short lifetimes virtually eliminated (compare plot (a) of figures 6.8 and 6.7). The inability to forecast low in-game progression is now limited to levels under 10 (figure 6.8 plot (b)). Still, the most problematic behaviour in terms of practical applications is the bad performance in playtime: random survival forest is capable of predicting for a larger range

of values, and to do so with less bias ((compare plot (c) of figures 6.8 and 6.7)), but the accuracy is still far from acceptable.

Finally, figures 6.6 and 6.9 can help assess the usefulness of conditional inference survival predictions. Interestingly, performance in lifetime and level is slightly degraded as compared to that of the other ensemble method considered (RSF), as made obvious by the larger spread around the diagonal in plots (a) and (b) of figures 6.6 and 6.9 (as compared to those in figures 6.5 and 6.8). When looking at playtime (plot (c) in figures 6.6 and 6.9), however, the situation is completely reversed: the performance of conditional inference survival ensembles is remarkably good at all scales, actually showing some bias towards overestimation (as opposed to the clear underestimation of the other two methods shown in plot (c) of figures 6.7 and 6.8).

Although the random survival forest yields slightly better lifetime and level predictions, it critically fails when predicting accumulated playtime to conversion. Only conditional inference ensembles are able to predict conversion after long accumulated playtime, and, as is clear from the log scatter plots in figure 6.9 (as compared to figures 6.8 and 6.7), both Cox hazards model and random survival forest have very obvious biases, systematically predicting significantly lower playtime than the actual outcomes.

In regards to level however, the random forest algorithm yields better predictions across all scales. Scatter plots also show that all models are incapable of predicting conversion when it happens in the first few levels. This however has very little practical relevance: progression through these levels is very quick, and happens almost instantly in terms of lifetime and playtime, so early detection of the potential of these users has very little added value. Random survival forests also give better predictions for lifetime, but mainly due to improved performance for short lifetimes, which are again a case of little real practical interest. Note also that all models are biased (although randoms forests less so) in that they tend to predict higher levels of conversion than those observed. This is also the case of conditional inference ensembles predicting accumulated playtime to conversion.

To sum up, all survival methods examined provide reasonable predictions, both of which players have paying potential, and of when this conversion will occur in terms of lifetime and level. Survival methods outperform the classical semi-parametric Cox regression, while adding churn as a competing risk slightly degrades the results. Although random survival forests tend to outperform conditional inference ensembles for level and lifetime, it is specially so when conversion happens early on, which are cases

of less practical relevance. On top of this, conditional inference ensembles are the only models to provide reasonable predictions on when conversion will occur in terms of accumulated playtime, hence this would be the preferred method to be used in production setups.

6.2 IMMINENT RISK OF PURCHASE CHURN

Once the issue of predicting conversion is settled, one can turn to trying to predict when players that indeed become PUs will stop purchasing (but will still be active in the game). Purchase churn modelling for individual players is going to be treated very much like login churn was in the last chapter (chapter 5). The modelling of purchase churn rate for the whole game was tackled in chapter 4.

Conditional inference forests can be used for binary classification of purchase churn, analogously as they can be used for (login) churn (see chapter 5 section 5.1). Such a binary classification would be a first step, detecting which of the current PUs of the game are in immediate risk of not spending any more. As was suggested for the case of login churn, these players could be targeted with specific particularly well suited content, or with discounts and promotions, in an attempt to prevent them from loosing their genuine PU status.

This method actually outputs a purchase churn probability per player (and not simply a predicted label), so the sensitivity to classify a player as in high risk of purchase churning can be fine tuned depending on the particular use intended for the predictions, and on whether it is more important to minimise false negatives or false positives (as was also described for regular churn in chapter 5 section 5.1).

As it is probably obvious by now, this is not the preferred approach in this thesis. The use of survival models to predict time, level and playtime to purchase churn will rather be explored, thus taking into account explicitly in the predictions how much each user is playing, and how skilful they are. This path will be explored in section 6.3, and how the inclusion or exclusion of zombies, churners and purchase churners from the training affects model performance in section 6.4. The latter will also assess how conditional inference classification is affected, reason why its use has been briefly described here. In section 6.4 table 6.4 the resulting AUC scores of VIP player purchase churn prediction (both including and excluding different types of churners) will be presented and discussed.

6.3 PREDICTING TIME, PLAYTIME AND LEVEL TO PURCHASE CHURN

Following the same reasoning and methodology proposed in chapter 5 section 5.2 for regular churn, conditional inference forests can be used to build survival purchase churn models to predict days after first login, in-game progression and cumulative playtime to purchase churn.

Figure 6.10 shows the Kaplan-Meier estimates of the purchase survival probability as a function of days since first login, level and playtime for VIP players. That is, it shows, for the period considered, the percentage of whales that have made at least one purchase in the previous 50 days.

These curves are very similarly shaped to the ones related to login survival plotted in figure 5.3, but with lower values of survival probability. This is indicating the main reason why whales stop purchasing is because they quit the game. Note that surviving as PU implies first surviving as player at all, and it is the remaining active trait which is dominant in explaining curve shape in both (survival as active player and as active spender) figures. Top spenders will typically continue spending throughout their whole history in the game, although a limited number of them will stop purchasing but continue playing. This accounts for the difference in values between the login curves of figure 5.3 and the purchasing curves of figure 6.10.

Actually, for relatively low level and playtime (plots (b) and (c) of figure 6.10), the purchase survival probability is very close even quantitatively to the survival login probability (plots (b) and (c) of figure 5.3). Only after level 200 or 1000 hours noticeable quantitative differences appear, with both curves saturating at levels above 25% in the case of login churn and well below that in the case of purchase churn. This indicates it is very rare for VIP players with less experience to stop purchasing without quitting the game. As they reach the highest progression and accumulated playtime, this is still rare but much more likely than at previous stages.

The case of lifetime is slightly different. Here, there is more variation in the shape of the curve that for progression and playtime (compare plot (a) of figure 6.10 to that of figure 5.3). The descend at the beginning (for shorter lifetimes) is steeper, while the value at which both curves saturate is closer (with one slightly above and the other slightly below 25%). This means players with heavy spending for a period of time after joining the game can indeed stop doing so later on while still playing (though probably with less frequency and intensity), while in terms of in-game progression and playtime this only happens towards the end. On the other hand, if they

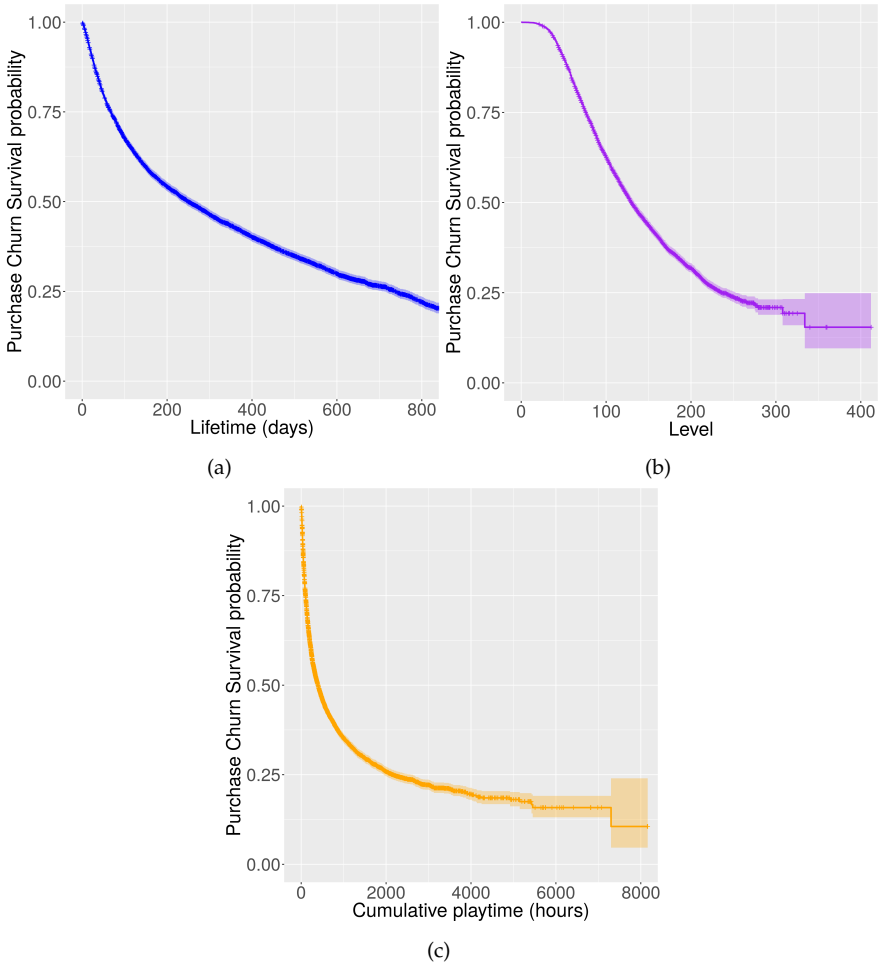


FIGURE 6.10: Kaplan–Meier estimates of purchase survival probability as a function of time since first login (a), game level (b) and accumulated playtime (c) for VIP players. Shaded areas represent 95% confidence intervals. Elaboration using data from AoI.

keep purchasing long enough (in days after first login), as opposed to what happens in level and playtime, they will most likely keep spending until they quit the game for good.

6.4 IMPACT OF ZOMBIES AND RESURRECTED

The exercise carried out in chapter 5 section 5.3 for churn, was repeated for purchase churn, using the same dataset, models (conditional inference forests and conditional inference survival ensembles, see chapter 2 sections 2.3 and 2.4), and verification metrics. Feature selection was carried out independently. The reader is encouraged to revisit chapter 5 section 5.3 for a reminder on the dataset characteristics and model specification.

Figure 6.11 shows the Kaplan-Meier estimates of the purchase survival probability as a function of days since first login, level and playtime for zombies, resurrected, purchase churners and normal (all the rest) VIP players. It is basically figure 6.10 stratified by type of churner. As expected given that all VIP login churners are also purchase churners (which is not necessarily true the other way round), these curves are steeper than the login purchase ones in figure 5.8.

For level and playtime (plots (b) and (c) of figure 6.11), the behaviour is qualitatively very similar to that of churn (plots (b) and (c) of figure 5.8), except maybe for very high values of both, where there is a large uncertainty for purchase churners that makes the situation difficult to assess. Namely, zombies and resurrected show the fastest descent in probability of still being active in both cases, with both groups displaying similar behaviour both qualitatively and quantitatively, with resurrected having a slightly larger purchase survival probability for all levels and playtime values. On the other hand, purchase resurrected players have the higher purchase survival probability of all groups (at least for all levels below 200 and playtimes below 5000 hours). Normal player behaviour lies pretty much in the middle: they do not disengage as rapidly as zombies or resurrected VIP players, but do so markedly faster than purchase resurrected ones, at least for most values. After level 200 and 5000 hours of playtime, the percentage of normal active players remain constant (at 25-30% in both cases).

In lifetime (plot (a) of figure 6.11) too, zombies are the more rapidly and purchase resurrected the more slowly disengaging players. Here, however, resurrected players don't show a quantitatively similar behaviour to zombies, and even have a higher purchase survival probability than normal players for the first approximately 100 days of lifetime. Again, this is the same pattern found in login churn (plot (a) of figure 5.8), although for a much shorter period and with very little difference between both groups. In purchase churn with lifetime, the purchase survival probability of purchase churners does not stabilise after some period, and continues decreasing at

approximately the same pace (which was the case of the resurrected for login churn). As the normal player's curve does become approximately constant after 600 days, their probability of still being genuine PUs is very close to that of purchase resurrected for lifetimes of more than 800 days.

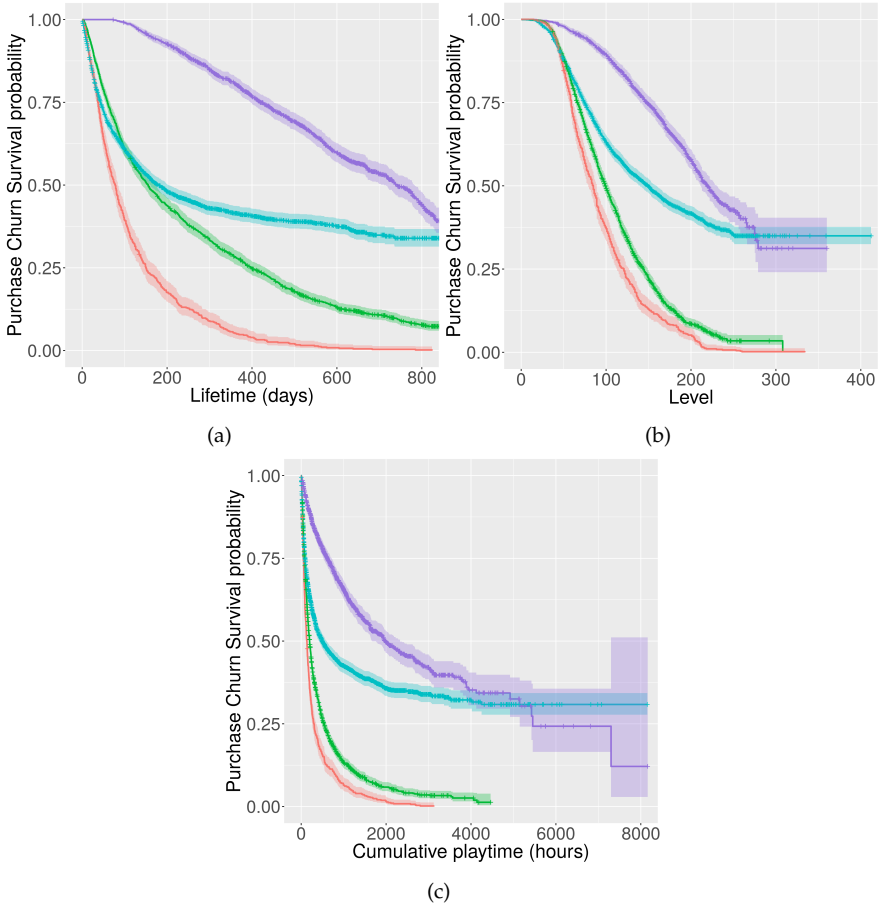


FIGURE 6.11: Kaplan–Meier estimates of purchase survival probability as a function of time since first login (a), game level (b) and cumulative playtime (c) VIP players. Curves are stratified by churner type: *normal* (blue), *zombie* (red), *resurrected* (green) and *purchase resurrected* (purple) players. Shaded areas represent 95% confidence intervals. Elaboration using data from AoI. The images have previously appeared in [132].

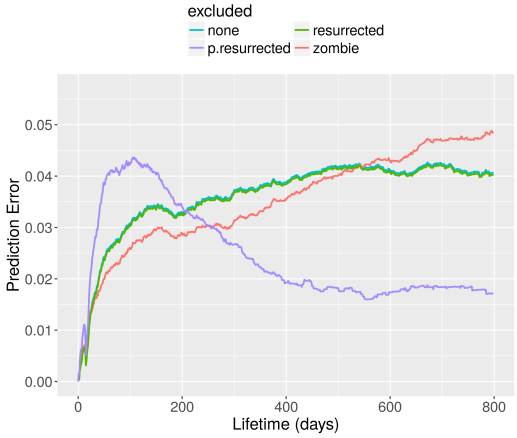
CHURN	Binary models (AUC)	Survival models (IBS)		
excluding from training		lifetime	level	playtime
none	0.69	0.070	0.080	0.077
zombie	0.69	0.055	0.067	0.086
resurrected	0.68	0.070	0.080	0.080
p. resurrected	0.72	0.065	0.076	0.062
zombie, resurrected	0.69	0.055	**0.057**	0.086
zombie, p. resurrected	0.72	**0.053**	0.067	**0.050**
resurrected, p. resurrected	0.73	0.065	0.068	0.057
zombie, resurrected, p.resurrected	0.73	**0.053**	**0.056**	**0.051**

TABLE 6.4: Area under the curve (AUC) for binary model and the integrated Brier score (IBS) for survival model (in terms of lifetime, level and cumulative playtime) for the different situations with regard to the training sample: including all users (*none*) vs. excluding zombie, resurrected or purchase resurrected players (or combinations of them). The best results for each model and variable are highlighted in bold.

Prediction accuracy results are summarised in table 6.4, with figures 6.12, 6.13 and 6.14 comparing prediction error curves on the different variables (lifetime, level and playtime respectively). As was the case for login churn (discussed in chapter 5 section 5.3), survival model results are much more sensitive (than classification model ones) to the removal or not of one or more of the special types of players in regards to churn. In this case however, the exclusion of purchase churners does have a (clear albeit small) positive impact in the binary model's performance.

When considering overall accuracy of survival model purchase churn predictions (see table 6.4), the single group whose removal improves performance in the three variables is purchase resurrected. Excluding login resurrected alone has no overall effect in lifetime and level, but slightly degrades playtime accuracy, while removing only zombies from the training improves performance in lifetime and level, but significantly and negatively affecting accuracy in playtime.

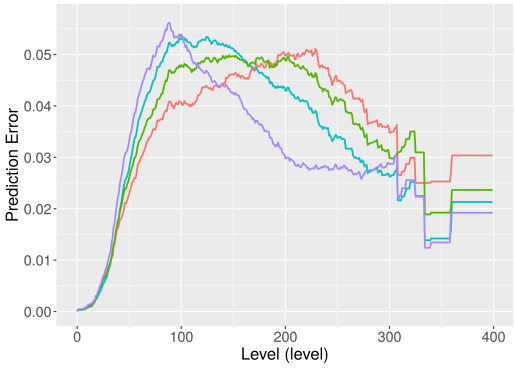
Looking at plot (a) of figure 6.12 (for lifetime), figure 6.13 (for level) and figure 6.14 (for playtime) to understand what it is happening, it is obvious that the removal of purchase churners very significantly reduces errors at medium to large values of all variables. For short values, however,



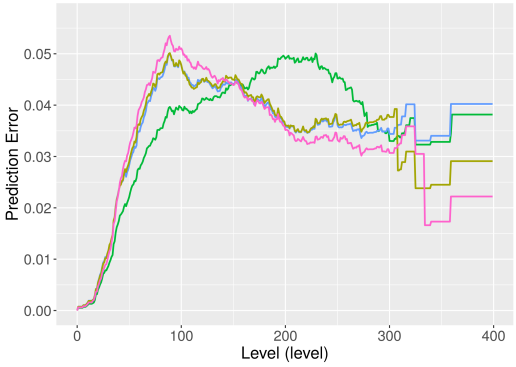
(a)

(b)

FIGURE 6.12: Prediction error curves for AoI *purchase* churn as a function of lifetime. The different lines represent model runs excluding zombies (red), resurrected (green) or purchase resurrected (purple) players (plot (a)) and combinations thereof (plot (b)) from the training sample. Combinations represented in plot (b) are: (i) resurrected and purchase resurrected (pink), (ii) zombies and purchase resurrected (brown), (iii) zombies and resurrected (green), and zombies, resurrected and purchase resurrected (blue). Elaboration using data from AoI. The images have previously appeared in [132].



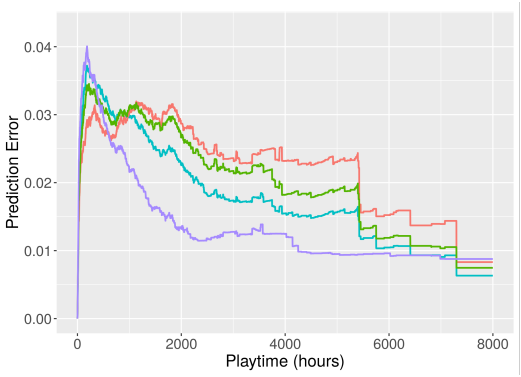
(a)



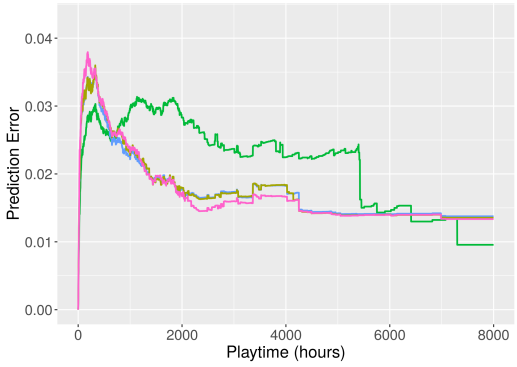
(b)

FIGURE 6.13: Prediction error curves *purchase* churn as a function of game level. The different lines represent model runs excluding zombies (red), resurrected (green) or purchase resurrected (purple) players (plot (a)) and combinations thereof (plot (b)) from the training sample. Combinations represented in plot (b) are: (i) resurrected and purchase resurrected (pink), (ii) zombies and purchase resurrected (brown), (iii) zombies and resurrected (green), and zombies, resurrected and purchase resurrected (blue). Elaboration using data from AoI. The images have previously appeared in [132].

the impact is reversed. While only short playtimes (below 250 hours) and in-game progressions are affected (below level 100), and with only a slight degradation in performance (particularly in playtime), the negative impact



(a)



(b)

FIGURE 6.14: Prediction error curves *purchase* churn as a function of playtime. The different lines represent model runs excluding zombies (red), resurrected (green) or purchase resurrected (purple) players (plot (a)) and combinations thereof (plot (b)) from the training sample. Combinations represented in plot (b) are: (i) resurrected and purchase resurrected (pink), (ii) zombies and purchase resurrected (brown), (iii) zombies and resurrected (green), and zombies, resurrected and purchase resurrected (blue). Elaboration using data from AoI. The images have previously appeared in [132].

of their exclusion is much more obvious in lifetime. There, performance is significantly worse for lifetimes of up to 200 days.

This is a general trend observed in all plots: improving accuracy at some scales by removing one of the types of churners, always comes at the cost

of degrading it for other ranges. If having to choose excluding only one of the groups (or none), however, unless having a particular interest in short lifetimes, the clear winner (at nearly all scales) for all variables is the elimination of purchase churners from the training set.

The effect of excluding different combinations of types of churners at different scales of the different variables can be studied using plot (b) of figure 6.12 (for lifetime), figure 6.13 (for level) and figure 6.14 (for playtime). All combinations excluding purchase resurrected display the same pattern, both qualitatively and quantitatively, which is qualitatively very close to that of excluding only purchase resurrected, but with slightly smaller errors for small values, and slightly larger errors for large values of all variables. The only remaining case, that of removing zombies and resurrected, but leaving purchase resurrected in, resembles very closely the case of excluding only zombies.

To sum up, when considering accuracy of survival model purchase churn predictions, removing zombies and/or resurrected players from the training has a similar qualitative impact as their removal from login churn models (as discussed in chapter 5 section 5.3). They improve predictions for all values of lifetime and for small values of level and playtime, while degrading them for larger values of playtime and in-game progression. Learning of the models is heavily influenced by purchase resurrected players, whose removal has an overall positive effect in the predictions in all three variables. Interestingly, this is due to a stark error decrease for longer values of the variables, while their exclusion actually degrades the forecasts for smaller values, specially so when considering days since first login. This could point towards the need of a more restrictive definition of purchase resurrection (along the lines of login resurrection), where players would be considered purchase resurrected after periods without any purchase that should be much longer than simply the purchase churn definition. Overall, the preferred setup would be to exclude zombies and purchase resurrected, with results nearly independent on whether login resurrected are included or not in the training.

This study therefore suggests, that if interested in overall accuracy and particularly so for extended lifetime, progression or playtime, the preferred option would be to exclude zombies, resurrected and purchase resurrected from the training. If the focus, however, is on lifetimes below 300 days, progressions below level 150, or playtimes below 500 hours, purchase resurrected players are better left in. In a production setup then, and taking into account the results for churn discussed in chapter 5 section 5.3, this

study would suggest to use in the training of the models the exact same sample as for login churn (all players except zombies and resurrected), but potentially excluding too all purchase churners.

6.5 PREDICTING PLAYER LIFETIME VALUE

The concept of *lifetime value (LTV)* (or *customer lifetime value (CLV)*) has been introduced in chapter 2 section 2.5, and is used to describe the expected total revenue coming from each customer. Having accurate estimates of it is of particular interest to any service not bounded by a contract, in which users can decide when and how many purchases to do. One of the first accounts of the term appears in [253], and the term became increasingly popular, as did the number of publications devoted to its study, in the context of marketing at the end of the 90s, as an increasing amount of computational power and collected data were available for its modelling [25, 81, 145]. It can be used to design specific marketing and promotion plans aimed at retention and expenditure maximisation, can help with the early detection of high value users [91], and is useful in predicting total expected revenue.

Free-to-play games with in-app purchases are the exact setting where having good LTV predictions is desirable. Basic approaches to its prediction making use of previous data include [68, 183]. There is a relatively limited number of papers exploring the use of ML methods. In [256], the authors combine a binary classification approach to predict whether players will spend or not in the future, with a regression model to estimate the number of purchases. The use of extreme gradient boosting both for classification and regression, together with the impact of social features in the problem, is explored in [76]. In [285], several ML methods are combined with the synthetic minority oversampling technique (SMOTE) [50]. The same approach is also used in [257] to generate individual player predictions of expenditure over the next year, in what is probably the only other work to this date using deep learning architectures as one of the proposed methods (in particular deep perceptron multilayer networks) to predict player outlay.

ML methods fall into the broad category of what can be defined as *predictive approaches* to LTV forecasting. These combine all past and present information available to try and predict the future behaviour of customers, and crucially, they intend (be it more implicitly or explicitly) to account for user purchasing behaviour changes. Examples of ML computed LTV outside video games are numerous, and these methodologies are becoming increasingly popular due to their success. See, for example, [48] (which uses

random forests in the context of an online fashion retailer), or [275] (that uses deep reinforcement learning to predict total donations for members of a potential donor mailing list).

Other models used traditionally in LTV estimation fall in the broad category of *historical approaches*, as opposed to predictive. The main distinction between both is that these base their best guess for each customer only in their previous purchasing history, i.e., they do not allow for changes in individual behaviour (other than ending the commercial relation with the service altogether). Such are the *RFM (recency, frequency, monetary value) models* [89], in which predictions for each user are based solely in when the last time they purchased was (recency), how often they have purchased in the historic data available (frequency), and how large their expenditure has been. Probabilistic RFM models assume customers will repeat the same purchasing patterns until they churn, reason why they also receive the name of "buy till you day" (BTYD) models. The churn prediction involved in these models can also be carried out in parametric probabilistic terms, as was proposed as early as in [247].

Examples of RFM models are those referred to as parametric models and described in chapter 2 section 2.5. The probability distribution considered to describe the problem is parameterised, with parameters to be estimated or learned form the data. A few of these models were used as benchmark in this study. These were compared to results produced with two different DNN architectures: the multilayer perceptron and a convolutional neural network (described in chapter 2 section 2.6).

6.5.1 *Dataset and model definitions*

From the data available for AoI (described in chapter 1 section 1.2), only the period from 2016-05-01 to 2017-05-01 was used for evaluation. This means, predictive accuracy was assessed using the 2505 players that were active, spent money and left the game during that period. This choice was made in order to allow enough previous data to be available to have the transaction history needed for the RFM methods. Purchasing history together with other past playing data, was used as to build features for the DNN architectures tested. As described in chapter 3 section 3.2, taking into account all modelling corresponds to PUs, the churn definition used is 9 days. After not login in for that duration, the player is considered to have quit the game, their transaction history finalised, and the final value of their LTV known.

Regarding the parametric models (see chapter 2 section 2.5), the Pareto/NBD, BG/NBD, BG/CNBD-k and MBG/CNBD-k models were tried. Two different approaches were explored as submodels for the monetary value of the predicted transactions: the use of gamma distributions, and simply taking the average value per transaction in the past for each player.

Parametric models need a fixed prediction horizon, which was set to one year (that is also approximately the period for which data is available for study). These models only use the player transaction history. Figure 6.15 shows, for a selection of players, the purchasing history during the training period (in black) and evaluation period (in grey). These kind of figures can help to quickly assess if the hypothesis of past purchasing patterns being repeated into the future appears to roughly hold for most players or not.

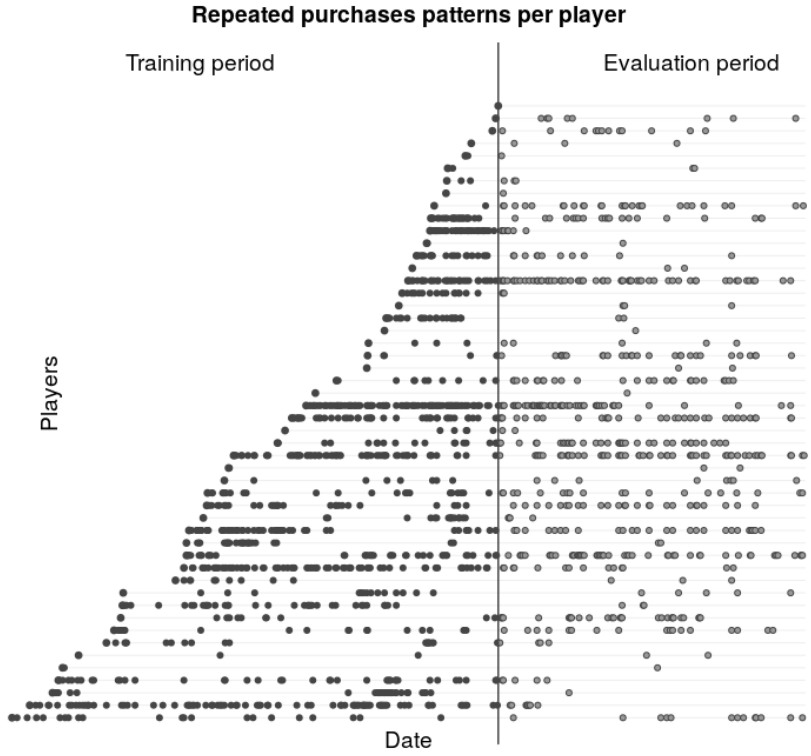


FIGURE 6.15: Purchasing patterns per player for a sample paying users for the training period and the evaluation period (test part). Elaboration using data from AoI. The image has previously appeared in [51].

Figure 6.16 shows the box plot of the average purchase value per player depending on the number of repeated purchases. While the average value is close in all cases, it tends to slightly increase with the number of repeated purchases, as does, even more markedly so, the spread.

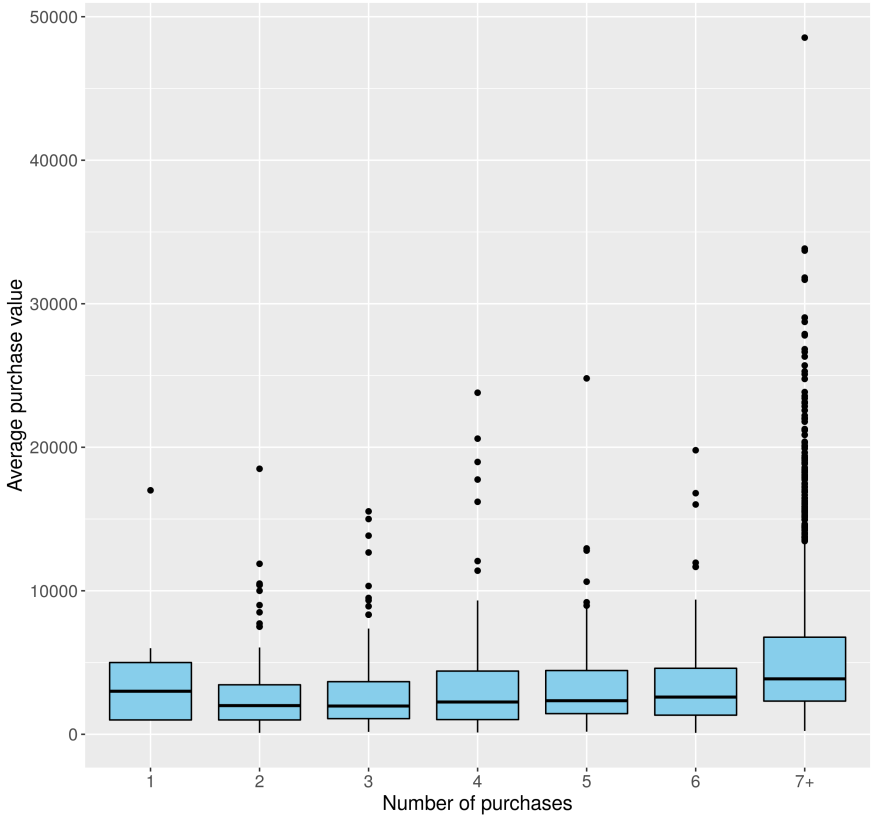


FIGURE 6.16: Box plot of the average purchase value per number of repeated purchases per all paying users. Elaboration using data from AoI. The image has previously appeared in [51].

Two deep learning architectures were considered: multilayer perceptron and convolutional neural network. The DMLP (see chapter 2 section 2.6.1) is the most simple example of a DNN, and has produced good results for a wide range of problems, so it is always useful to include it as a benchmark when comparing to other more sophisticated architectures. CNNs (see chapter 2 section 2.6.2) are potentially of particular interest in this case, as

their nature allows them to learn from correlations in the input data. This makes it possible for them to work directly with input features that are time series (which is the case of most relevant video game data), not requiring feature engineering as an initial step. Another candidate architecture that has been used to predict LTV, and that will be described in chapter 7 (but that was not used in this particular study), is the LSTM (see chapter 2 section 2.6.3), also because their design is specifically well suited to deal with data of time series nature.

The DMLP used five fully connected layers: an input layer with 203 neurons (one per input feature), followed by three hidden layers with 300, 200 and 100 nodes respectively, and finally the output layer with a single node (corresponding to the LTV). The specific CNN architecture used is depicted in figure 6.17: an input layer, followed by a convolutional layer with 32 filters of size 7, followed by a max-pooling layer to prevent overfitting [245], followed by two additional convolutional layers with 16 and 1 filters of size 3 and 1 respectively, a flatten layer, three fully connected layers with 300, 150 and 60 nodes, and finally the single output node. In both cases, the training algorithm chosen was adaptive stochastic gradient descent (ADAM) (see chapter 2 section 2.6.5), and Xavier initialisation (see chapter 2 section 2.6.4) was used. The activation functions used were sigmoids.

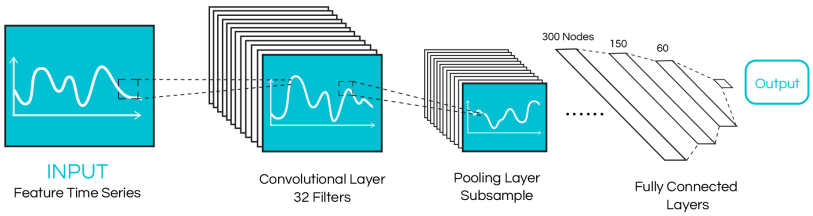


FIGURE 6.17: Structure of the convolutional neural network used to model LTV. The image has previously appeared in [51].

Player behaviour logs were used to build player time series of daily logins, playtime, in-game progress, purchases ... Generic enough data is preferred as to ensure the same model definition can be directly translated to other titles. These time series were used directly as input features for the CNN. For the DMLP, similarly to what has been described in the case of decision and survival ensembles, these time series were further processed using statistics to arrive at the features to be used (for example, average level-ups between purchases, maximum and minimum daily playtime in

the last month, number of purchases in the first week, average number of days played consecutively in the whole history...). Data from players that had churned before the beginning of the evaluation period (2016-05-01) was used to train the models. At every epoch, the models were trained using 80% of these players, and predictions performed and validated for the remaining 20%. The learning process comes to an end once the predictive accuracy has not increased for 20 epochs, as this method of determining the number of epochs to use (*early stopping*) is known to prevent overfitting [229]. The network weights are then set to those that produce the lowest errors, and are used to predict for the 2505 players that make up the evaluation dataset.

6.5.2 *LTV prediction results*

As metrics to assess the goodness of the predictions, the RMSLE, NRMSE, SMAPE and percentage error were selected (see chapter 2 section 2.9 for definitions and descriptions of these and other metrics). Results are summarised in tables 6.5 to 6.7. Tables 6.5 and 6.6 collect all the metrics for all the models for the training and evaluation sets respectively, that is the validation metrics as computed when comparing predicted and observed LTV for users used in the training (80% of those available in the dataset) and those that were not (remaining 20%) The model name for the RFM methods used codifies the model used to predict the number of purcahses, and after the + sign, how the value of the remaining transactions is computed. Table 6.7 compares the percentage error the models make when predicting for all PUs as compared to only the top spenders (defined here as the 20% of the PUs that spent more during the year considered for model evaluation).

Note that depending on the metric defined and the accuracy standards required, all of these could fall into the been generally good, or being generally bad categories. This becomes clear when comparing, for example, both percentage errors. While the average deviation of predictions is, for all models, under 10% of the largest observed LTV (error %), SMAPES are much higher in all cases. Note that while the SMAPE does not have a straight forward interpretation, it is always bounded between 0 and 200%, which indicates a clear room for improvement in all cases.

What is out of question is that DNNs are performing better than any of the traditional parametric models, and they do so consistently across all metrics. The use of NRMSE allows for comparison to other cases with different scales, so it can be checked that in particular the DNN performance is similar to that discussed in [257] for high-value players. Furthermore,

Model	RMSLE	NRMSE	SMAPE	Error %
Pareto/NBD + average	9.42	1.89	95.87%	6.20%
Pareto/NBD + gamma	9.43	1.91	96.29%	6.24%
MGB/CNBD-k + average	3.41	1.72	75.44%	5.52%
MGB/CNBD-k + gamma	3.55	1.77	78.58%	5.71%
BG/CNBD-k + average	4.13	1.69	76.82%	5.43%
BG/CNBD-k + gamma	4.24	1.74	79.83%	5.63%
BG/NBD + average	9.48	1.89	96.35%	5.43%
BG/NBD + gamma	9.49	1.92	96.67%	5.63%
DMLP	1.78	1.07	75.08%	3.90%
CNN	1.74	1.11	72.75%	3.90%

TABLE 6.5: Error measures for the LTV training

Model	RMSLE	NRMSE	SMAPE	Error %
Pareto/NBD + average	9.35	1.88	95.65%	8.96%
Pareto/NBD + gamma	9.37	1.88	96.35%	9.01%
MGB/CNBD-K + average	3.46	1.68	75.53%	7.88%
MGB/CNBD-K + gamma	3.61	1.73	79.67%	8.08%
BG/CNBD-k + average	4.41	1.65	76.22%	7.85%
BG/CNBD-k + gamma	4.24	1.71	79.72%	8.06%
BG/NBD + average	9.37	1.88	96.06%	8.96%
BG/NBD + gamma	9.39	1.88	96.80%	9.03%
DMLP	1.82	1.12	72.99%	5.82%
CNN	1.84	1.05	73.76%	5.72%

TABLE 6.6: Error measures for the LTV prediction

deep learning are more accurate than the RFM models across all scales, as indicated by the particularly significant improvement of the RMSLE metric. It is also interesting to note that the DMLP and the CNN are close in regards to accuracy. This has very likely a lot to do with the fact that both

Model	All PU Error %	Top spenders Error %
Pareto/NBD + average	8.96%	33.35%
Pareto/NBD + gamma	9.01%	33.39%
MGB/CNBD-K + average	7.88%	29.61%
MGB/CNBD-K + gamma	8.08%	30.54%
BG/CNBD-k + average	7.85%	29.46%
BG/CNBD-k + gamma	8.06%	30.45%
BG/NBD + average	8.96%	33.34%
BG/NBD + gamma	9.03%	33.38%
DMLP	5.82%	15.76%
CNN	5.72%	15.64%

TABLE 6.7: Prediction error compared for all PUs vs top spenders only

used very similar input information on the players (even if with different amount of processing).

From the parametric models, those considering erlang-distributed number of purchases (and thus more regular transaction frequencies) clearly outperform those that assume randomness. The distribution chosen for the dropout process does not appear to have any significant impact. The introduction of a gamma-gamma submodel for the economic value predicted for each transaction shows practically the same accuracy as relying on the past average, so the latter would be the preferred method to be used in production, as the introduction of this additional complexity is not justified by the results.

Not only do both DNNs clearly outperform parametric methods, but their relative accuracy becomes larger with increasing player expenditure. All parametric models significantly and systematically underestimate top spender LTV, which makes them particularly ill suited for the endeavour of VIP player detection. This is highlighted by the results shown in table 6.7, with both DNNs halving the errors of the parametric approaches. Another shortcoming of the classical RFM models is that they also predict no future purchases for a significant number of players that do go on in reality to continue spending money.

It is however hardly a surprise that deep learning techniques show a much higher accuracy than simple probabilistic models. Not only are the

models and their learning algorithms more sophisticated, but more importantly, they are able to use much more information about the players and their behaviour to learn how these relate to their purchasing engagement. Moreover, because of the nature of ANNs, the accuracy of these models is expected to continue to improve when more data is used (either a longer historic period used for training, or for games with more players), while other statistical or ML models' performance saturates rapidly after a certain point. The ease with which CNNs can be used with nearly raw data, hardly requiring any feature engineering, would also become more important for larger histories and/or games, probably pointing to that architecture as the prefer candidate for productionalisation of LTV predictions for large games.

6.6 SUMMARY AND CONCLUSIONS

Purchasing engagement can be studied using similar approaches and methodologies as those used for regular play engagement. As was the case for churn, survival models are an interesting tool when addressing purchase churn, and they can be used to describe each player by a set of predictions in time, level and playtime. Purchase churn is also markedly different for different types of players, in particular for the different types of churn profiles that have been discussed: zombies, resurrected, purchase resurrected, and the rest. As was the case for login churn, players displaying very little activity (while remaining officially active) appear to introduce noise, and models perform better when they are excluded from the training set. The inclusion or not of those that return to the game after a long period of inactivity have no significant impact, while removing purchase churners positively impacts accuracy in all variables (in both cases, contrary to the case of regular churn).

Besides studying purchase churn as a proxy to purchasing engagement, survival models can also be successfully used to detect paying potential and to predict after how many days, hours played, and at which level, will every user make their first transaction (if this will happen at all). Survival ensembles outperform the parametric Cox hazards model, and only conditional inference survival ensembles (as compared to RSF) show a decent performance in playtime.

The total expected outlay gives an additional dimension to characterise purchasing engagement. With regards to LTV prediction, two deep learning architectures, CNN and DMLP were proposed, and their accuracy compared to that of some of the most popular conventional parametric approaches.

CNN emerges as the best option. While both DNN architectures have similar performances, CNN can work with raw sequential data directly, which could be potentially critical if it is to be used in near real time for very large games. Not only does deep learning achieve higher accuracy than all traditional probabilistic models, but this is particularly so when considering top spenders, which makes them specially well suited for VIP player detection. An additional DNN architecture which also indicated to deal with sequential data -the LSTM (see chapter 2 section 2.6.3)- was also tested (after this study was finished) to predict LTV for VIP players only, and was found to perform comparably to the CNN and DMLP. While those results have not been presented here because the dataset used was different (the DNN was trained with top spenders only), the LSTM predictions are used for player profiling in next chapter (chapter 7).

7

PROFILING PLAYERS USING PREDICTIONS

All creative people want to do the unexpected.
— Heidy Lamarr

Player profiling consists essentially in grouping players according to their behaviour and characteristics, as was already discussed in chapter 3. In an increasingly competitive market with increasingly diverse demographics playing games, it is crucial to be able to profile players appropriately, in order to meaningfully target them and cater the game to their needs. Much of the past player profiling has often focused in expenditure, as the goal was mainly to pinpoint high value users. However, the use of additional behavioural information can help characterise players engagement, play style, interests, skills and motivations, thus greatly enriching the picture.

The last chapters (chapters 5 and 6) have been devoted precisely to the characterisation of user's play and purchasing engagement with different individual player behavioural predictions. This chapters explores how some of these -those of expected LTV, and of days, playtime and level to churn- can be used and combined to provide insights on the player population and game dynamics, and to produce meaningful player profiles.

The use of predictions (rather than actual values) should allow to profile players since their very first steps in the game, making use of all present and also past information (even of players who quit the game long ago). The rest of the chapter focuses on whales, as the predictions described in previous chapters have been fine tuned to perform particularly well for this group of players. It would however be straight forward to generalise this methodology to apply it to other types of players, or to use additional predictions to better characterise their behaviour.

All content of this chapter is an original contribution of this thesis. The approach was first presented in [72].

7.1 METHODOLOGY

Each player was characterised by a collection of predictions. Survival curves in lifetime, in-game progression and cumulative playtime are used to de-

scribe each player's engagement, while their purchasing behaviour is summarised in their expected LTV, computed using deep learning.

The LTV predictions used in this chapter were computed in the same spirit albeit with a different architecture as those discussed in last chapter (chapter 6 section 6.5). Time series of player actions and of transaction histories were built and used as input of two LSTM (see chapter 2 section 2.6.3) layers respectively, considering only VIP players. The outputs of these two layers were concatenated and forwarded to a fully connected layer, that was followed by three additional fully connected layers, with a final single output node yielding the LTV prediction. This architecture is schematically shown in figure 7.1.

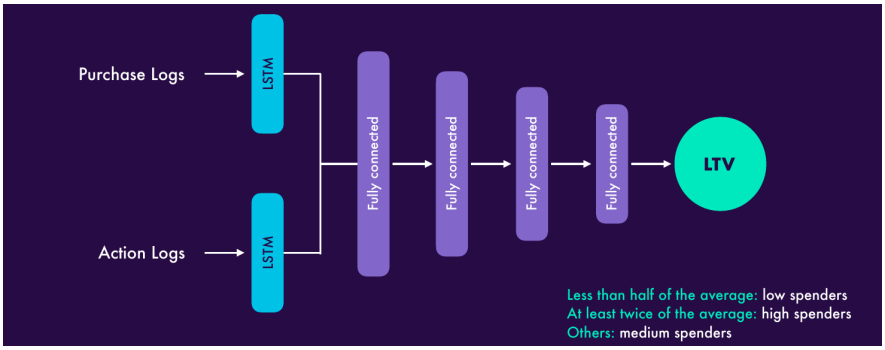


FIGURE 7.1: Schematic representation of the LSTM architecture used to predict VIP player LTV and classify them into *low, medium* or *high* expected LTV groups.

LTV predictions will be used to classify players into three groups: *high*, *medium* and *low* spending groups. Different ways of grouping players can be useful for different purposes. In what follows, players with expected outlay less than half the expected average are considered low spenders, those with expected outlay at least twice the expected average are the high spenders, and the rest are medium spenders.

Conditional inference survival ensembles are used to generate predictions of days, playtime and level to churn, as was described in chapter 5 section 5.2. The grouping is not done only in terms of the predicted value (as given by the median of the survival curve), but making use of the different quartiles of each curve, to classify them as corresponding to *short*, *medium* or *long* lifespans. Furthermore, players whose survival probability remains always above 50% (i.e., for which the median value and hence a numeric

prediction does not exist), are considered to be *loyal* in that variable. Players for which the lowest probability value of the survival curve is greater than or equal to 25% but smaller than 50%, and with a median value greater than or equal to the population's average have *long* lifespans. There are three different ways a player can be classified as having a *medium* lifespan: (i) if the lowest value of their survival probability curve is between 25% and 50%, but the median value is below the population's average; (ii) if the lowest value of their survival curve is between 0 and 25%, whenever the value at which this 25% probability is reached is greater than or equal to the population's average; or (iii) when 0% survival probability is attained, provided the median value is greater or equal than the average value computed only taking players into account that have survival probability always above zero. The rest are *short* lived players. These classification criteria are summarised in figure 7.2.

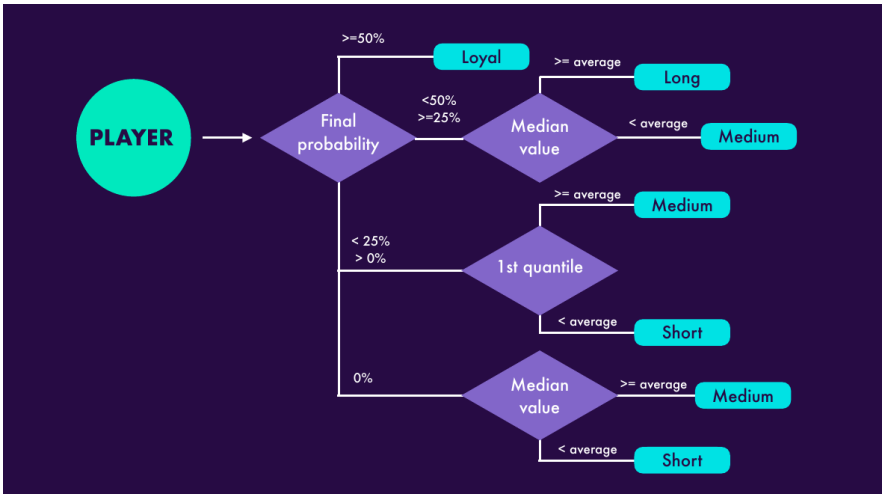


FIGURE 7.2: Schematic representation of the classification followed to assign each player and variable (i.e. each survival curve) to one of the various lifespan groups (*short*, *medium*, *long* and *loyal*).

This exercise was carried out considering only AoI VIP players (check chapter 3 for details on this and other definitions), training the models with data from October 2014 to April 2017 (comprising 3265 whales in total). Predictions were run (and their relationships plotted and discussed in some detail in what follows) for the 1771 VIP players that remained active as of

May 1, 2017, i.e., those whales that have connected to the game at least once in the previous 9 days.

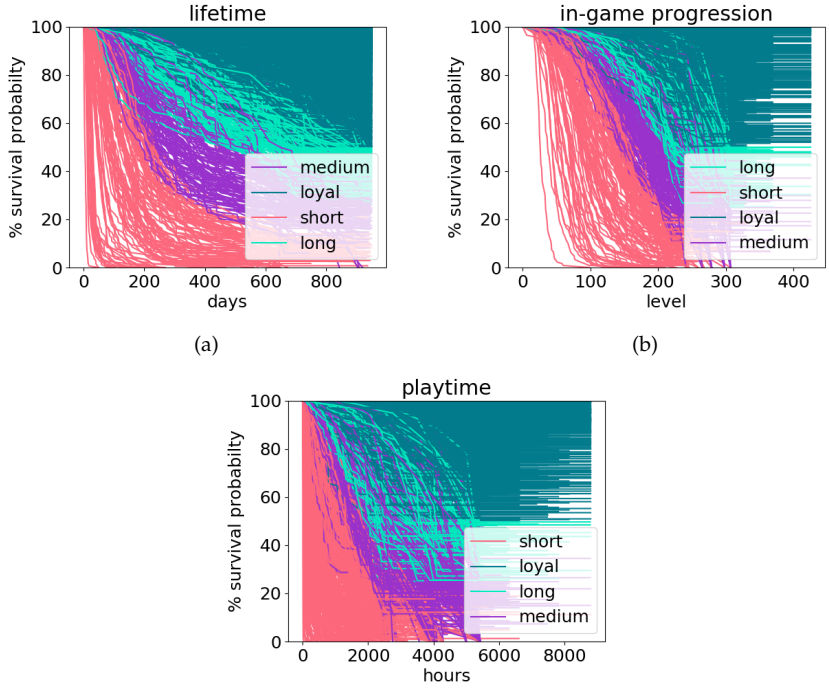


FIGURE 7.3: Survival curves for VIP players in terms of lifetime (days since first login; (a)), in-game progression (game level reached; (b)) and accumulated playtime (hours played; (c)). Colours distinguish the various lifespan groups (*short*, *medium*, *long* and *loyal*) for the corresponding variable. Own elaboration using AoI predictions. The images have previously appeared in [72].

Of course other groupings are possible and may be more useful depending on the specific needs of the profiling. However, this grouping seems to work well as can be seen in Figure 7.3, which shows the survival curves for all the players considered in terms of lifetime, in-game progression and playtime, with colours distinguishing the different groups. That the grouping also works can also be seen in Figure 7.4 which shows histograms corresponding to median vales (when they exist, i.e., for non-loyal players)

for those three variables, and also of the predicted LTV. Again, colours highlight the different groups. The histograms show predictions in the different variables are distributed very differently: approximately uniformly in lifetime, monotonously decreasing in LTV, skewed (i.e., asymmetric with respect to the mean) to larger values in level, and to smaller in playtime.

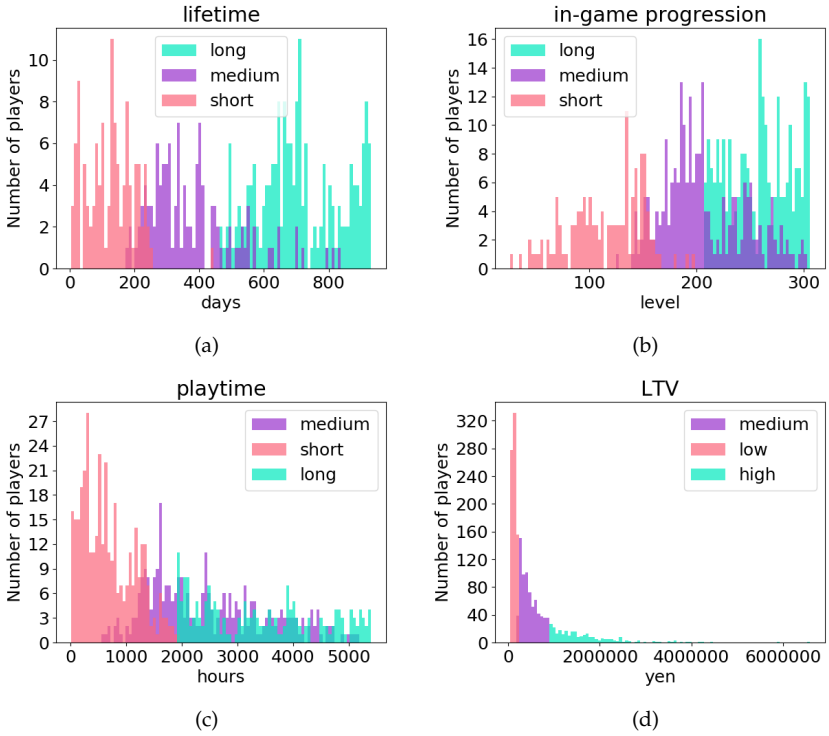


FIGURE 7.4: Histograms of the predicted lifetime (days since first login; (a)), in-game progression (game level reached; (b)), playtime (hours played; (c)) and LTV (outlay in local currency; (d)) for VIP players. Players are classified as described in the text, with groups shown in different colours. All players except those labelled as *loyal* (for whom the median value of the survival curve does not exist) are shown. Own elaboration using AoI predictions. The images have previously appeared in [72].

7.2 EXPLORING RELATIONS BETWEEN PREDICTIONS

Grouping players according to predictions in each of the variables can be of course useful in itself: by using predictions, players of one type or another can be spotted early on and targeted appropriately. Furthermore, the use of probability curves in the case of survival modelling, used as a proxy of login engagement in this case, also contributes to a richer picture. Only a way of grouping players in terms of their survival curves is discussed here, but an infinite number is possible, and can take into account more subtle information than what the most likely value is. One could for example, for a particular exercise, group players in terms of their probability of still being active after a fixed period of time. Or make a difference between players that will disengage gradually but constantly (with an approximately stable decline or curve's slope throughout the predicted period), to those that will have a nearly constant probability of survival for some period, that will then drop abruptly.

An even richer landscape arises when the relation between predictions between two or more of the variables for each player is also taken into account. This can provide a lot of insights on the game dynamics and the different types of players present. For example, in Figure 7.5 the relation between predicted lifetime (in days) and playtime (in hours) is explored. Additional information about predicted LTV is added to both plots as the area (each point represents a player that has lifetime and playtime predictions, i.e., players not loyal on those variables), and to the top plot also as colour (which represents the LTV grouping of the player). The bottom plot differs from the top one in that the colour codes belonging to predicted level grouping instead.

As expected, Figure 7.5 shows how the spread in playtime increases as predicted lifetime does: periods of inactivity and difference in play frequency and typical session length between players, account for increasingly stark differences in accumulated playtime as time passes. The spread increases in an approximately linear manner, with most players lying on the region corresponding to an average daily play of 20 minutes to 5 hours.

It is also obvious from both plots, that, although large LTV predictions tend to belong to players with longer predicted lifetime, it is possible to find players with a relatively large LTV prediction for low predicted lifetime and/or playtime. Similarly, players with low LTV abound across all scales.

In regards to expected in-game progression (colour coded in the bottom plot), there is an obvious and expected correlation between playtime and

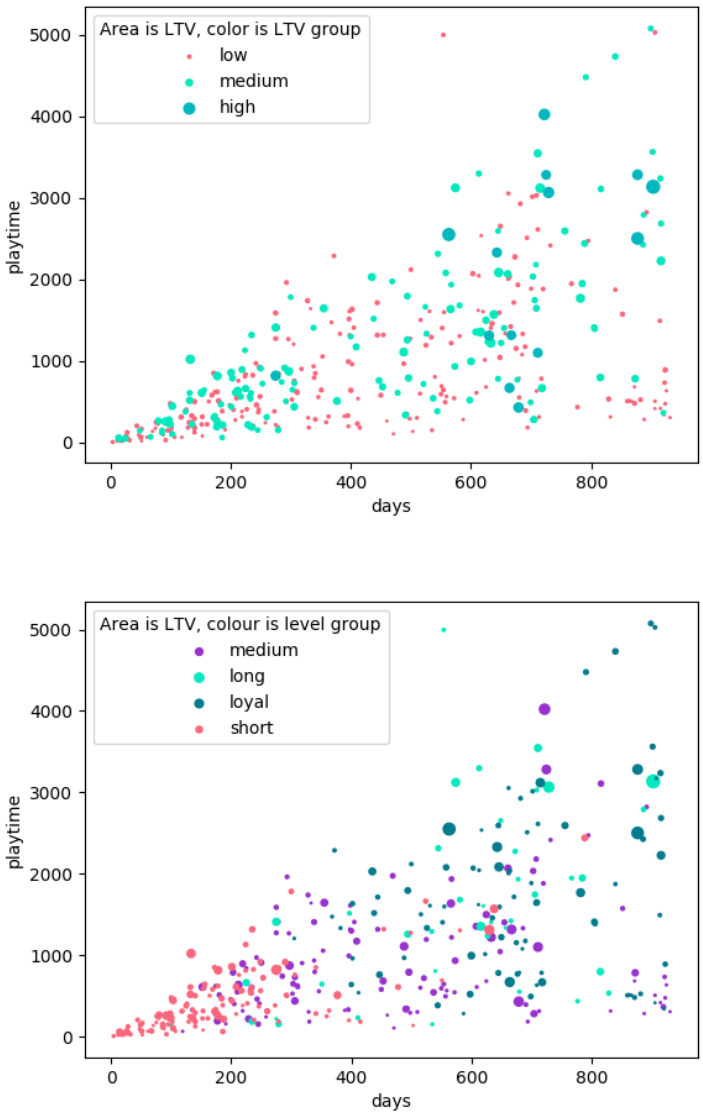


FIGURE 7.5: Playtime (in hours) versus lifetime (in days) predictions (median survival values) for all VIP players *non-loyal* in both variables. Colour represents grouping in terms of predicted LTV (top) and game level (bottom). The area of the circles is proportional in both cases to the expected LTV. Own elaboration using AoI predictions. The top image has previously appeared in [72].

predicted final level. It is, for example, rare, to find players with short expected final level and predicted lifetime above 400 days or predicted playtime above 1000 hours. However, variations in the player's skills still account for significant deviations: there are players with very long predicted lifetimes that are expected to be short lived in level (see, for example, the player at around 800 days and nearly 2500 hours in the short level group), while others are predicted to be long lived or even loyal for relatively short or medium expected lifespans (with some users in the long level group appearing soon after the 200 day mark, and the first loyal around 300).

This kind of analysis also facilitates outlier behaviour detection. Figure 7.5, for example, shows a player with very high expected playtime as compared to their predicted lifetime and the behaviour of the rest of the group (nearly 5000 hours of playtime in less than 600 days). The models are therefore predicting this player will play for over a year and a half, everyday, for a regular workday. This player is in need of disengaging rather than engaging strategies. This type of analysis also highlights how this type of extreme behaviour is actually found seldom in the real data.

Figures 7.6 and 7.7 focus on studying the relation between predicted playtime and in-game progression. The only difference between both plots in Figure 7.6 is that the colour codes the LTV prediction group in the top plot and the predicted lifetime group in the bottom plot. In both plots, the area of the circles is proportional to the predicted LTV. Figure 7.7 looks differently because, although still depicting the same playtime-level relationship, in this plot the area is proportional to lifetime predictions, therefore loyal players in this variable (i.e., all darkest blue circles from the bottom plot in 7.6) need to be excluded. Area of the circles is also proportional to expected days of lifetime for this plot.

It is clear from all three plots that progression through the first levels of the game is very quick and easy, hence the steep slope and the relatively low spread. The slope becomes much less steep around level 100. For even longer playtime and higher levels (over level 200), the relation between both predictions flattens, indicating a typical pattern in many games where higher levels need much longer to go through. The spread also becomes much larger, reflecting mainly that the varying skills of different players become increasingly important in predicting in-game progression for longer playtime.

Both plots in figure 7.6 seem to indicate that, as was the case with playtime, in the case of level there still seems to be some correlation with LTV but not that much: relatively large LTV is predicted for players with

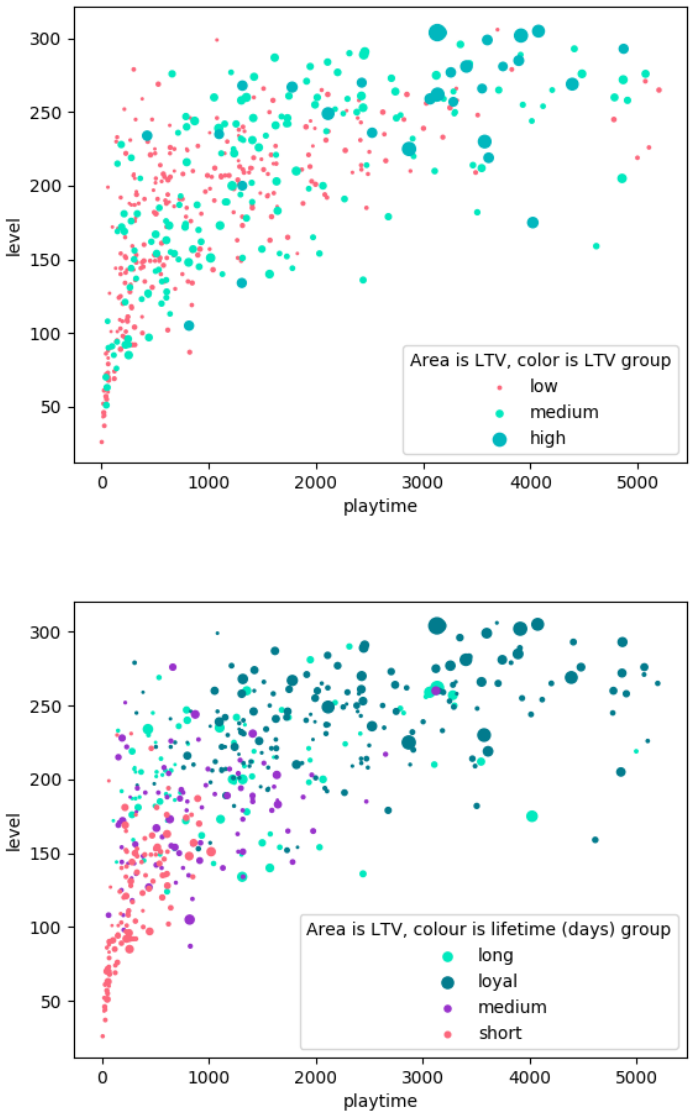


FIGURE 7.6: Game progression (in level) versus playtime (in hours) predictions (median survival values) for all VIP players non *loyal* in both variables. Colour represents grouping in terms of predicted LTV (top) and lifetime (bottom). The area of the circles is proportional to the expected LTV. Own elaboration using AoI predictions. The bottom image has previously appeared in [72].

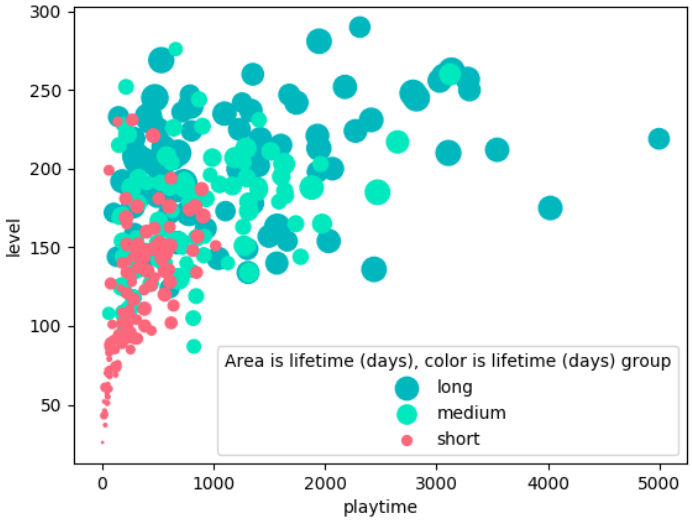


FIGURE 7.7: Game progression (in level) versus playtime (in hours) predictions (median survival values) for all VIP players non *loyal* in all variables modelled using survival models (lifetime, level and playtime). Colour represents grouping in terms of predicted lifetime. The area of the circles is proportional to the expected lifetime. Own elaboration using AoI predictions.

limited predicted in-game progression, while small LTV are found across all scales. This pattern does not seem, however, to be symmetrical: players with low LTV can reach the highest possible levels, while no player with high LTV can be found with expected level under 100 (and only four below 200).

Bottom plot together with Figure 7.7 again indicates, as was already unveiled when studying Figure 7.5, that although there is an obvious and strong correlation between lifetime and playtime, it still allows for significant variation. In particular, players loyal in lifetime populate all the range from 1000 to 5000 hours of playtime, and are typically predicted to get to levels over 200 (and in all cases above 150).

The predictions can also be normalised (subtracting the average and dividing by the standard deviation of the distributions depicted in figure 7.4), so that they are distributed around zero with unity variance. This can make the interpretation of individual predictions more straight forward to understand for certain uses, as it makes easy to detect values above or below average, as they will be positive or negative respectively. The absolute values can also be directly interpreted in terms of how many standard deviations away form the average each predicted value is.

Figure 7.8 shows the same playtime-level relation that has been explored in figures 7.6 and 7.7, this time using normalised values for both of them. Instead of the grouping, the colour of this plot is a scale representing normalised predicted lifetime, and the area of the circles is proportional to the expected LTV. Figure 7.8 zooms into the first and third quadrants of the plot, with players with expected playtime and level below average shown in the top plot, and those with both predictions above average in the bottom plot. They exemplify an equivalent and similar way of performing the same analysis that has been done above, arriving thus at the same conclusions regarding the much smaller times needed to complete low levels as compared to high ones, the impact of player skills, and significant variation of the lifetime-playtime relation across players.

Note how this representation highlights an interesting fact: for players expected to play and progress less than average, the relative variability is much greater in their predicted progression (two and a half standard deviations from the average) than in their predicted playtime (only one standard deviation from the average). On the contrary, for those expected to play and progress above average, this pattern is reversed, with some players with expected playtime more than five standard deviations away

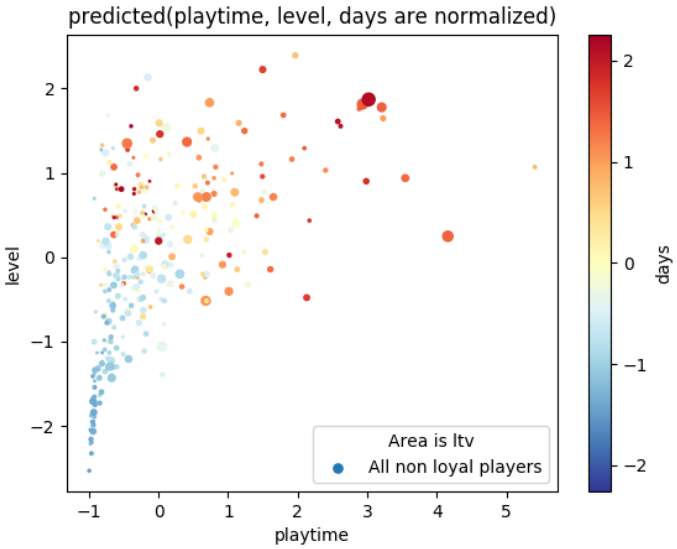


FIGURE 7.8: Normalised game level versus playtime predictions for all AoI VIP players *non-loyal* in level or lifetime. Positive (negative) values therefore correspond to players with predictions above (below) the average. The normalised predicted lifetime is shown as a colour scale, with larger than the mean values depicted in red shades, and smaller ones in blue. The area of the circles is proportional to the expected LTV.

from the average, but (again) only two and half standard deviations away from the average predicted level.

This is in accordance with the fact already observed in the predicted progression and playtime histograms (figure 7.4 plots (b) and (c)) of being skewed to larger and smaller values respectively. The skewness (asymmetry with respect to the mean) shifts the mean value towards larger/smaller values, hence incrementing the relative distance to the smallest/largest values in the sample. In this particular case, it probably simply reflects that time availability is constrained much more tightly than individual skills or playing ability.

The bottom plot in figure 7.8 suggests, that, for players with both predicted playtime and progression above average, despite the large spread, there seems to be a nearly two to one relationship in the playtime to level relation. That is, there is a larger concentration of players being approximately nearly twice as far from the mean expected playtime, than from the mean expected progression. When looking at the top plot in figure 7.8, one sees for players with both predicted playtime and progression below average, this still approximately holds for users close enough to the average (always with a high variability around this trend). For players more than 1.5 standard deviations of the average progression, and between 0.8 and 1 standard deviations away from the average playtime, the relation between both variables shows much less spread, with smaller variations in the deviation in playtime associated which a much higher variation in the deviation in level.

It is also interesting to note that, despite all the variability observed in the lifetime-playtime-progression relations, very seldom are users with both predicted playtime and progression below average found to have above average predicted lifetimes (warm coloured dots in figure 7.8 top plot). Likewise, very few players have above average predicted playtime and progression, while also having below average expected lifetime (cold coloured dots in figure 7.8 bottom plot).

The relationship between expected level and expenditure is explored in detail in Figure 7.10, where the area of the circles is proportional to predicted lifetime in days, while the colour codes the grouping in playtime predictions. It shows that, while all players with high expected outlay are also expected to make significant progress in the game before leaving, players with much smaller predicted LTV can do just as good. This suggests the game could be relatively fair, in the sense that although money could be helping to get faster through the levels (although this correlation could

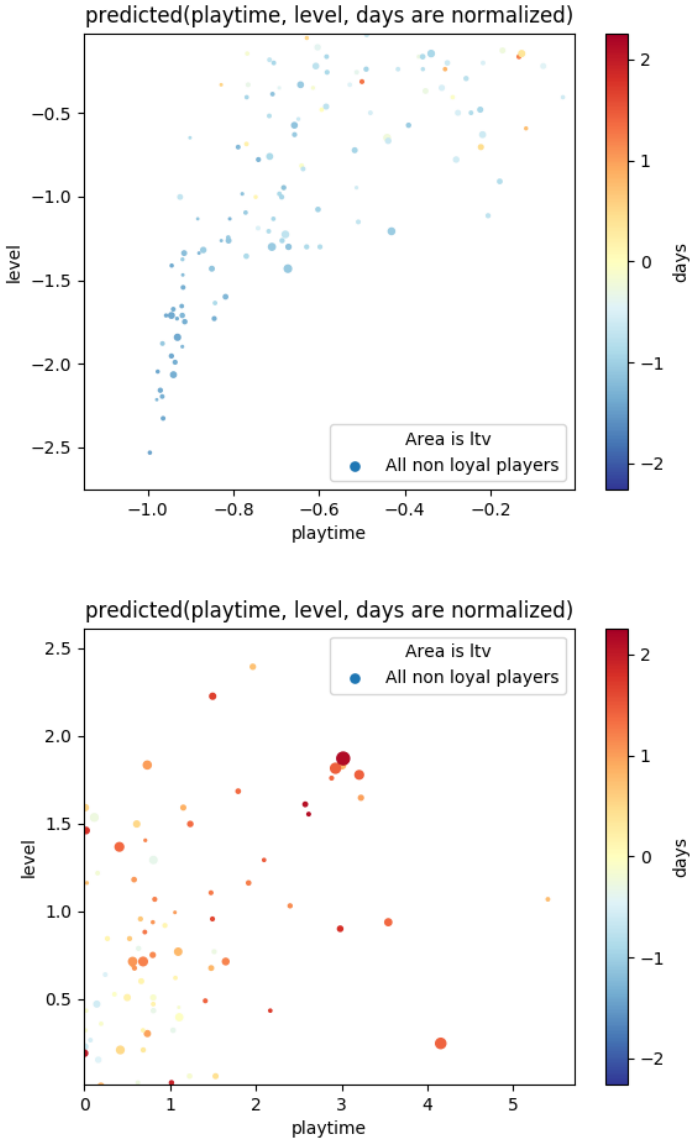


FIGURE 7.9: Zoom into two areas of figure 7.8. Normalised game level versus playtime predictions for all VIP players *non-loyal* in level or lifetime, and with both predictions below average (top), or above average (bottom). The normalised predicted lifetime is shown as a colour scale, and the area of the circles is proportional to the expected LTV. Own elaboration using AoI predictions.

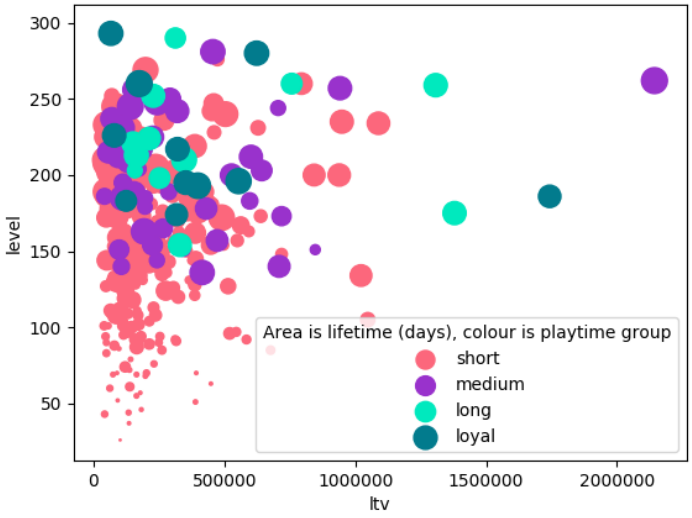


FIGURE 7.10: Game progression (in level) versus LTV (in local currency) predictions for all VIP players *non-loyal* in level or lifetime. Colour represents grouping in terms of predicted playtime. The area of the circles is proportional to the expected lifetime. Own elaboration using AoI predictions. The image has previously appeared in [72].

also be explained if, for example, players that reach higher levels tend to spend more), the analysis shows progression seems to have much more to do with the time spent playing and the skill of the player.

7.3 SKILFUL PLAYERS

As has been mentioned throughout the chapter, the combination of several of these classifications can be used to profile groups of players of particular interest. For example, players that are predicted to be loyal in terms of expected level to be attained in the game, are those who will probably finish the game or get to the highest possible levels. Players that are non-loyal in playtime are those for which the models can output an accumulated playtime prediction (while loyal players in playtime are expected to play *indefinitely*). This implies that those players that are simultaneously loyal in level and non-loyal in playtime are conceivably the most skilful of the engaged players (and more so the shorter the expected playtime is). There are 385 players fulfilling those conditions in the dataset considered.

The sort of exploration carried out in this chapter can be restricted to this group of interest (or other defined in similar ways) to understand their dynamics and to compare it to the general player behaviour. Figure 7.11 shows the histograms of predicted lifetime, playtime and LTV for these players. Lifetime and playtime predictions are distributed differently as compared to the total VIP population (see Figure 7.4). For skilful players, playtime instead of lifetime is more evenly distributed, and lifetime becomes now skewed to larger values. This makes sense, in that players that will reach such high levels will still typically play relatively longer than the average in both playtime and lifetime, hence making the previously skewed to shorter values (playtime) more uniform, and the previously uniform (lifetime), skewed to larger values. It indicates that for this type of players, even those that are more skilled among them (with short expected playtime), will do so typically over relatively extended times since their first login as compared to the general VIP population. These are players that could be bored by their quick progression and fail to log into the game until many days later, then making rapid in-game progression in a short playtime, and then keep repeating this behaviour pattern.

Figure 7.12 explores the lifetime-playtime relation of skilful players further. Area and colour of the circles code LTV predictions. It confirms the patterns just described, and for example, it makes obvious that high playtime is always associated with high lifetime, while players with short

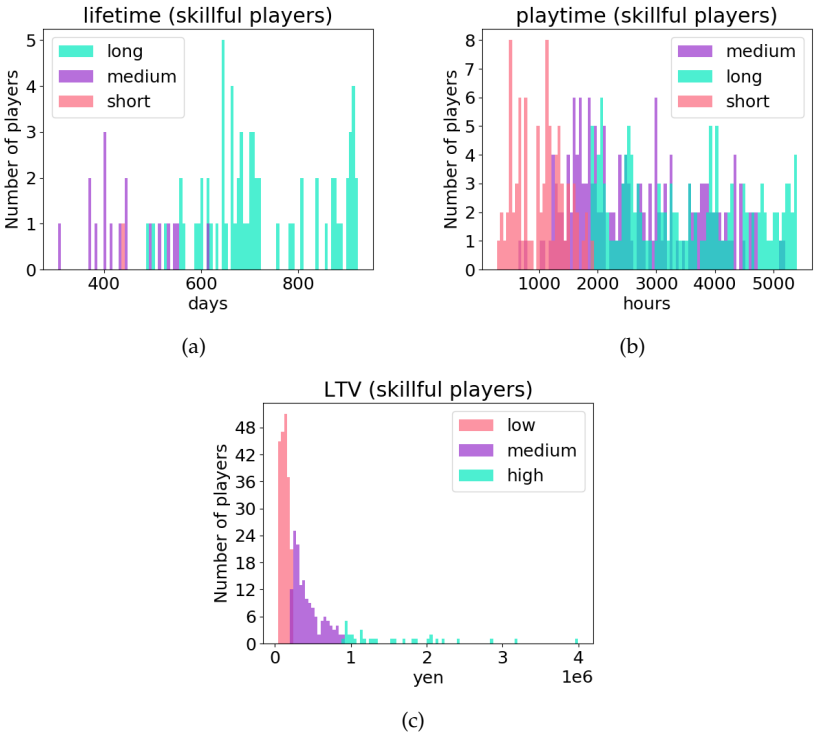


FIGURE 7.11: Histograms of predicted lifetime (days since first login (a)), playtime (hours played; (b)) and LTV (outlay in local currency; (c)) for VIP players *loyal* with respect to level and *non-loyal* in terms of playtime. Colours represent different groups for the corresponding variable. Own elaboration using AoI predictions. The images have previously appeared in [72].

expected playtime can be found across all lifetime scales. It also confirms previous discussions on the relative fairness of the game: LTV is distributed similarly as in the general population under study, and there are still many low spending players to be found in this group across all lifetime and playtime predictions.

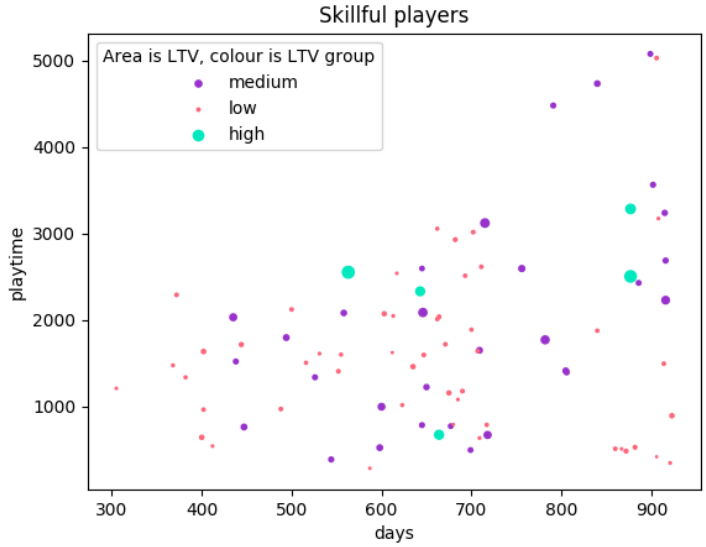


FIGURE 7.12: Playtime (in hours) versus lifetime (in days) predicted values for all VIP players *non-loyal* in both variables and *loyal* in terms of level. Colour represents grouping in terms of expected LTV and the area of the circles is also proportional to LTV. Own elaboration using AoI predictions. The image has previously appeared in [72].

7.4 SUMMARY AND CONCLUSIONS

This chapter has explored the possibility of using predicted engagement -expected lifetime, accumulated playtime, final level and total outlay as a proxy to it- to profile players. Segmentation using any of the variables studied can be useful in itself, but even more interesting seems to be analysing the relationship between the different predictions, and combining them to produce meaningful player categories. This exercise allows for an easy detection of outlier behaviour. It spots distinct behavioural patterns, and can help drawing qualitative conclusions about game dynamics and player population.

It has been discussed how to use churn and LTV predictions, but similarly conversion and purchase churn curves (described in chapter 6 sections 6.1 and 6.3) can be added to enrich the picture, and characterise purchasing behaviour in more detail. This can become particularly useful when using this approach to study all players (and not only top spenders). The methodology presented constitutes a very promising approach to generating richer profiling landscapes.

8

PERSONALISED ITEM RECOMMENDATIONS

If I can't work with you, I will work around you.
— Annie Easley

Methodologies to study individual user play and purchasing engagement have been described in chapters 5 and 6, and it has been suggested they could be used to decide when to target players. This has been proposed as a first step towards a greater personalization in games. Content, items, actions... could be made available, notifications sent, or discounts offered, at the *right time* for each player.

Another step in the same direction is personalising the content offered. This chapter is going to outline how item recommendations specific to every player can be issued in a production setup. The method proposed is able to work with hundreds of thousands of items and users. The results can be used, for example, to personalise game content (such as by making every player find their favourite -as opposed to a fixed or random- level one sword), to re-engage players by offering them free or discounted items particularly well suited to their taste, or to select the order in which items are shown to each player in the purchasing or selecting screen, making it easier for them to find items they really like. It is designed with games in which an enormous amount of virtual items are available in mind. These typically can be purchased by players using some sort of in-game currency (that in turn needs to be earned or purchased).

A method to recommend items in video games, when the collection available is not that large, was presented in [27], using AoI as illustrative dataset. Having a limited amount of items is typically the case in games where these are bought directly in real money. In that work, ensembles of ERTs (described in chapter 2 section 2.3.3) and DNNs (described in chapter 2 section 2.6) were trained to predict with what probability would each player purchase each item next, and those probabilities used as indications of the preferred items for that player at that stage in the game. The analysis favoured the use of ERT ensembles in production, as they had similar performance, while being more efficient computationally. This method however breaks down when dealing with literally thousands of items,

which is the case for many games (particularly when dealing with items that are bought in in-game currency).

The methodology described in this chapter extends the work presented in [27] to make it suitable to work in very high dimensional item spaces. This work has not been previously communicated in print due to the absence of an appropriate dataset available for publication to exemplify its uses and test its effectiveness (AoI has not a large enough collection of items). For that same reason, here only an outline of the methodology is given. The research leading to the operationalization of this system was carried out by Anna Guitart, Shi Hui Tan, and África Periáñez, together with the author of this thesis.

Section 8.1 explains how item clustering can be used to reduce the dimensionality of the problem. Section 8.2 deals with how to issue cluster recommendations per player, while section 8.3 focuses on ordering items within each cluster according to individual user preferences. Section 8.4 describes how to combine both to yield probabilities per player and item in the original dimension space. The validation metrics used are indicated in section 8.5, and the content of the chapter summarised in section 8.6

8.1 REDUCTION OF THE ITEM SPACE DIMENSIONALITY

The first step is to reduce the dimensionality of the item space. This is achieved by applying a clustering algorithm. The idea is to group *similar* items together, with this similarity playing a role in how the recommendations are issued. Features characterising the items themselves were used. For example, categories such as Armour, weapons, healing potions, magic items, etc. can be defined, each of them with subcategories such as helmet or shield for Armour, or sword or axe for weapons, and so on. Additional features can be used to describe the looks, material, resistance, price... Besides these categorical characteristics, play data was also fed into the clustering, to help identify which type of players, and in which stages or parts of the game, each item is more coveted or used.

Different feature selections and choices of clustering algorithm will arrive at different groupings. It is difficult, and to some point arbitrary, to decide which one is better. An objective measure would be given by the final performance of the recommendation system, but this is also not straight forward to determine, as there is no way to find out what the actual favourite items of each player are (as will be discussed in section 8.5). Validation metrics will be nonetheless defined, and they will play an important role in decid-

ing which clustering to select. The intended use of the recommendations however, is also to be taken into account. Certain applications might call for a clustering which strongly depends on the category or subcategory (e.g. swords together), others on the type of players that usually pick them (e.g. frequent players with character type warrior). This can be modulated through the feature selection, and by giving them adequate weights within the distance computation.

Game experience typically plays a crucial role in this type of recommendations, as they need to be *level appropriate*. Giving an inexperienced player a very powerful item, for example, might make them very happy, but *break the game*, by giving them an unfair competitive advantage with respect to other players, or by rendering the game boring due to the absence of matching challenges. On the other hand, an item that might have been much cherished by a player in the past, may become completely worthless and useless after a certain level of the game is left behind. This can be taken into account by making an extensive use of level related features (both in the clustering and for the ML cluster predictions described in section 8.2).

The final choice of clustering algorithm was HDBSCAN applied on the Gower's similarity matrix (see chapter 2 section 2.7). This is a conceptually sound choice, in that it enables the use of both quantitative and qualitative features. It also allows for clusters with complex shapes in the dimension space. Critically, it identifies outliers, which is important for an item recommendation system, as many games have objects that are very unique, non interchangeable with any other, and which need to be given individualised attention even in spaces with very high dimensionality. HDBSCAN determines the number of clusters, and requires the specification of few parameters, minimising the need for fine tuning. This choice was the one that produced the most reasonable groupings after detailed inspection of many different cases, and the best model performance after the whole recommendation process was completed and evaluated using the metrics described in section 8.5.

8.2 CLUSTER RECOMMENDATION

Once the item clusters were defined, the idea was to assign a probability to each user and item group. A predictive ML model was trained in the same spirit that was followed for the individual items in [27], to compute the probability, for each player, that the next item they will buy belongs to each cluster. Feature engineering made use of both player, cluster, and

player-cluster information. In addition to ERT ensembles (see chapter 2 section 2.3.3), XGboost (see chapter 2 section 2.3.4) was also tested, but found to perform worse (according to the metrics that will be described in section 8.5). DNN architectures were not tried out in this case. Decision tree ensembles were the preferred option because the ERT system described in [27] was already running in production for games with a limited number of items, and the goal was to extend this scheme to be able to deal with high dimensional item spaces too.

8.3 PREFERENCES WITHIN EACH CLUSTER

A recommendation system can also be run per cluster. Given that there can be a relatively large amount of item groups, and that the clustering itself has already made an extensive use of play features, it was decided to use a *simple* CF approach (see chapter 2 section 2.8). This family of algorithms is single feature based (previous purchase history in this case), and has proven to give good recommendations in many different contexts.

As was described in chapter 2 section 2.8, CF is particularly well suited to deal with systems were there is explicit feedback from the users (i.e., some review mechanism). Alternatively, a measure of implicit response can be defined. In this case, this needs to reflect the previous preferences of each player for the different items, and is thus constructed out of their transaction history. Several measures were tested, such as the total number of purchases per item, or total money spent in each item. The best results were obtained for the purchase count percentage, i.e., for each user, the percentage of their total purchases that corresponded to that item defines their *review* of that object.

As for the method in particular to carry out the CF, ALS matrix factorisation (see chapter 2 section 2.8) was chosen, as it is well suited to the problem, and outperformed the other methods considered (namely nearest neighbours and bayesian ranking). A different CF model was trained for each item cluster (and for all players). The CF scores were then normalised to generate a probability for the preference of each player for items within that cluster as $e^{s_i} / \sum_i (e^{s_i})$, for s_i the score of item i (this normalisation provided better results than others that were tried).

8.4 FROM CLUSTER TO ITEM SPACE

If items are meaningfully clustered, having a system yielding both good cluster recommendations, and item scores within each cluster, can already be a powerful tool. If it is to be used to target players with an item of their liking, their preferred cluster can be selected and an item from it simply picked at random. If the item has to be of a certain specific type, or the order in which to show them is desired, the scores yielded by the CF in an appropriate cluster can be used.

Both can also be combined to recover probabilities in the higher dimensional space. There is an associated probability per user and cluster (with probability one for each user when summing over all clusters), and an associated probability per user and item for each cluster after normalising the CF scores (with probability one for each user when summing over all items within each cluster). We can therefore arrive at a final probability per user and item (with probability one for each user when summing over all items in the game). Using conditional probabilities, the probability of player u choosing item i (p_{ui}) can be expressed as the probability of user u choosing an element of the cluster c ($p(c)$ as given by the ERT ensemble) to which item i belongs times the probability of choosing that item conditioned to having picked that cluster:

$$p_{ui} = p(c)p(i|c) \tag{8.1}$$

where $p(i|c)$ is zero when item i does not belong to cluster c, and the probability computed normalising the CF scores of cluster c ($e^{s_i} / \sum_{i \in c}(e^{s_i})$) otherwise.

Similarly to the output of the models described in [27], there is now a probability associated to each player-item, which can be interpreted as the probability, for the user, of that item being the next one they will purchase.

8.5 VALIDATION METRICS

Measuring the success of a recommendation system is not trivial. The only way to really determine if it works or not, is to have its purpose perfectly determined (for example, increase retention or playtime), and to then carry out A/B testing to see if there are significant differences in the indicator of interest when using the ML issued recommendations as compared to the previous system at work.

The system described in this chapter, however, is designed as a multipurpose generic methodology (that can be tweaked through feature selection and parameter fine tuning for different purposes). The basic models (both the ERT ensemble and the CF) are trained on user's previous purchase history, because it is considered to be a good proxy to what they like. In particular, the cluster ERT ensemble's goal is to predict to which cluster will the next purchased item belong, while the CF is concerned with the percentage of each player's total purchases that will be associated to each item. Note that there is a difference between predicting which item a player will buy, and which would they like best, which is what makes these systems' performance difficult to assess without proper experimentation.

In regards to the CF, the *expected percentile ranking* is used to evaluate how good the resulting preference order is, as described for example in [151]. This is a recall rather than precision oriented metric, which intends to take into account the uncertainty on the explanations to why a player does not purchase a certain item. While it is assumed that a user buys an item only if they like it, the lack of purchase could be due to a dislike for the object, but also to not having enough money to acquire it, or even to not knowing it exists. For each player and cluster, there is a list of items ordered by preference. For each user, the items have a percentile ranking ranging from 0% for the preferred item according to CF, to 100% for the last one in the list. The expected percentile ranking is given by:

$$\overline{rank} = \frac{\sum_{u,i} r_{ui} rank_{ui}}{\sum_{u,i} r_{ui}} \tag{8.2}$$

with $rank_{ui}$ the percentile ranking of item i and player u, and r_{ui} the associated observed response variable. The lower this score, the better, with 50% being the threshold value separating those recommendations performing better and worse than random ones.

The metrics described in [27] were deemed appropriate to assess, both the cluster prediction quality, and the final item probabilities. They were also used to discriminate between methodologies concerning the former, and to carry out fine tuning for both. These are:

- Ratio of times where the next predicted cluster/item per player was the next cluster/item purchased. The same considering top 2 and top 3 cluster/items.
- Ratio of times where the next predicted cluster/item per player was purchased in the next day where the user made at least a purchase. The same considering top 2 and top 3 cluster/items.

- Ratio of times where the next predicted cluster/item per player was purchased in the next window of time (selected here to be next week and next month). The same considering top 2 and top 3 cluster/items.

These scores were found to be significantly improved when applied to clusters with respect those obtained for AoI applying ERT ensembles to single items (see table 1 in [27]). Remarkably, even when evaluating item probabilities by combining the ERT and CF results, the performance was still very similar to that of the low dimensional spaces without clustering.

8.6 SUMMARY AND CONCLUSIONS

An approach has been proposed that combines clustering to group items into meaningful groups, ML predictions carried out to assign a probability per user and cluster (of next item to be purchased being in that cluster), and a CF for each group to translate cluster probabilities into item probabilities. It can be used to issue item recommendations in a production setup in an efficient and timely manner, even for very large number of available items in games with lots of players.Unfortunately, the research leading to this system running in operations was carried out with datsets not available for publication (AoI has a limited number of items). Hence, the methodology has been described, but no concrete results presented.

Depending on the purpose of the recommendations, the probabilities per player and item can be handled in different ways. If the aim is for example, to offer a free item to increase player retention, each user could be offered their favourite within a value range. It can also be used in subtler ways to personalise the game and make it more enjoyable. For example, selecting the weights in Gower's distance for each characteristic wisely, this method can be nudged to produce clusters where the type of item and level of the game where they are useful are crucial. Then if, for example, in a role type game, at some stage players always find a particular item because it is an element needed to successfully continue in the game (e.g. additional protection against attacks), this system could be used to give each player their favourite equivalent item (shield, Armour, protective magic spell...), even before they themselves are even aware that that is in fact their favourite one.

The recommendation of items has been discussed and is certainly of relevance in the context of video games. Methodologies working under the same principles can be developed to build systems that can recommend

actions, opponents, events... catered to each player's individual tastes. This opens a door to a lot of possibilities for game content personalization.

9

INTERDEPENDENT CHOICES AND VIDEO GAMES

I like crossing the imaginary boundaries people set up between different fields – it's very refreshing.
— Maryam Mirzakhani

This chapter analyses in-game behaviour from a fundamentally different perspective than the rest of the thesis. First and foremost, the aim is to understand qualitatively, rather than quantitatively, some of the processes at play, and the implications these may have collectively, rather than on individual players. Secondly, the results presented here are not concerned with the dynamic evolution itself of quantities in the game, as the rest of the content largely is. It considers the application of models from discrete choice theory -or their equivalent condensed matter formulation- described in chapter 2 section 2.10. As was explained there, these actually assume the system is in statistical equilibrium, in order to understand the properties emerging in the context of interdependent choices.

Players are constantly making choices that affect the status and evolution of the game: whether to login, take part in a particular event, subscribe to a certain service (ad-free playing for example), acquire a certain item, interact with another player... are only a few examples. In this sense, some of these choices and their study with the proposed framework here, are intimately related with much of the work already presented. Both whether a player remains active in the game (studied from different perspectives in chapters 4 and 5), and whether they remain non-PU or become PU, and how many purchases they will make (studied in chapters 4 and 6), can be understood as discrete choice problems. Player conversion, churn, and purchase churn -which have been the focus of much of what precedes- can therefore be studied formally and qualitatively using the framework proposed in chapter 2 section 2.10.

As was described there, discrete choice theory (or its equivalent statistical mechanical formulation), is particularly well suited to include the effect of social influence, i.e., to understand how the tendency of individuals to emulate the behaviour of their peers affects the collective outcome. The decisions described above regarding play, participation in events, item purchases... are definitely prone to be affected by the perceived choices of

other players, whenever users can interact in some way. The term *social games* will be used here loosely to describe any such game (and not necessarily one in which social interactions are a crucial element). This interaction can be very explicit -such as in role type games where players can actually band together or fight against each other-, implicit -like a puzzle game where the scores of other players are shown-, or anything in between.

In this chapter, decision interdependence is considered, meaning interrelation between two choice or spin like variables. This can represent two different groups of players making a choice, where each player is influenced by the (perceived) decision of the members of their group differently than by those of the other group (*non-local model* or *group interdependence scenario*). It can also be understood as describing a situation where each player in a single group is making two choices that affect each other (*local model* or *individual interdependence scenario*), in that they either tend to reinforce (positive interdependence) or exclude (negative) each other.

Examples where the group interdependence model might be useful are, for example, the study of daily login for players connected socially vs those who are not, participation in a particular event for PUs vs non-PUs, purchases of a particular item for players belonging or not to a certain clan in the game... Examples where the local model could be applicable include the study of simultaneous participation in two different events, purchase of two different related items, or of the reciprocal impact of participation in a certain event type and purchase of a particular type of item.

Note that both models can be used to very generally understand playing frequency and purchasing behaviour in social games, and the role user interaction plays in the collective outcome. That is, they provide a framework to qualitatively study the fraction of active and spending players, which are obviously intimately related to conversion, churn and purchase churn. The non-local model can be applied to any of the indicators (regarding login or purchasing activity) when dividing the population of interest into two groups (more vs less connected or experienced, two different player types as can be selected in the game, ...). Maybe even more interestingly, the local model can be employed to investigate the relationship between play and purchase decisions in social games, and how games can be in different *phases* in regards to these.

Note that by qualitative rather that quantitative study, what is in fact meant is that this part of the thesis does not follow the data driven approach of all the previous content. This chapter is the only one in which the data is not used to fit a particular model to then make quantitative predictions.

Its aim is rather to understand why different games (or the same game at different moments in time) show different collective behaviours in terms of how their player preferences are distributed, the strength of the social coupling, and on how the interaction between both groups or choices is. A limited number of attempts to explain real data with this type of models can be found in the literature [26, 32, 98, 105, 119, 201, 287]. While a general framework to study these problems will be developed, only radically simple models will be analysed and discussed in detail. These should be considered as toy models with the only goal of uncovering basic qualitative features emerging from player and choice interaction. It would therefore be unrealistic to expect them to provide quantitatively accurate descriptions of any real case. They can however be considered a first step in that direction.

In section 9.1 we motivate further the interest of studying two coupled Ising models form a binary choice perspective, and describe the general model resulting. In sections 9.2 and 9.3 the equations of state are deduced for the non-local and local models respectively. The phase diagrams for unbiased populations (those where individuals have no inherent preference and are guided only by their social interactions) are then discussed in section 9.4. The local case is further extended to non-homogeneous populations (those where there are preferences that are distributed in a certain way among the players) in a deterministic scenario (*zero temperature* or no free will or completely rational decision making case) in section 9.5. The chapter ends with a summary and conclusions in section 9.6. All content of this chapter is, unless clearly stated otherwise, an original contribution of this thesis. The models as they apply to general social sciences problems (rather than specifically in the video game context) were first described in [71, 74]. The results presented in section 9.5, while unpublished until now, had been presented at a couple of conferences.

9.1 COUPLED CHOICES AND COUPLED ORDER PARAMETERS

Coupled spin systems have been studied in statistical physics to describe plastic crystals and phase transitions among other phenomena [103, 106, 108, 110, 155, 171, 221, 259]. Two coupled Ising type models can also represent a system of interdependent binary choices. Consider the Hamiltonian:

$$H = -\frac{1}{N}\sum_{(i,j)}\left(J^s_{ij}s_is_j - J^t_{ij}t_it_j\right) - \frac{1}{N}\sum_{(i,j)_k} k_{ij}s_it_j - \sum_i \left(h^s_i s_i + h^t_i t_i\right) \qquad (9.1)$$

where s_i and t_i are the interdependent choice/spin variables, J^s_{ij}, $J^t_{ij} > 0$ the intra-couplings quantifying the strength of the spin interactions or social pressure, and k_{ij} the inter-coupling quantifying the interdependence between both choice variables. The summation over (i,j) is over all pairs $(i,j) = 1 \leq i < j \leq N$, while that over $(i,j)_k$ can correspond to two different options associated to two different coupling schemes and models. Considering an infinite range interaction $\sum_{(i,j)_k} s_i t_j = \sum_{1 \leq i < j \leq N} s_i t_j$ is equivalent to coupling both variables through their expectation value, in a system studied by Korutcheva [171] in the context of plastic phase transitions. We will refer to such a system as having *non-local inter-coupling* between the two choices. *Local inter-coupling* occurs for $\sum_{(i,j)_k} s_i t_j = N \sum_i s_i t_i$ with $k_{ij} = k_i \delta_{ij}$. Finally, the external fields h^s_i and h^t_i are the agent's *idiosyncratic willingness to adopt* (IWA) or opinion fields for choice s and t respectively [1].

The *non-local inter-coupling* or *group interdependence model* can be understood as representing two groups which are making the same choice under social influence of their peers (members of the same group), when their choice is also affected by their (accurate) perception of the average behaviour of the other group. The *local inter-coupling* or *individual interdependence model* can be thought to represent a single population, where all its members are making two choices that depend on each other for each individual. In both the local and non-local model, k_{ij} can be both positive or negative, indicating that the choices tend to be the same or opposite respectively. Solutions for the constant inter- and intra-coupling case are studied in detail for both models for the zero IWA case (homogeneous unbiased populations), in section 9.2 for group interdependence and in section 9.3 for individual interdependence, and their phase diagrams are described in 9.4. In section 9.5 some cases of the zero temperature for two coupled RFIMs (quenched disorder introduced in both IWAs) are considered for the local model.

Examples of interest from the general social sciences realm that can be studied qualitatively using the non-local model include, for example, voting for a certain political party from (self identified) left or right wing voters, having a full time job for individuals with or without kids, public opinion on a certain matter in neighbouring regions, school drop out in two schools of the same district, use of owned vs free software in two professional sectors, or demand of any product or brand between two groups of people

1 In a demand context, these can be rewritten, as for the single product or item case, as $h^s_i = b^s_i + p_s$, $h^t_i = b^t_i + p_t$ with b^s_i, b^t_i the *idiosyncratic willingness to pay (IWP)* and p_s, p_t the prices. This would allow to study how the prevalence of the buying choice varies with the price, i.e., the *demand curves*.

where there will be some kind of interaction between the choice in both groups (different to the interaction within the group).

The possible applications of the local model can seem relevant for even more interesting social sciences problems: voting a particular party for the Congress and Senate, families with children where all adult members have a full time job or where one becomes a stay home parent, public opinion on two given matters, school dropout and teenage pregnancy, buying a smartphone and computer of the same brand, hiring the same company to provide mobile and home internet connection services...

In the realm of video games, whenever there is social interaction in the sense that real players *see each other* (be that literally or through scores, for competition and/or collaboration), the non-local setup can be used to analyse the outcome of any relevant binary choice when considering the interactions between two groups of players. The local model provides a framework in which to understand how any single choice the players are making can be affected by another one they are simultaneously making.

Relevant binary choices within the game can be continuing to play, taking part in an event, purchasing an item, hiring a service, taking a certain action, interacting with another player.... When grouping players, segmentations taking into account social characteristics might be more relevant or be better suited for this framework, for example, dividing players that choose to interact with other players frequently vs those who don't, or dividing players according to the topology of their social interaction networks, or explicitly belonging to different social groups in the game. However, as far as it is assumed that any grouping is done according to some measure of similarity, that could make the two groups distinct enough that they will be affected differently from members of their own group than from the other, it can be applied too. This would include, for example, PUs and non-PUs, high skilled players and less skilled players, more and less experienced players...

Focusing on the generic choices of continuing to play or not, and spending or not, and defining groups along the lines of the segmentation strategies that have been outlined throughout this thesis, these models can be used to formally study quantitatively the same problems that have been analysed quantitatively throughout the previous chapters (namely player conversion, churn and purchase churn), and to explicitly analyse the impact of social interactions in the expected collective outcome, as will be discussed throughout the next sections.

The extension of this work to more than two groups or choices would be interesting and, in principle, while technically challenging, straight forward.

The system described by (9.1) has both average choices as its two (coupled) order parameters:

$$\begin{aligned} s &= \tfrac{1}{N}\textstyle\sum_i s_i = 2\mu_s - 1 \\ t &= \tfrac{1}{N}\textstyle\sum_i t_i = 2\mu_t - 1 \end{aligned} \tag{9.2}$$

with μ_s, μ_t the *fractions of adopters* (fraction of the population with $s_i = 1$ and $t_i = 1$ respectively).

Depending on the value of the *magnetisation or mean choice vector* (s, t), which has the two order parameters as components, the system can be described to be in one of three phases:

1. *Unpolarized* or *paramagnetic phase* ($s = 0$, $t = 0$): Half of the population deciding in favour and half against for both decisions at all times (disorder in both spin variables).

2. *Completely polarised* or *ferromagnetic phases* ($s \neq 0$, $t \neq 0$): There is some alignment in both variables, with four different cases depending on the sign of the two components of the magnetisation vector.

3. *Partially polarised* or *mixed phases* ($s = 0$, $t \neq 0$ *or* $s \neq 0$, $t = 0$): Order in one of the variables but not in the other, with again four different cases depending on which of the order parameters is zero and the sign of the non zero one.

Considering then the average magnetisation or choice vector (s, t), it will have zero norm for unpolarized phases and norm between 0 and 1 otherwise. It will only have value 1 in the complete absence of statistical fluctuations, i.e, for zero social temperature. If en euclidean norm in $\mathbb{R}^2$ is considered, then the first and third quadrants represent completely polarised systems where both components have the same sign, while the second and the fourth those where they have opposite sign. Partially polarised states are those with average magnetisation vector lying in one of the axes, and the paramagnetic phase is represented by the origin.

9.2 NON-LOCAL MODEL FOR HOMOGENEOUS POPULATIONS

Let us considered the system governed by the Hamiltonian:

$$H = \sum_{(i,j)} \left(-\frac{J_s}{N_s} s_i s_j - \frac{J_t}{N_t} t_i t_j - \frac{k(N_s + N_t)}{2N_s N_t} s_i t_j \right) - \sum_i (h_s s_i + h_t t_i) \qquad (9.3)$$

where summations over (i, j) are $1 \leq i < j \leq N_s$ for the first term, $1 \leq i < j \leq N_t$ for the second term and $1 \leq i \leq N_s$, $1 \leq j \leq N_t$ for the mixed term. Summations over i are $1 \leq i \leq N_s$ for the fourth term and $1 \leq i \leq N_t$ for the last term.

This system can represent two groups, one made up by N_s type s players, one by N_t t type players, in a social game. These could be for example, PUs and non-PUs, players of the same game in two different countries, players belonging to two different in-game clans or races, or player types (as in warrior or healer, for example), or groups of very different experience and/or skills. All players are making a binary choice, ranging from the basic and general play or not and purchase or not, to more specific ones related to taking any particular action, participating in any particular event, adopting some particular strategy, or purchasing any particular item. Within each group, all players are assumed to have the same inherent preference, which is the payoff associated to the decision making in the absence of social and choice interaction. Note this payoff can related to an explicit reward or penalisation in the game, be that in experience points, game lives or virtual currency. It can also represent a more abstract satisfaction for the real flesh and blood players.

All players have an additional reward if they mimic players of their group in their decision, with this urge being of equal strength for all members of the same group, but with possibly different intensity from group to group. In addition, there is an extra bonus or penalisation to their payoffs depending on how their choice compares to that of all members of the other group. These infinite range interactions (both within the group and with members of the other) are well approximated in the thermodynamic limit (large enough N_s and N_t) by the mean field theory (as will be described in section 9.2.1). This is then representing the case where each player is affected only by their (accurate) perception of what the average behaviour in each of the groups is (rational expectations or perfect knowledge of the system).

Both these (intra- and inter-group) interactions with the average values, can represent some sort of fuzzy social influence, in that they would mimic the desire of individual players to conform to the norm (or differentiate themselves from another group perceived as contrarian). For example,

players will be more likely to purchase a particular item if they see many other players in the game who already own it, and be more strongly affected by players in their same time zone than in others, as they will be more likely to interact with them. A particular event where they can find other players with whom to associate will be more attractive the more players that are taking part in it, but more so if these are of a similar level or experience. In this case, the inter-group coupling might even be negative, as the participation of a very large number of very experienced players might indicate the newbie that the event is not appropriate for them (and vice versa). As has been already noted, in certain video games, these interactions, that account in the end to additional terms in the player's utility, could actually reflect actual gains or losses for the players, in terms of in-game currency or experience points for example.

9.2.1 *Model definition*

Using mean field theory (on all intra- or inter-coupling terms) for equally sized groups (which will always be the case in the thermodynamic limit), equation (9.3) can be rewritten as

$$\begin{aligned} H =& \frac{NJ_s}{2}s^2 + \frac{NJ_t}{2}t^2 + Nkst - \\ & - (J_s s + kt + h_s)\sum_i s_i - (J_t t + ks + h_t)\sum_i t_i \end{aligned} \tag{9.4}$$

This model was first studied in detail in [171] in the context of plastic phase transitions. The results of that work are summarised here, then extended with the numerical analysis of the unbiased case in subsection 9.2.3 and section 9.4, and the implications of its use in a binary choice (and in particular video game) context discussed.

The partition function corresponding to the Hamiltonian in equation 9.4 for the representative canonical ensemble ($Z = Tre^{-\beta H}$ where Tr indicates sum over all possible spin configurations) can be expressed

$$\begin{aligned} Z =& e^{-\beta(\frac{N}{2}s^2 J_s + \frac{N}{2}t^2 J_t + Nkst)} \cdot \\ & \left[2\cosh\left(\beta\left(J_s s + kt + h_s\right)\right) 2\cosh\left(\beta\left(J_t t + ks + h_t\right)\right)\right]^N \end{aligned} \tag{9.5}$$

where $\beta = \frac{1}{K_B T}$, K_B is Boltzmann's constant and T the temperature (which in this case accounts for statistical fluctuations). The system's free energy

density ($f = F/N$ with F the free energy $F = K_B T \log(Z)$) will be therefore given in the mean field approximation by

$$\begin{aligned} f =& \frac{1}{2} J_s s^2 + \frac{1}{2} J_t t^2 + kst - \\ & \frac{1}{\beta} \ln\left(2\cosh\left(\beta\left(J_s s + kt + h_s\right)\right)\right) - \frac{1}{\beta} \ln\left(2\cosh\left(\beta\left(J_t t + ks + h_t\right)\right)\right) \end{aligned} \tag{9.6}$$

with stable states of the system those that minimise the free energy.

9.2.2 *Equations of state: solutions and stability*

The free energy in equation 9.6 has first derivatives in the order parameters s and t given by

$$\frac{\partial f}{\partial s} = J_s s + kt - J_s \tanh\left(\beta\left(J_s s + kt + h_s\right)\right) - k \tanh\left(\beta\left(J_t t + ks + h_t\right)\right) \tag{9.7}$$

$$\frac{\partial f}{\partial t} = J_t t + kw - J_t \tanh\left(\beta\left(J_t t + ks + h_t\right)\right) - k \tanh\left(\beta\left(J_s s + kt + h_s\right)\right) \tag{9.8}$$

and so its critical points are the solutions to the system of equations of state:

$$\begin{aligned} a\left(s - \tanh\left(\beta\left(J_s s + kt + h_s\right)\right)\right) = 0 \\ a\left(t - \tanh\left(\beta\left(J_t t + ks + h_t\right)\right)\right) = 0 \end{aligned} \tag{9.9}$$

where:

$$a = J_s J_t - k^2 \tag{9.10}$$

There will be two differentiated cases:

1. *Degenerate case ($a = 0$):* The resulting equation of state (besides $J_s J_t - k^2 = 0$) is:

$$\begin{aligned} & J_s s + kt - J_s \tanh\left(\beta\left(J_s s + kt + h_s\right)\right) \\ & -k \tanh\left(\beta\left(\tfrac{k^2}{J_s} t + ks + h_t\right)\right) = 0 \end{aligned} \tag{9.11}$$

2. *Non degenerate case ($a \neq 0$):* With system of equations of state:

$$\begin{aligned} s &= \tanh[\beta\left(J_s s + kt + h_s\right)] \\ t &= \tanh[\beta\left(J_t t + ks + h_t\right)] \end{aligned} \tag{9.12}$$

This work focuses in the non degenerate case, as it is of interest for a much larger region of the parameter space. The type of critical points that solutions to the system of equations 9.12 are (and hence their stability) will be determined by the Hessian of the free energy 9.6, which has elements:

$$\begin{aligned}\frac{\partial^2 f}{\partial s^2} =& J_s - \beta {J_s}^2 \frac{1}{\cosh^2[\beta\,(J_s s + kt + h_s)]} \\ & - \beta k^2 \frac{1}{\cosh^2[\beta\,(J_t t + ks + h_t)]} \end{aligned} \tag{9.13}$$

$$\begin{aligned}\frac{\partial^2 f}{\partial t^2} =& J_t - \beta {J_t}^2 \frac{1}{\cosh^2[\beta\,(J_t t + ks + h_t)]} \\ & - \beta k^2 \frac{1}{\cosh^2[\beta\,(J_s s + kt + h_s)]} \end{aligned} \tag{9.14}$$

$$\begin{aligned}\frac{\partial^2 f}{\partial s \partial t} =& k - \beta k J_s \frac{1}{\cosh^2[\beta\,(J_s s + kt + h_s)]} \\ & - \beta k J_t \frac{1}{\cosh^2[\beta\,(J_t t + ks + h_t)]} \end{aligned} \tag{9.15}$$

Stable solutions (minima) will be those with positive definite Hessian, i.e., those with positive determinant of the Hessian:

$$\det(\mathcal{H}) = a\,(1 - \beta J_s \gamma_s - \beta J_t \gamma_t) + \beta^2 a^2 \gamma_s \gamma_t \tag{9.16}$$

and positive first diagonal element:

$$\frac{\partial^2 f}{\partial s^2} = J_s - \beta {J_s}^2 \gamma_s - \beta k^2 \gamma_t \tag{9.17}$$

where $\gamma_s = \frac{1}{\cosh^2[\beta(J_s s + kt + h_s)]}$ and $\gamma_t = \frac{1}{\cosh^2[\beta(J_t t + ks + h_t)]}$. That is, critical points will be minima when both equation 9.16 and equation 9.17 evaluated on them are positive, maxima for positive equation 9.16 and negative equation 9.17, and saddle points whenever equation 9.16 is negative.

9.2.2.1 *Unbiased populations*

To continue with the analytical study of the system, let us now analyse the situation for the unbiased case, i.e., when the opinion field in both groups is zero $h_s = h_t = 0$. This will be the case when players have no inherent preference, and any reward or penalisation (be it concrete or abstract) is

due to imitation of one's group and the interaction between the average choice in both groups.

First, it is obvious that the unpolarized state $(s,t) = (0,0)$ can only be a solution of the system of state equations 9.12 for unbiased populations. For high enough temperatures ($\beta \to 0$), it is the only solution and it is stable.

In the absence of statistical fluctuations ($T = 0$), the situation is reversed, and there is now always complete order, with either (1,1) and (-1,-1) for $k > 0$ or (1,-1) and (-1,1) for $k < 0$ being the (stable) solutions of the system, both with identical probability (physically equivalent). At finite non zero temperature, the situation is qualitatively similar, although as there are now statistical fluctuations, both components of the average magnetisation will be less than 1 in absolute value (being closer to 1 the lower the temperature is). Solutions always appear in pairs (with both components of the same sign for positive k and of opposite sign for negative k). Further more, the system 9.12 is invariant under $k \to -k$ by changing the sign of one of the components of the average magnetisation vector, so the absolute values of the critical points are the same for k and $-k$. Note equations 9.16 and 9.17 also share this symmetry, so these pairs of solutions will also have the same stability.

Partially polarised or mixed phases can only exist for $k = 0$, i.e., when the system is made up by two uncoupled Ising models. This describes the situation where two unbiased groups are making a choice subject only to social influence from within their group. Mixed phases appear in this case when, for the given temperature, one of the Ising models is in a ferromagnetic or polarised phase ($J > K_B T$), while the other is unpolarized ($J < K_B T$). The non-zero component will therefore always be that associated to a higher intra-coupling constant J. The uncoupled system still has the same symmetries, so solutions always appear in couples, both with same component zero, and with the other component the same absolute value and opposite signs.

To study critical regions where the stability of the solution $(s,t) = (0,0)$ changes, equations 9.12 can be linearized for $|s| \ll 1$ and $|t| \ll 1$. At finite temperature in the unbiased case this yields:

$$\begin{aligned} s &= \beta\,(J_s s + kt) + O(s^3, t^3, s^2 t, s t^2) \\ t &= \beta\,(J_t t + ks) + O(s^3, t^3, s^2 t, s t^2) \end{aligned} \tag{9.18}$$

which leads to the expression

$$l(\beta) = a\beta^2 - (J_s + J_t)\beta + 1 = 0 \tag{9.19}$$

The roots of $l(\beta)$ in equation 9.19 are the points in the parameter space where the stability of the paramagnetic solution changes. These are given by the expression:

$$\beta = \frac{J_s + J_t \pm \sqrt{(J_s + J_t)^2 - 4a}}{2a} = \frac{J_s + J_t \pm \sqrt{(J_s - J_t)^2 + 4k^2}}{2a} \tag{9.20}$$

There will therefore be some regions for which $l(\beta)$ has a single positive root (and thus physically plausible and relevant to the discussion, at $\beta_b = \frac{J_s+J_t+\sqrt{(J_s-J_t)^2+4k^2}}{2a}$) and others where it has two ($\beta_b = \frac{J_s+J_t+\sqrt{(J_s-J_t)^2+4k^2}}{2a}$ and $\beta_c = \frac{J_s+J_t-\sqrt{(J_s-J_t)^2+4k^2}}{2a}$, with $\beta_c < \beta_b$).

As it is the stability of the paramagnetic phase what is now under study, equations 9.16 and 9.17 need to be considered (evaluated at $(s,t) = (0,0)$, i.e., $\gamma_s = \gamma_t = 1$) which will have in this situation the form:

$$\det(\mathcal{H}) = a\,(1 - \beta J_s - \beta J_t) + \beta^2 a^2 = a l(\beta) \tag{9.21}$$

$$\frac{\partial^2 f}{\partial s^2} = J_s + \beta \left({J_s}^2 - k^2 \right) \tag{9.22}$$

The Hessian in equation 9.21 also changes sign at the roots of $l(\beta)$ (equation 9.19), and the first diagonal element in equation 9.22 is positive only when $\beta < \frac{J_s}{{J_s}^2+k^2}$. Solving for J_s

$$\beta^2 < \frac{1}{4k^2} \tag{9.23}$$

$${J_s}^- < J_s < {J_s}^+ \tag{9.24}$$

with

$${J_s}^\pm = \frac{1 \pm \sqrt{1 - 4\beta^2 k^2}}{2\beta} \tag{9.25}$$

Putting all this together, this means a distinction between two different regimes can be made:

- *Strong coupling regime* ($k^2 > J_s J_t$): There is single physically relevant value β_b separating regions where the paramagnetic phase is a saddle point of the free energy of those where it is a maximum, so the unpolarized solution is never a stable state in this regime. For $T = 0$,

the pair of fully polarised critical points (with absolute value 1 and relative signs depending on the sign of k) are also saddle points. This is indicating that no stable solutions exist at all in this region, as will be corroborated by the numerical analysis in section 9.2.3.

- *Weak coupling regime* ($k^2 < J_s J_t$): Here β_b is still related to a change in the instability type of the unpolarized solution, from being a maximum (for $\beta > \beta_b$) to being a saddle point (for $\beta_b > \beta > \beta_c$) of the free energy, while β_c is an actual critical point where a second order phase transition takes place, as the paramagnetic phase becomes a stable solution (minimum of the free energy) for $\beta < \beta_c$. When looking at the zero temperature case, the pair of fully polarised solutions with absolutely no disorder ($(1,1)$ and $(-1,-1)$ for $k > 0$ and $(1,-1)$ and $(-1,1)$) are now also minima and thus stable states of the system, both with equal probability. The four of them are minima for $k = 0$ in the zero temperature case. Actually, there are four minima until at high enough temperatures they merge into the two mixed phase solutions. At zero temperature but non zero inter-coupling values break the symmetry of the four minima privileging two of them. There is a first order phase transition at $k = 0$, and for low enough values the other two will still be metastable states (local minima of the free energy with a higher value that the other two stable solutions). This suggests this will also be qualitatively the picture for non zero temperatures below the critical one, with components of the average choice having absolute values (strictly) between 0 and 1. Again, all this is corroborated by the numerical analysis that will be described in section 9.2.3.

The degenerate case ($k = J_s J_t$) corresponds to where the system changes from being able to reach equilibrium to having no stable solutions at all.

9.2.3 *Numerical analysis for unbiased populations*

Solutions to the system of equations of state in the non degenerate case (equation 9.12) were computed using the Newton-Raphson algorithm, and their stability assessed by evaluating equation 9.16 and equation 9.17 at each of them. Solutions showing convergence problems or with no invertible Jacobian were discarded, with number of iterations used between 100 and 1000, and convergence considered to have been attained when the absolute value of each of the components of the function was smaller that the

tolerance considered, which was set to $5 \cdot 10^{-6}$. To simplify the parameter space to study, J_s was set to 1, and thus the rest of couplings are measured in terms of it. For each collection of values of the parameters considered, different values for the initial average choices (each component between -1 and 1 with 0.1 steps) were considered in order to exhaust the solution space.

A couple of plots have been selected to illustrate the dependence of the average choice or magnetisation vector on each of the parameters in what follows. Additional auxiliary plots to better understand the phase diagrams that will be described in section 9.4 can be found in appendix A section A.1. In all plots, green is used for the component s and blue for component t, with dark colours (points) denoting stable solutions (minima of the free energy 9.6), and lighter colours non-stable solutions (asps x for saddle points of the free energy and crosses + for maxima). Note that all solutions and both components are plotted together, which could lead to some confusion. Whenever $k > 0$ the solutions will appear in pairs with either both components positive or both components negative, while for $k < 0$ the valid pairs have components with opposite signs.

9.2.3.1 *Dependence on temperature*

Figure 9.1 shows an example of the behaviour of (s, t) vs $K_B T$ for a system in the strong coupling regime (in particular for $J_s = 1$, $J_t = 0.6$, $k = \pm 0.8$, $a = -0.04$ and thus $K_B T_c = \frac{2a}{J_s + J_t + \sqrt{(J_s - J_t)^2 + 4k^2}} = 1.62$). As the previous analysis about the stability of the paramagnetic solution, and of the completely ordered one for zero temperature, was hinting, there are no stable solutions at all. The free energy of equation 9.6 has a saddle point in $(s, t) = (0, 0)$ for temperatures above T_b. For temperatures below it, it becomes a maxima and two additional symmetric ferromagnetic saddle points appear (with relative sign of the components depending on the sign of k).

Figure 9.2, on the other hand, depicts the situation on the weak coupling regime ($J_s = 1$, $J_t = 0.6$, $k = \pm 0.15$, $a = 0.58$, $K_B T_c = 1.05$ and $K_B T_b = 0.55$). As was discussed in section 9.2.2.1, T_b is the temperature separating regions where the unpolarized state is a maximum of the free energy from those where it is a saddle point, while T_c is a critical point at which there is a second order phase transition and the paramagnetic phase becomes the ground state of the system. Below T_c there are two critical ferromagnetic phases which are the absolute minima of the free energy. Below a certain

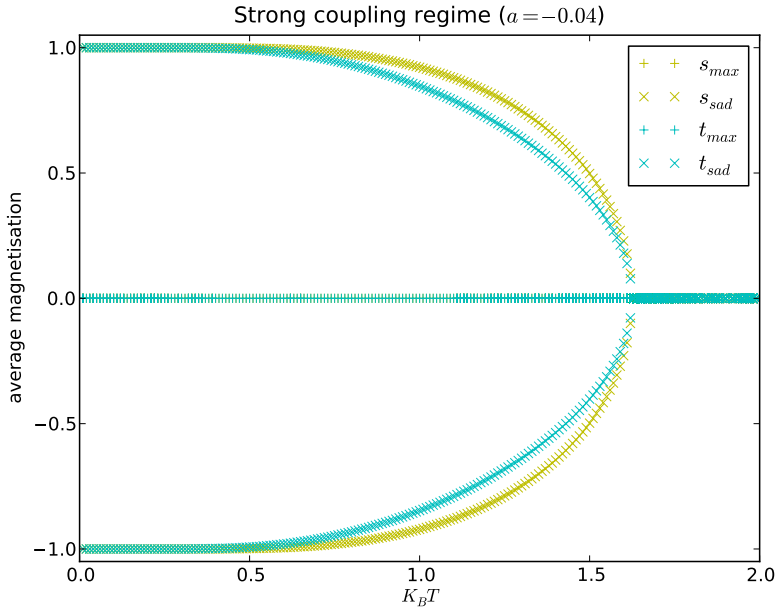


FIGURE 9.1: Dependence on temperature of the numerically calculated average magnetisation in the strong coupling regime: $J_s = 1$, $J_t = 0.6$, $k = \pm 0.8$ ($K_BT_c = 1.62$). Different solutions are plotted for temperatures between 0.01 and 2 every 0.01 (K_BT). Magnetisation is plotted in light green for s and light blue for t using asps ($\times$) for saddle point solutions and crosses ($+$) for maxima of f. Own elaboration using numerically computed solutions to the equations of state. The image has previously appeared in [71].

spinodal temperature T_a, an additional pair of saddle points and local minima appear, in which the magnetisation component associated with the lowest intra-coupling has opposite relative sign with respect to the other component as that dictated by the sign of k. These correspond to metastable states, as the free energy still has a higher value than for the other minima when evaluated on these points.

9.2.3.2 *Dependence on inter-coupling*

In terms of the inter-coupling k, values at which the nature of the paramagnetic critical point changes can be expressed as:

$$k_c = \pm\sqrt{J_s J_t - \frac{1}{\beta}(J_s + J_t) + \frac{1}{\beta^2}} \tag{9.26}$$

Figures 9.3 and 9.4 show the value and type of the magnetisation components that are critical points of the free energy against k, for $J_s = 1$, $J_t = 0.6$, and $K_B T = 1.2$ (high temperature) and $K_B T = 0.35$ (low temperature) respectively. Note that in both cases there are no stable solutions in the strong coupling regime $k^2 > J_s J_t$. Both plots are symmetric in k, however whenever there are pairs of ferromagnetic solutions, the relative sign between the two components of the magnetisation will be different for negative and positive values of k.

The values k_c of equation 9.26 are those (together with those delimiting the degenerate state) where the sign of the determinant of the Hessian 9.16 changes for the unpolarized state. This means there are two additional such values except for $K_B T$ having values between the two intra-coupling constants ($J_t < K_B T < J_s$ for $J_t < J_s$), i.e., between the two uncoupled critical temperatures. Equation 9.17 for the first diagonal element of the Hessian, on the other hand, changes sign for the paramagnetic phase for the two values satisfying $k^2 = J_s(J_s + K_B T)$.

The high temperature plot of figure 9.3 depicts the situation when the temperature is above both critical uncoupled ones. Here the paramagnetic phase is stable for $k^2 < k_c^2$, a saddle point for $k_c^2 < k^2 < J_s J_t$ and a maximum for $k^2 > J_s J_t$. At k_c there is a second order phase transition (if $|k_c| < J_s J_t$), as a pair of ferromagnetic solutions appear as minima of the free energy. These will lose their stability and become saddle points in the strong coupling regime.

The low temperature plot in figure 9.4 illustrates the dependence with k for temperatures below both the uncoupled critical ones. Here, while

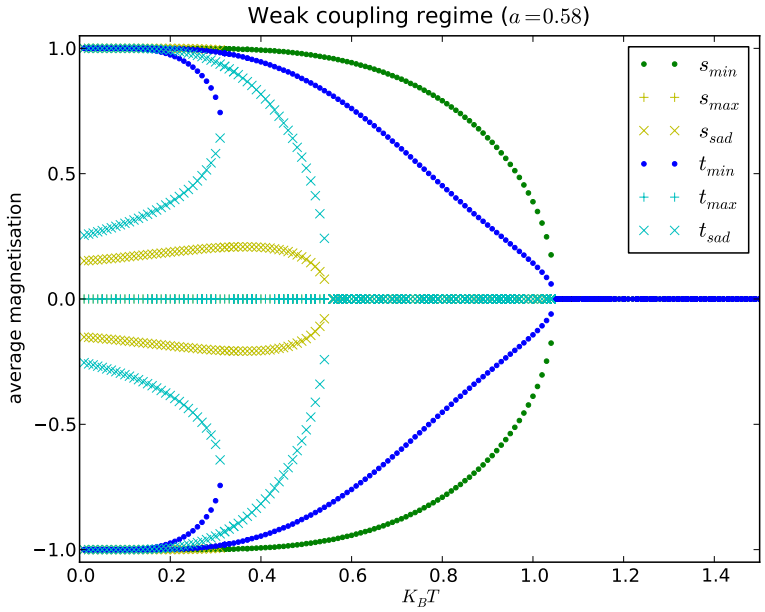


FIGURE 9.2: Dependence on temperature of the numerically calculated average magnetisation for the weak coupling regime: $J_s = 1$, $J_t = 0.6$, $k = \pm 0.15$ ($K_BT_b = 0.55$, $K_BT_c = 1.05$). Different solutions are plotted for temperatures between 0.01 and 1.5 every 0.01 (K_BT). Magnetisation is plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image has previously appeared in [71].

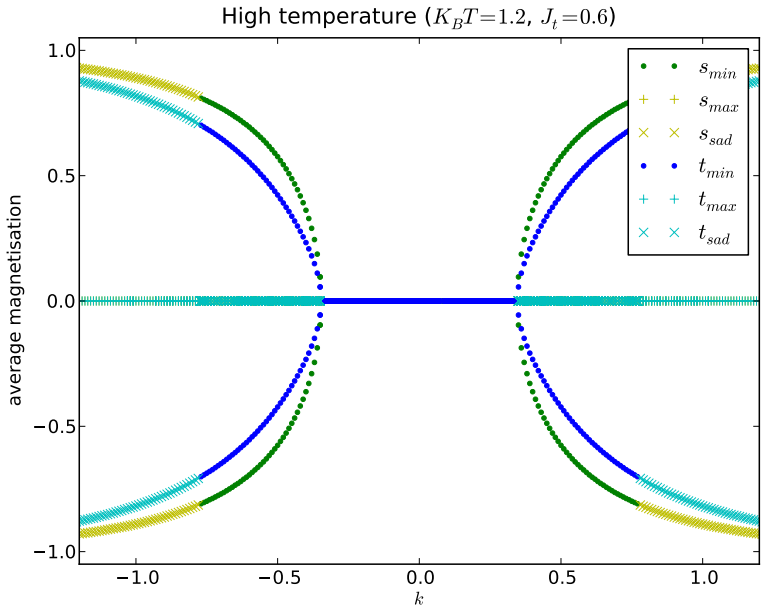


FIGURE 9.3: Dependence on inter-coupling of the numerically calculated average magnetisation for $J_s = 1$, $J_t = 0.6$, $K_BT = 1.2$ ($k_c = \pm 0.35$). Degenerate case (limiting value between both coupling regimes) for $|k| = \sqrt{J_s J_t} = 0.77$. Different solutions are plotted for k between -1.2 and 1.2 every 0.01. Magnetisation is plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image has previously appeared in [71].

$|k_c| < J_s J_t$, the unpolarized solution will be a a maximum except for $|k|$ between the critical $|k_c|$ and the degenerate value $J_s J_t$, where it is a saddle point. There is a completely polarised solution that is stable in the weak coupling regime, with absolute values closer to one the higher the absolute value of k (and components with relative signs as dictated by the sign of k). Between the two critical values of k (of the same absolute value and opposite sign), additional pairs of critical points (saddle points) appear, and for absolute values below the spinodal value $k^2 < k_a^2$, two more metastable ferromagnetic solutions emerge, where the component associated to lower intra-coupling has lower absolute value and opposite sign of what the sign of k would be favouring.

No plot is shown for the intermediate case $J_t < K_B T < J_s$ here because the behaviour will be that of the low temperature one (figure 9.4), but with no saddle point ferromagnetic solutions or metastable states at low k, and with the paramagnetic phase always being a saddle point in the weak coupling regime. See figure A.3 (d), figures A.4 (g), (h) and (i) and figure A.4 (a), (b) and (c) in appendix A section A.1 for some examples.

9.2.3.3 *Dependence on intra-couplings*

Finally, when considering dependence on the intra-couplings, roots of the polynomial $l(J_s)$ will be given by:

$$J_t^c = \frac{\beta(k^2\beta + J_s) - 1}{\beta(J_s\beta - 1)} \tag{9.27}$$

This means that, besides the change between weak and strong coupling at $J_t = \frac{k^2}{J_s}$, depending on the values of the inter-coupling and the temperature, there will be one or no positive critical value J_t where the sign of the determinant of the Hessian (equation 9.16) for the paramagnetic phase changes: the second derivative in equation 9.17 is negative for $K_B T < \frac{k^2 - J_s^2}{J_s}$ and positive for $K_B T > \frac{k^2 - J_s^2}{J_s}$.

There are two qualitatively different situations, as portrayed for $J_s = 1$, $k = \pm 0.3$ by figure 9.5 for high temperature ($K_B T = 1.5$, $J_t^c = 1.32$) and figure 9.6 for low temperature ($K_B T = 0.4$, $J_t^c = 0.55$). In the former, there are no stable solutions for $J_t < 0.09$ (strong coupling regime) as the only critical point, the unpolarized state, is a saddle point. For values of the intra-coupling constant placing the system in the weak coupling regime, but below the critical value $J_t^c = 1.32$, the paramagnetic phase is still the only critical point but is now stable. At the critical value there is a second

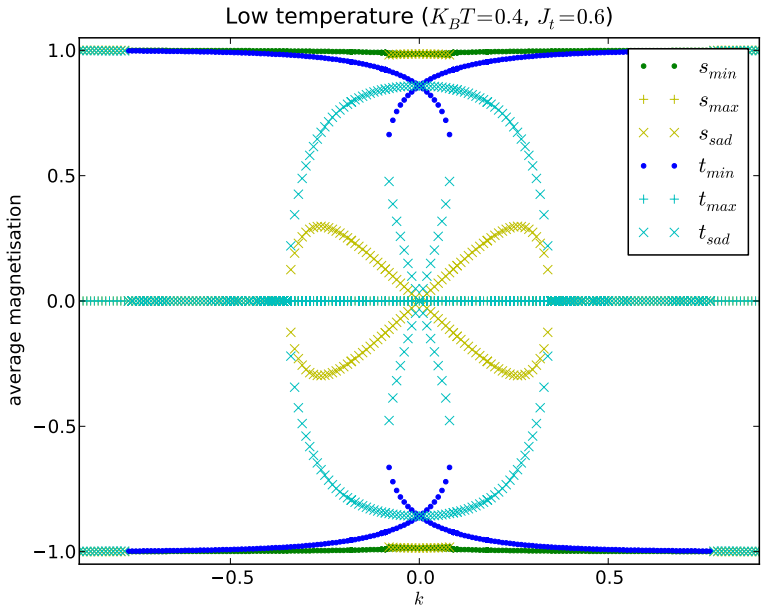


FIGURE 9.4: Dependence on inter-coupling of the numerically calculated average magnetisation for $J_s = 1$, $J_t = 0.6$, $K_BT = 0.4$ ($k_c = \pm 0.35$). Degenerate case (limiting value between both coupling regimes) for $|k| = \sqrt{J_s J_t} = 0.77$. Different solutions are plotted for k between -0.9 and 0.9 every 0.01. Magnetisation is plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image has previously appeared in [71].

order phase transition and it becomes again a saddle pint, with a pair of ferromagnetic minima appearing.

For low enough temperatures (see figure 9.6) the unpolarized phase is always either a maximum or a saddle point, and the ferromagnetic phase is the only stable solution (and this only in the weak coupling regime). For values above the coupling J_t^c, two new saddle point ferromagnetic branches appear (and the paramagnetic solution becomes again a maximum), and for even higher values, at the spinodal point J_t^a the ferromagnetic metastable states appear.

9.3 LOCAL MODEL FOR HOMOGENEOUS POPULATIONS

Let us now turn to the system described by two Ising models coupled locally, with dynamics governed by the Hamiltonian:

$$H = \sum_{(i,j)} \left(-\frac{J_s}{N} s_i s_j - \frac{J_t}{N} t_i t_j \right) - \sum_i (k s_i t_i + h_s s_i + h_t t_i) \tag{9.28}$$

where summations over (i, j) are $1 \leq i < j \leq N$ and summations over i are $1 \leq i \leq N$.

Its *translation* to a social utility representation of a binary decision making problem in social video games would be that of (a single group of) players making two choices, both of which are subject to some sort of social pressure in that all individuals are rewarded for aligning each of their choice with the same choice of the other players. All players have the same IWA for a given choice, but it can be different for the two different decisions (i.e., there are two constant opinion fields). Besides the payoff provided through imitation of their peers in each choice (social utility), and that associated to their inherent preferences (private utility), which are considered to be the same across all players in this case, there is an additional bonus or penalisation mediated through the inter-coupling k for each individual if they align (for $k > 0$) or not align ($k < 0$) their choices. This individual interdependence is in the case considered also constant for all players, i.e. all players have the exact same incentive to align (or not) their decisions. In the thermodynamic limit this system is well described by the Weiss mean approximation, which, as was discussed for the single Ising model in chapter 2 section 2.10, is equivalent to all players wanting to align their decisions to what they perceive to be the average behaviour of the group, when their perception is accurate (rational expectations).

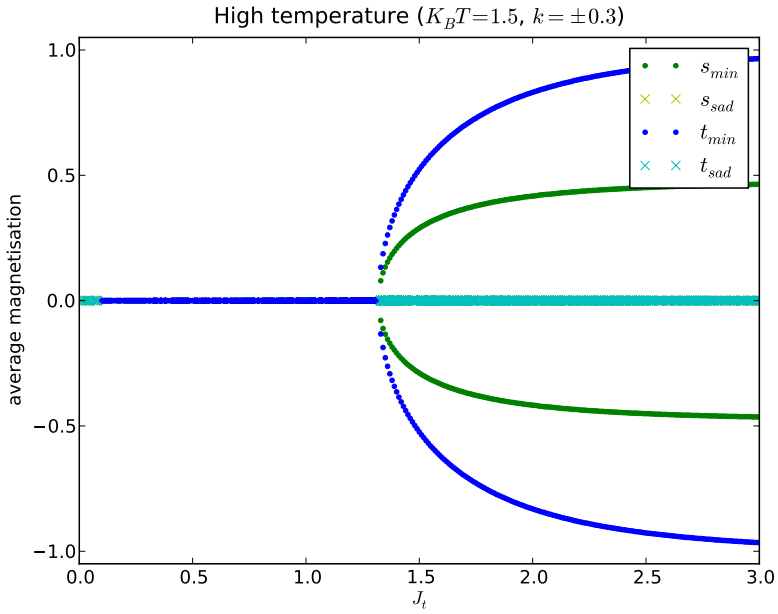


FIGURE 9.5: Dependence on the intra-coupling J_t of the numerically calculated average magnetisation at low temperature ($J_s = 1$, $k = \pm 0.3$, $K_BT = 1.5$, $J_t^c = 1.32$). Degenerate case (limiting value between both coupling regimes) for $J_t = \frac{k^2}{J_s} = 0.09$. Different solutions are plotted for J_t between 0 and 3 every 0.01. Magnetisation is plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asps ($\times$) for saddle points, non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image has previously appeared in [71].

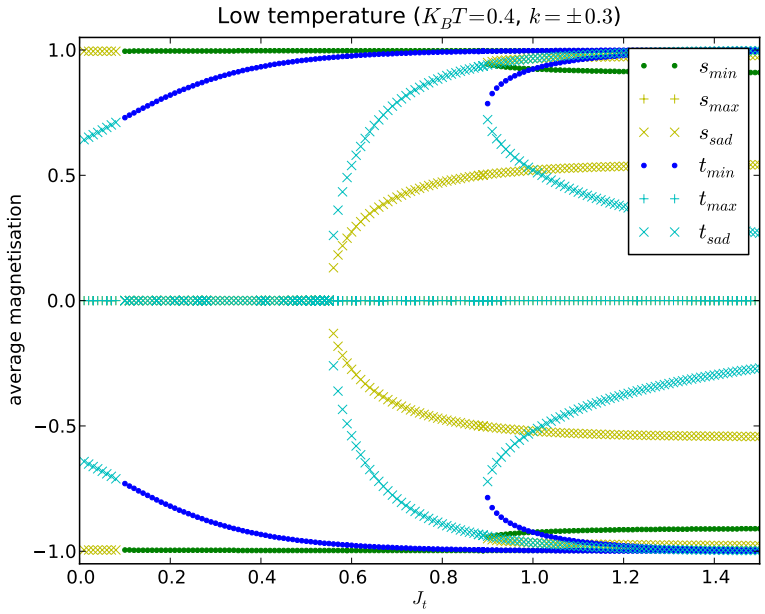


FIGURE 9.6: Dependence on the intra-coupling J_t of the numerically calculated average magnetisation at low temperature ($J_s = 1$, $k = \pm 0.3$, $K_BT = 0.4$, $J_t^c = 0.55$). Degenerate case (limiting value between both coupling regimes) for $J_t = \frac{k^2}{J_s} = 0.09$. Different solutions are plotted for J_t between 0 and 1.5 every 0.01. Magnetisation is plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image has previously appeared in [71].

Whether to play, coupled to deciding to pay for add-free play or additional in-game content, is a perfect general example of when such a system is relevant in the realm of video games. Other more particular examples could be related to purchasing two matching items (e.g. a sword and shield of the same type, with positive k) or two items that can not be used simultaneously (negative k), adopting a particular strategy (for example attack vs defence) and purchasing a particular item (for example better defensive gear), taking part in two related events... Payoff can measure an abstract quality, like the fun or satisfaction that would be experienced by the player by wearing matching cool garments in the game, or the expected better results in taking part in two events with similar dynamics. It could also be more tangible, like say, if combining the same type helmet and shield gives the player additional protection than the simple combination of both, or when the player gets an additional in-game reward for completing two related but independent tasks than that provided by the some of each of them.

The system with Hamiltonian given by equation 9.28 for $h_s = h_t = 0$ and for random inter-coupling k is studied in [110] in the context of plastic phase transitions. While the methodology and some results of that work is used in the analysis that follows, it focuses in symmetric distributions for k, and so most of the results there presented are of no use to our discussion. The rest of this section follows the same lines of the previous one (section 9.2), focusing now in the solutions of Hamiltonian 9.28 (instead of 9.3). In section 9.4 the results will be used to describe and compare the phase diagrams of both models (non-local and local).

9.3.1 *Model definition*

The Hamiltonian in equation 9.28 can also be expressed as:

$$H = \frac{J_s + J_t}{2} - \frac{J_s}{2N}\left(\sum_i s_i\right)^2 - \frac{J_t}{2N}\left(\sum_i t_i\right)^2 - \sum_i \left(k s_i t_i + h_s s_i + h_t t_i\right) \quad (9.29)$$

Following [110], the partition function of the representative canonical ensemble can be computed exactly in the thermodynamic limit yielding

$$Z = \frac{\beta N}{2\pi}(J_s J_t)^{\frac{1}{2}} e^{-\frac{\beta}{2}(J_s + J_t)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ds\, dt e^{-\beta N g(s,t)} \quad (9.30)$$

with

$$g = \frac{1}{2}J_s s^2 + \frac{1}{2}J_t t^2 - \frac{1}{\beta}\ln[2e^{\beta k}\cosh\left(\beta\left(J_s s + J_t t + h_s + h_t\right)\right) + 2e^{-\beta k}\cosh\left(\beta\left(J_s s - J_t t + h_s - h_t\right)\right)] \tag{9.31}$$

where $g(s,t)$ is the *free energy functional* such that:

$$f = \lim_{N\to\infty}\frac{F}{N} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ds\, dt\, g(s,t) \tag{9.32}$$

for F the free energy and f the free energy density of the system.

9.3.2 *Equations of state: solutions and stability*

The free energy and the free energy functional share the same critical points, which can be found using g derivatives:

$$\frac{\partial g}{\partial s} = J_s s - \frac{J_s e^{\beta k}\sinh\left(\beta(J_s s + J_t t + h_s + h_t)\right) + J_s e^{-\beta k}\sinh\left(\beta(J_s s - J_t t + h_s - h_t)\right)}{e^{\beta k}\cosh\left(\beta(J_s s + J_t t + h_s + h_t)\right) + e^{-\beta k}\cosh\left(\beta(J_s s - J_t t + h_s - h_t)\right)} \tag{9.33}$$

$$\frac{\partial g}{\partial t} = J_t t - \frac{J_t e^{\beta k}\sinh\left(\beta(J_s s + J_t t + h_s + h_t)\right) + J_t e^{-\beta k}\sinh\left(\beta(J_t t - J_s s + h_t - h_s)\right)}{e^{\beta k}\cosh\left(\beta(J_s s + J_t t + h_s + h_t)\right) + e^{-\beta k}\cosh\left(\beta(J_t t - J_s s + h_t - h_s)\right)} \tag{9.34}$$

yielding the system of equations of state:

$$\begin{aligned} s &= \frac{\tanh\left(\beta(J_s s + h_s)\right) + \tanh\left(\beta(J_t t + h_t)\right)\tanh\left(\beta k\right)}{1 + \tanh\left(\beta(J_s s + h_s)\right)\tanh\left(\beta(J_t t + h_t)\right)\tanh\left(\beta k\right)} \\ t &= \frac{\tanh\left(\beta(J_t t + h_t)\right) + \tanh\left(\beta(J_s s + h_s)\right)\tanh\left(\beta k\right)}{1 + \tanh\left(\beta(J_s s + h_s)\right)\tanh\left(\beta(J_t t + h_t)\right)\tanh\left(\beta k\right)} \end{aligned} \tag{9.35}$$

which can be written as:

$$\begin{aligned} s &= \frac{\alpha_s + \alpha_t\alpha_k}{1 + \alpha_s\alpha_t\alpha_k} \\ t &= \frac{\alpha_t + \alpha_s\alpha_k}{1 + \alpha_s\alpha_t\alpha_k} \end{aligned} \tag{9.36}$$

with $\alpha_s = \tanh(\beta(J_s s + h_s))$, $\alpha_t = \tanh(\beta(J_t t + h_t))$ and $\alpha_k = \tanh(\beta k)$.

The stability of critical points of the free energy will be determined by its Hessian, with components:

$$\frac{\partial^2 g}{\partial s^2} = J_s - \frac{\beta J_s^2 \gamma_s \left(1 - \alpha_t^2 \alpha_k^2\right)}{(1 + \alpha_s \alpha_t \alpha_k)^2} \tag{9.37}$$

$$\frac{\partial^2 g}{\partial t^2} = J_t - \frac{\beta J_t^2 \gamma_t \left(1 - \alpha_s^2 \alpha_k^2\right)}{(1 + \alpha_s \alpha_t \alpha_k)^2} \tag{9.38}$$

$$\frac{\partial^2 g}{\partial s \partial t} = -\frac{\beta J_s J_t \gamma_s \gamma_t \alpha_k}{(1 + \alpha_s \alpha_t \alpha_k)^2} \tag{9.39}$$

where $\gamma_s = \frac{1}{\cosh^2(\beta(J_s s + h_s))}$ and $\gamma_t = \frac{1}{\cosh^2(\beta(J_t t + h_t))}$.

The determinant of the Hessian is:

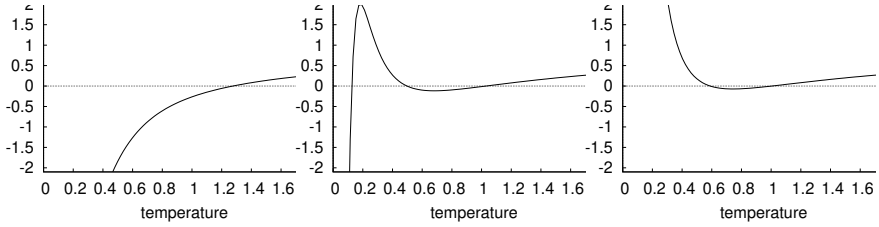
$$\begin{aligned} det(\mathcal{H}) = J_s J_t [1 &- \frac{\beta \left(J_t \gamma_t (1 - \alpha_s^2 \alpha_k^2) + J_s \gamma_s (1 - \alpha_t^2 \alpha_k^2)\right)}{(1 + \alpha_s \alpha_t \alpha_k)^2} \\ &+ \frac{\beta^2 J_s J_t \gamma_s \gamma_t \left((1 - \alpha_s^2 \alpha_k^2)(1 - \alpha_t^2 \alpha_k^2) - \gamma_s \gamma_t \alpha_k^2\right)}{(1 + \alpha_s \alpha_t \alpha_k)^4}] \end{aligned} \tag{9.40}$$

Solutions of the system of equations 9.36 are the critical points of the free energy, but only those where the Hessian is positive definite are minima and thus stable solutions. These are those with positive first diagonal element (equation 9.37) and positive determinant (equation 9.40). Those with positive determinant and negative first diagonal element are maxima, and finally those critical points where the determinant is negative are saddle points of the free energy.

9.3.2.1 *Unbiased population*

Let us considered the case where players do not have any inherent preference, i.e., $h_s = h_t = 0$. This is the case when any reward or penalisation associated to the decisions players are making are either purely social or due to the choice interaction.

The system presents similar characteristics to the non-local model (that was discussed in section 9.2.2.1), in that the paramagnetic phase is always a solution at finite temperature, and the only one at high enough temperature. The symmetry in k and thus the appearance of solutions in pairs, with the relative signs of both components of the average magnetisation



(a) (b) (c)

FIGURE 9.7: Plots of the function $l(K_BT) = J_sJ_t\left(1-\tanh^2(\frac{k}{K_BT})\right)\frac{1}{(K_BT)^2} - (J_s + J_t)\frac{1}{K_BT} + 1$ against temperature (K_BT). $J_s = 1$, $J_t = 0.6$ and different values of inter-coupling are considered. In (a) $k = 0.8$ and l has one root. In (b) $k = 0.15$ and l has two roots. In (c) $k = 0.05$ and l has three roots. Own elaboration. The image has previously appeared in [71].

depending on the sign of k, also holds in this case. The corresponding pair of ferromagnetic solutions with no disorder at all (i.e., with unitary norm) are the only solutions at zero temperature. Mixed phases are also here only possible when $k = 0$, where the behaviour is exactly the same as that displayed by the group interdependence scenario (both systems reduced to two uncoupled Ising models). Note that here, unlike for the non-local case, there are no metastable ferromagnetic solutions for zero temperature.

To study how the stability of the paramagnetic phase varies, the linearized system of equations of motion is considered:

$$\begin{aligned} s &= \beta J_s s + \beta J_t \tanh(\beta k)t + O(s^3, t^3, s^2t, st^2) \\ t &= \beta J_t t + \beta J_s \tanh(\beta k)s + O(s^3, t^3, s^2t, st^2) \end{aligned} \tag{9.41}$$

which can be simplified to

$$l(\beta) = J_sJ_t\left(1-\tanh^2(\beta k)\right)\beta^2 - (J_t + J_s)\beta + 1 = 0 \tag{9.42}$$

The polynomial $l(\beta)$ has here again roots where the type of the critical point the paramagnetic phase constitutes changes. In this case however, depending on the region of the parameter space considered, there will be one, two or three solutions, as illustrated by figure 9.7.

Evaluated at $(s,t) = (0,0)$, $\gamma_s = \gamma_t = 0$ and $\alpha_s = \alpha_t = 0$, and so the determinant of the Hessian (equation 9.40) and the first diagonal element (equation 9.37) are:

$$det(\mathcal{H}) = J_s J_t[1 - \beta\,(J_s + J_t) + \beta^2 J_s J_t\left(1 - \alpha_k^2\right)] = l(\beta) \tag{9.43}$$

$$\frac{\partial^2 g}{\partial s^2} = J_s - \beta J_s^2 \tag{9.44}$$

Upon inspection, the determinant of the Hessian changes sign too at the roots of $l(\beta)$. By looking at the signs of equation 9.43 and equation 9.44 for $T \to 0$ and $T \to \infty$, it is inferred that, depending on the number of roots of $l(\beta)$ as given by equation 9.42:

- If $l(\beta)$ has three roots $\beta_c < \beta_b < \beta_{b'}$, then there is a second order phase transition at β_c, as for $\beta < \beta_c$ (temperatures above the critical one), the paramagnetic phase is a minimum of the free energy, while for $\beta_c < \beta < \beta_b$ it is a saddle point. Between β_b and $\beta_{b'}$ it is a local maximum, while for $\beta > \beta_{b'}$ it becomes again a saddle point.
- If $l(\beta)$ has two roots $\beta_c < \beta_b$, there is still a second order phase transition at β_c, with the unpolarized state being the ground state of the system for $\beta < \beta_c$, and then a saddle point for $\beta_c < \beta < \beta_b$ and a maximum of the free energy above β_b.
- If $l(\beta)$ has one root, β_c is a critical point separating regions where $(0, 0)$ is a minimum from those where it is a saddle point.

Regardless of the number of roots, it is always the smallest (or only) value of β where there is a phase transition and the stability of the unpolarized phase changes.

9.3.3 *Numerical analysis for unbiased populations*

As in section 9.2.3, here too the Newton-Raphson algorithm is used to compute solutions to the system of equations 9.36, and their stability assessed by evaluating equations 9.43 and 9.44 at them. The same convergence criteria, iterations, and initial values for the average choice vector components are considered, with J_s set to 1 here too (and thus the rest of couplings measured in terms of it).

The plots follow the same conventions, namely, green is used for the component s and blue for component t, with dark colours (points) denoting stable solutions (minima of the free energy functional of equation 9.31), and lighter colours non-stable solutions (asps x for saddle points of the free

energy and crosses + for maxima). The symmetry in k must also be taken into account to correctly interpret how the solutions are paired. Additional auxiliary plots can be found in the appendix A section A.2.

As will be seen, qualitatively the situation is very similar to that of the non-local model described in section 9.2.3. Both phase diagrams will be discussed and compared in section 9.4. There is one major difference: as has been seen, for the non-local model, there are regions in the parameter space where the system becomes frustrated and there is no stable state. This is never the case for the local model. As the coupling of both choices is not social and is through each player, they can always successfully decide whether aligning them or not is the best option, and the frustration scenario is never arrived at.

9.3.3.1 *Dependence on temperature*

As for the non-local model, figure 9.8 shows the dependence with temperature in strong coupling (here referring to the region of the parameter space where $l(\beta)$ in equation 9.42 has a single root), and figure 9.9 for weak coupling (regions of the parameter space where $l(\beta)$ has at least two roots, three in this particular case). No explicit distinction is made between situations where $l(\beta)$ has two roots from those where it has three, as there is no change in stability of the solutions, and the former can simply be considered a particular case of the latter with $T_{b'} \to 0$. For both plots $J_s = 1$ and $J_t = 0.6$, with $k = \pm 0.8$ for figure 9.8 and $k = \pm 0.15$ for figure 9.9 (same parameter values than figures 9.1 and 9.2 for the non-local case).

Figure 9.8 is very similar to figure 9.1, but now saddle points have become local minima, and local maxima saddle points, indicating a stable unpolarized phase for temperatures above the critical one, and a pair of stable completely polarised solutions (with relative signs of the components that of k) for temperatures below, and a second order phase transition taking place at $K_B T_c = 1.28$ (value computed numerically using the Newton-Raphson algorithm). Note the value of the temperature at which the change of critical point type occurs is slightly lower than in the non-local case.

There are two qualitative differences between figure 9.9 and its non-local counterpart 9.2. There are now three values of the temperature where the nature of the paramagnetic critical point changes $K_B T_c = 1.03$, $K_B T_b = 0.50$ and $K_B T_{b'} = 0.13$ (all values computed numerically using the Newton-Raphson algorithm). As compared to the non-local case, here for $T < T_{b'}$, $(0, 0)$ becomes a saddle point, and an additional pair of ferromagnetic maxima appear, but this has no implications regarding the stability of

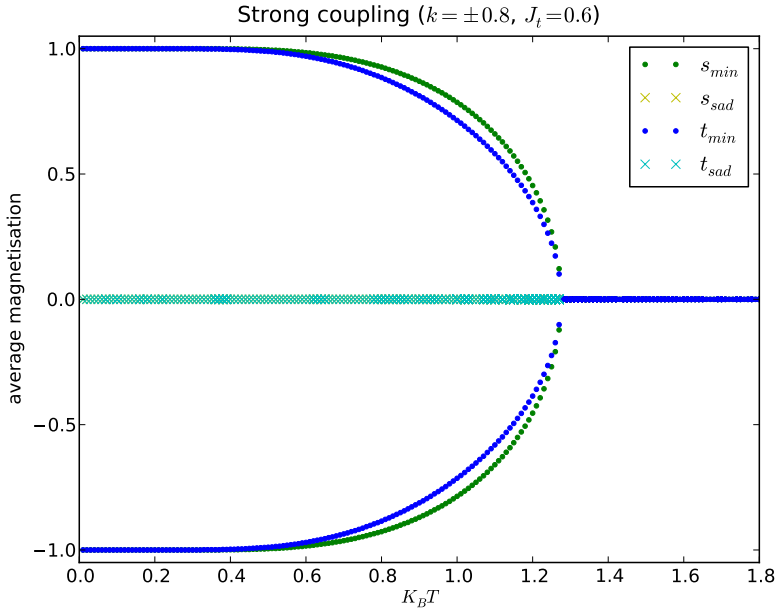


FIGURE 9.8: Dependence on temperature of the numerically calculated average magnetisation for $J_s = 1$, $J_t = 0.6$,$k = \pm 0.8$ ($K_B T_c = 1.28$). Different solutions are plotted for temperatures between 0.01 and 1.8 every 0.01 ($K_B T$). Magnetisation is plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asps ($\times$) for saddle point, non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image has previously appeared in [71].

solutions. More importantly, the ferromagnetic metastable solutions that here too appear, do not exist until zero temperature (fact that had already been noted after the analytical study of the system of equations of state), but rather for $T_{a'} < T < T_a$. Again here, as for the strong coupling case, the values of T_c and T_b are lower than in the non-local case, i.e., the region where the ferromagnetic solution is the ground state is smaller, and so is the region where metastable states exist.

9.3.3.2 *Dependence on inter-coupling*

The roots of the polynomial $l(\beta)$ in equation 9.42 in terms of k, where the determinant of the Hessian changes sign for the paramagnetic phase, are in the local case given by:

$$k_c = \pm K_B T \tanh^{-1}\left(\sqrt{1 - \frac{K_B T(J_s + J_t)}{J_s J_t} + \frac{(K_B T)^2}{J_s J_t}}\right) \tag{9.45}$$

The situation is qualitatively similar to that of the non-local model, with a remarkable difference besides the already stressed existence of at least a stable equilibrium in all the parameter space. Here again the uncoupled model critical values of the temperature appear and determine whether the paramagnetic phase changes from minimum to saddle point (for temperatures above both the uncoupled critical ones), or is never stable but changes from saddle point to maximum for low enough values of k (for temperatures below both of them). Now there is an additional threshold value $K_B T$ above which the paramagnetic phase is always the ground state of the system, with no change in its stability regardless of the value of k. There is no plot in this section illustrating this behaviour, but examples can be found in the appendix A section A.2 in figures A.12 plot (a) and A.13 plots (a) and (b).

Figure 9.10 depicts the behaviour of the free energy critical points for varying values of k for temperature above both uncoupled critical values ($J_s = 1$, $J_t = 0.6$, $K_B T = 1.2$ and $k_c = \pm 0.58$), but lower than its sum. It displays basically the same behaviour as figure 9.3 did for the non-local model, besides the fact that stable solutions do not disappear in the strong coupling regime, and so the pair of ferromagnetic solutions continue to be minima (and the paramagnetic solution to be a saddle point) of the free energy for arbitrarily large values of k.

Figure 9.11 displays a similar behaviour than its non-local counterpart depicted in figure 9.4 for low temperature ($J_s = 1$, $J_t = 0.6$, $K_B T = 0.4$

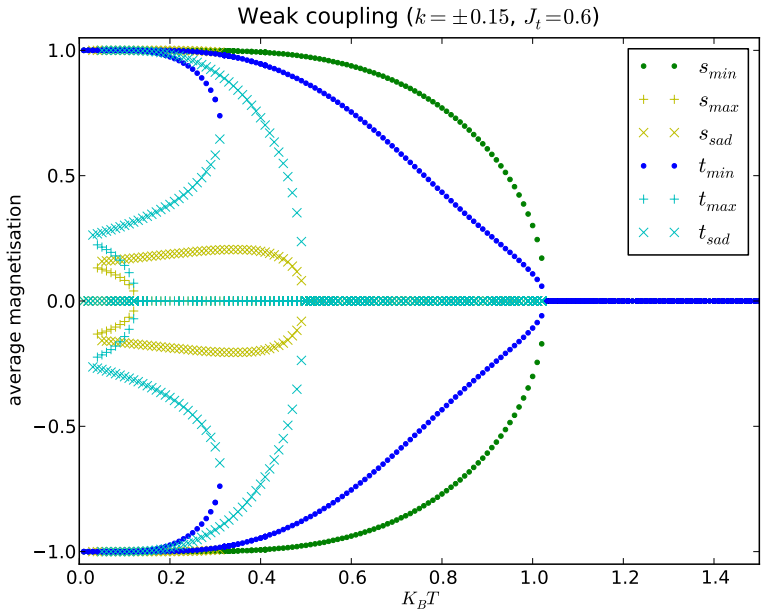


FIGURE 9.9: Dependence on temperature of the numerically calculated average magnetisation for $J_s = 1$, $J_t = 0.6$,$k = \pm 0.15$ ($K_B T_c = 1.03$, $K_B T_b = 0.50$ and $K_B T_{b'} = 0.13$). Different solutions are plotted for temperatures between 0.01 and 1.5 every 0.01 ($K_B T$). Magnetisation is plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image has previously appeared in [71].

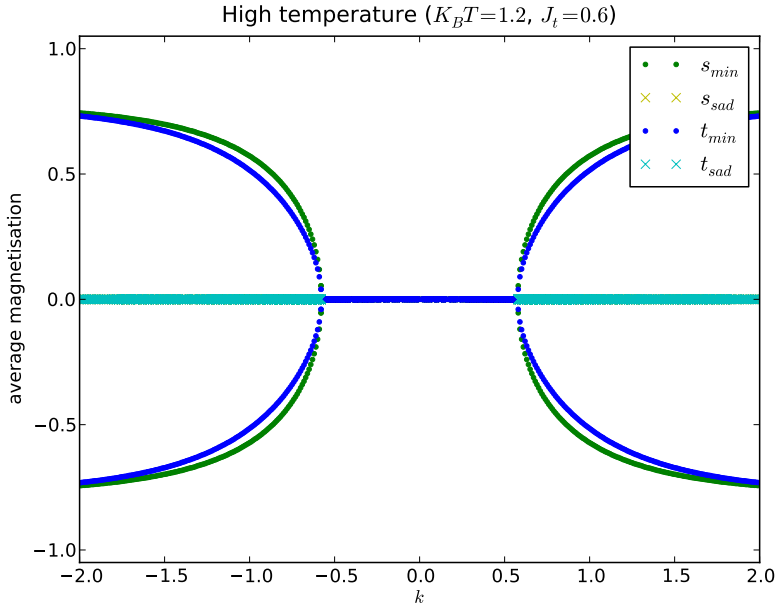


FIGURE 9.10: Dependence on the inter-coupling k of the numerically calculated average magnetisation at high temperature for $J_s = 1$, $J_t = 0.6$ and $K_BT = 1.2$ ($k_c = \pm 0.58$). Different solutions are plotted for k between -2 and 2 every 0.01. Magnetisation is plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asps ($\times$) for saddle point, non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image has previously appeared in [71].

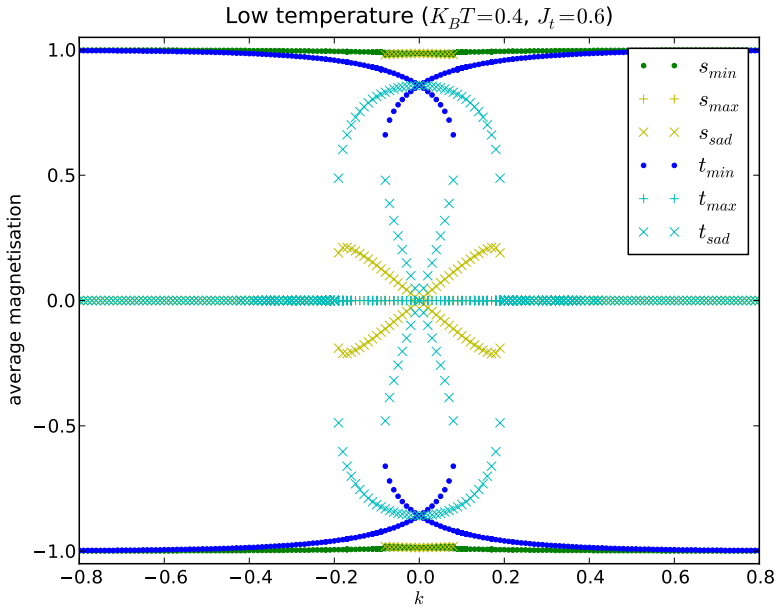


FIGURE 9.11: Dependence on the inter-coupling k of the numerically calculated average magnetisation at low temperature $J_s = 1$, $J_t = 0.6$ and $K_BT = 0.4$ ($k_c = \pm 0.19$). Different solutions are plotted for k between -0.8 and 0.8 every 0.01. Magnetisation is plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image has previously appeared in [71].

and $k_c = \pm 0.19$), besides the change when entering in the strong coupling regime

9.3.3.3 *Dependence on Intra-couplings*

The dependence on J_t of the polynomial l as defined in equation 9.42 yields as roots:

$$J_t^c = \frac{K_BT(J_s - K_BT)}{J_s(1 - \alpha_k^2) - K_BT} \tag{9.46}$$

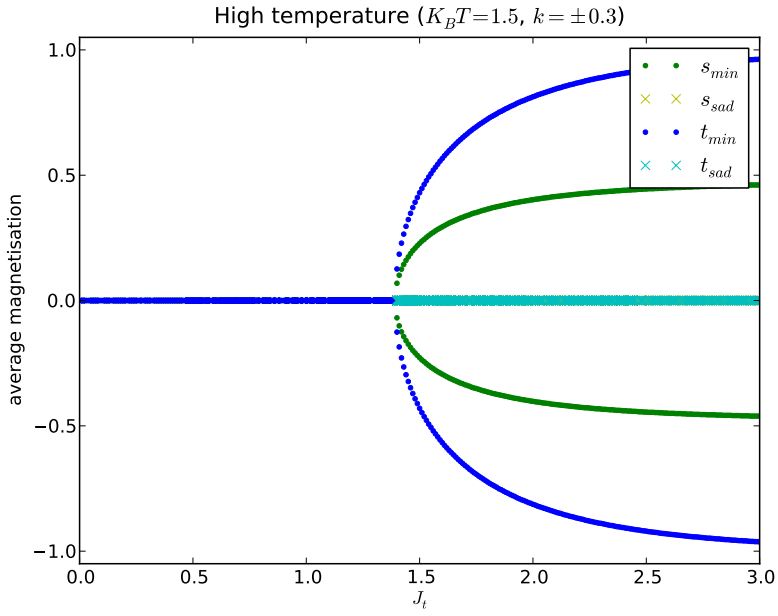


FIGURE 9.12: Dependence on the intra-coupling J_t of the numerically calculated average magnetisation for $J_s = 1$, $k = 0.3$, $K_BT = 1.5$ ($J_t^c = 1.39$). Different solutions are plotted for J_t between 0 and 3 every 0.01. Magnetisation is plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asps ($\times$) for saddle point, non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image has previously appeared in [71].

This means there are three qualitatively different behaviours depending on the region of the parameter where the system is. For $J_s < K_BT$ there is a second order phase transition at the only (positive and thus relevant to this discussion) critical value determined by equation 9.46, dividing regions where the paramagnetic is the only solution and it is stable, from those where it is a saddle point and the ground state will be one of two possible ferromagnetic solutions. This is depicted in figure 9.12 for $J_s = 1$, $k = 0.3$, $K_BT = 1.5$ ($J_t^c = 1.39$). It is the same qualitative behaviour as described for the non local model in figure 9.5 without the stability problems associated to the strong coupling regime when there is group interdependence.

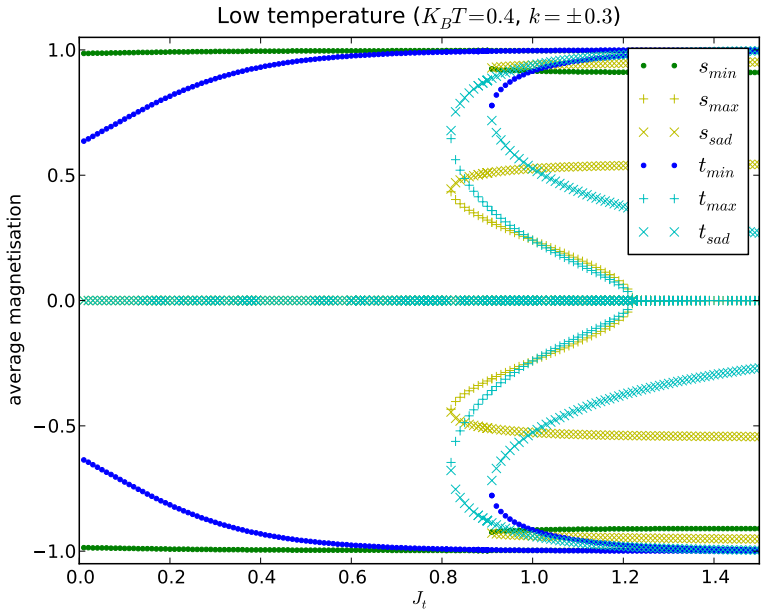


FIGURE 9.13: Dependence on the intra-coupling J_t of the numerically calculated average magnetisation for $J_s = 1$, $k = 0.3$ and $K_BT = 0.4$ ($J_t^c = 1.22$). Different solutions are plotted for J_t between 0 and 1.5 every 0.01. Magnetisation is plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp (×, for saddle points) or cross (+, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image has previously appeared in [71].

For $J_s > \frac{K_BT}{1-\alpha_k^2}$, J_t^c is no longer a critical point, as the paramagnetic phase can now only be a saddle point or a maximum. The ferromagnetic phase is the stable solution for all values of J_t, with additional saddle point branches and metastable states appearing for large enough values of the coupling. This is depicted in figure 9.13 for $J_s = 1$, $k = 0.3$, $K_BT = 0.4$ ($J_t^c = 1.22$), and is again very similar qualitatively to the situation for the non-local model depicted by figure 9.6, but without the stability problems of the strong coupling regime.

The final possible distinct behaviour occurs when $K_B T < J_s < \frac{K_B T}{1-\alpha_k^2}$, which can in fact be considered as a limit case of the behaviour just described as low temperature (only the ferromagnetic state is stable at all values of the coupling), but without additional saddle point or metastable polarised branches (no relevant J_t^c). It is not depicted graphically in this section, but examples can be found in the appendix A section A.2 in figure A.14 plot (i) and figure A.16 plots (h) and (i).

9.4 PHASE DIAGRAMS FOR HOMOGENEOUS UNBIASED POPULATIONS

This section discusses and compares the phase diagrams of the group interdependence (non-local k coupling) and individual interdependence (local k coupling) models in the case of unbiased populations ($h_s = h_t = 0$), as well as their implications when discussing interrelated choices in the context of video games. It contains figures depicting the type and number of solutions that are stable for the different cross sections of the parameter space.

All the plots show numerically computed solutions of the systems of equations of state 9.12 (non-local model) and 9.36 (local model) that are stable, i.e., minima of the free energy (equation 9.6) for group interdependence, and of the free energy functional (equation 9.31) for individual interdependence. The unpolarized or paramagnetic solution is always shown as a green dot, while pairs of ferromagnetic or completely polarised cases are shown as blue asps (x) when both average magnetisation components have the same sign, and blue crosses (+) when they have opposite signs. These always appear in pairs. Red triangles represent mixed or partially polarised phases, which also appear in pairs as has been discussed. Regions where blue asps (x) and crosses (+) are superposed are those where a pair of metastable states (with relative sign between the components of the average magnetisation vector different from what would be expected from the sign of k) exist, besides the ferromagnetic ground state pair (with relative sign as indicated by k).

Both models show qualitatively a remarkably similar behaviour, except for a critical difference that has been already repeatedly stressed: for the group interdependence or non-local model, there are no stable solutions for $k^2 > J_s J_t$.

Besides the numerically calculated solutions, some analytical results are also plotted. The *degenerate curve* $k^2 = J_s J_t$ is shown as a thick solid line for the non-local model, and separates the strong coupling regime (where

no stable solutions exist) from the weak coupling regime. Note regions in the strong coupling regime are those where the choice making process is dominated by the inter-choice interaction. Players have then a larger urge to imitate (or differentiate themselves from) the other group's choice than to emulate their own peers.

When considering both intra and inter choice interactions as purely social in nature, this would simply be pointing at ill defined groups. Making the influence of in-group members stronger than that of outsiders seems reasonable when defining player segments or groupings to study decision making processes in social games. The payoff associated to the interaction between groups can however be a tangible reward or penalisation (say, for example, players get additional experience points by participating in an event where many other players of the other group are taking part). In this case, the absence of stable solutions would be pointing to a real lack of satisfactory stable collective state, and the frustration preventing the system for reaching a state of equilibrium could be very real. It would probably be indicating some sort of non desirable in-game planing or design.

Imagine the case of participation in an in-game event, where players can form alliances and confront each other, of two groups of users of very different experience levels. If the event involves fighting other users, players will typically be more attracted to take part if there is a significant amount of users of a similar level, both to provide a well balanced challenge and a reasonable chance at winning. Depending on the event dynamics, the presence of users of the other group can also be attractive ($k > 0$). This will be particularly so if it favours collaboration across different experience levels, but could simply be due to accomplished players looking for an easy win, and/or newly arrived ones for an exciting challenge with the possibility of a major win (if player skill and not only game experience plays a role). Game dynamics could be such that it actually discourages joining events with players of a different experience level ($k < 0$). Imagine for example, there is really no chance of winning (not even experience points) for the new comers, and also absolutely no gain whatsoever in confronting and defeating amateur players to the seasoned ones. These are all reasonable game dynamics.

Very different would be cases in which, for example, very experienced players were rewarded more for beating unexperienced ones than their peers, while newbies also received a larger compensation for this than for overpowering their own equals. Or those in which the preference for similarly experienced players was so big it completely prevented players

from mixing, virtually compartmentalising the game into categories from which there is no escape. These would both correspond to undesirable game dynamics, both described by the strong coupling regime ($k^2 > J_s J_t$).

The *mixed segment* is the only region where the partially polarised pair of solutions are the ground states, and is plotted as a dashed line. It is the segment for $k = 0$ between the two uncoupled critical values $J_t < \beta^{-1} < J_s$ when $J_t < J_s$ (or $J_s < \beta^{-1} < J_t$ when $J_s < J_t$). It is associated to first order phase transitions in $k = 0$.

It corresponds, in the group interdependence scenario, to the regime in which two groups are making a decision affected by social interaction from within the group, but completely isolated from the other ($k = 0$). This would be the case, for example, of players always playing at different times (imagine opposite sides of the globe with very different time zones). For the individual interdependence scenario, this corresponds to the case of unrelated decisions, such as picking between two aesthetically different but otherwise identical items, and deciding to confront or not another player encountered.

The *critical curve* (where there is a second order phase transition between polarised and unpolarised phases) is shown as a solid thin line. It is given by the combined conditions:

$$\begin{aligned} &\beta^{-1} > \max\{J_s, J_t\} \\ &2a\beta = J_s + J_t \pm \sqrt{((J_s - J_t)^2 + 4k^2)} \end{aligned} \tag{9.47}$$

in the non-local model or group interdependence case, and by:

$$\begin{aligned} &\beta^{-1} > \max\{J_s, J_t\} \\ &J_s J_t \left(1 - \tanh^2(\beta k)\right) \beta^2 - (J_t + J_s)\beta + 1 = 0 \end{aligned} \tag{9.48}$$

in the local model or individual interdependence case.

Note the critical curve delimits the areas of the parameter space where there will be spontaneous magnetisation or choice alignment due to social pressure once the opinion fields are turned on, and thus the regions where there might be additional metastable states (at low enough values of the IWAs) associated to a first order phase transition at zero opinion field, where very rapid drastic shifts in the average choices are to be expected. These are the regions where, when moving to study players with inherent preferences on the choices, but still relatively small as compared to the payoffs provided through social imitation and/or choice interaction, a richer phenomenology with additional combinations of possible fractions of adopters will appear.

The concrete values of the parameters for which the cross sections are plotted have been selected to be representative of all possible phenomenology (that has already been generally discussed in sections 9.2.3 and 9.3.3 including that of non stable critical points). For all plots $J_s = 1$ (which is equivalent to measuring all other couplings in terms of it).

Figure 9.14 shows the $J_t - \beta^{-1}$ section at $J_s = 1$ and $k = 0.3$ for the non-local (a) and local (b) models. Note they are remarkably similar both qualitatively and quantitatively, with the very significant difference of the non-local model not having stable solutions for low enough values of J_t (where the system will be in the strong coupling regime). If both Ising models were uncoupled, the section would have as critical curve the slope 1 line across the origin. Each of the choices could be in two (physically equivalent and thus equally probable) ferromagnetic states, of equal absolute value. Combining the two possibilities for each choice would yield four possible states in all the ferromagnetic region. The choice coupling, be that through group or individual interdependence, extends the ferromagnetic region, and selects only two of the four ferromagnetic states by determining the relative sign between both components of the average magnetisation vector. For some regions however, there are two additional metastable states (with relative sign different from that dictated by k) associated to first order phase transition at $k = 0$. The shift in the critical curve and extended ferromagnetic phase implies that there will be a range of temperatures for which the paramagnetic solution will never be stable regardless of the value of the rest of the parameters. In the individual interdependence case, the inter-choice coupling accounts for regions in the ferromagnetic phase even for no social coupling $J_t = 0$.

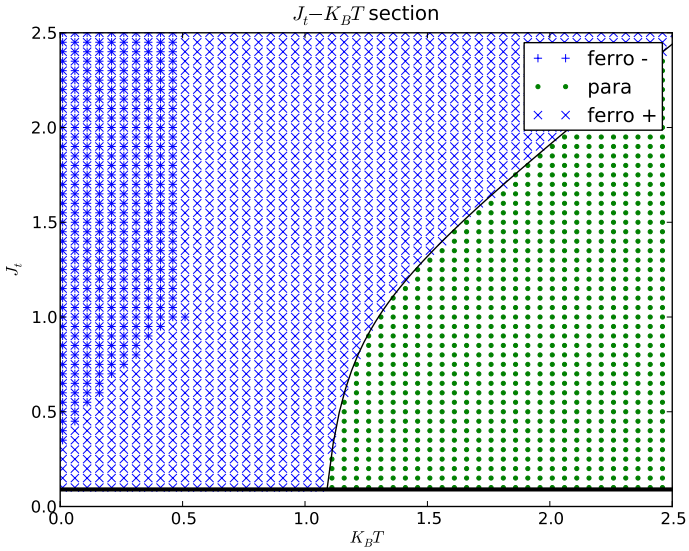
Consider the example of players being either warriors or sorcerers, picking between two outfits (non-local model). None of the choices brings any particular advantage, but players are assumed to be influenced by what appears to be more fashionable (i.e. choices the other players are making), particularly within their group. Additionally, all of these players are deciding whether or not to participate in two different events (local model). Here, participating and succeeding in both simultaneously, does bring users an additional reward. However, it also prevents players from taking part in other two such events. Chances of success increase the more players involved in the event.

For any given intensity of the choice coupling (desire to resemble or differentiate from the other group, additional reward possible from simultaneous participation), for large enough statistical fluctuations with respect

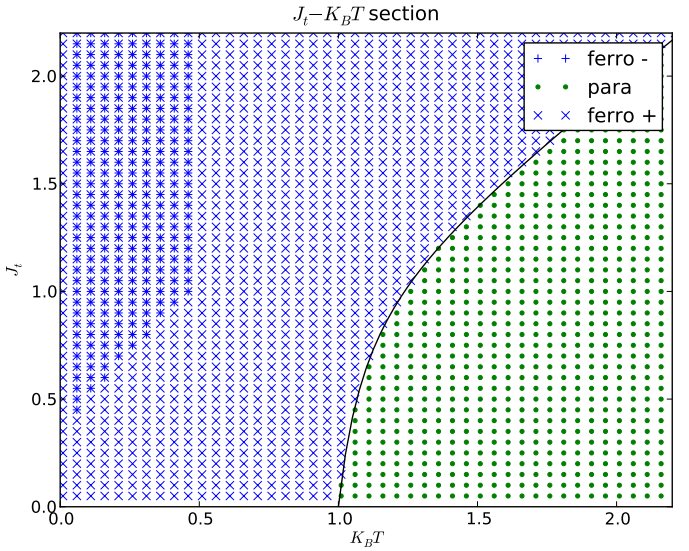
to social pressure, approximately half of the players will be favouring each decision (unpolarized or paramagnetic phase). For the region of the phase space with large enough social coupling (and given that the choice coupling k is positive in this example), a majority in both groups will be choosing the same outfit, and a majority of players will either be taking part in none or both of the events (completely polarised or ferromagnetic phase with both average magnetisation components having the same sign). That is except statistical fluctuations are small enough, and intensity of social interactions strong enough as compared to the desire to resemble or differentiate from the other group, or the additional reward possible from simultaneous participation. Then, a strong desire to conform to the same group could bring about situations in which the collective state contradicts the sign of k (additional pair of ferromagnetic solutions with components of the average magnetisation having different sign). These correspond, however, to metastable states, and will decay eventually as statistical fluctuations manage to get the system out of the local minimum.

It is also worth noting that for the non-local case there are metastable solutions for zero temperature, while this is not the case for the local model. This fact can also be easily checked analytically. It is actually easy to understand this distinctive characteristic intuitively as arising from the nature of the choice interdependence. In the completely deterministic case of zero temperature, without statistical fluctuations at all, in the individual interdependence scheme each player will know beyond any doubt how to align both their choices (in relation to one another, there will still be a pair of possible solutions). When there is group interdependence instead, although there is still going to be a collectively more convenient pair (the absolute minima of the free energy), rational individual players trying to maximise their payoff still will not be sure on how to best align their own choices. Hence the existence of metastable states even at extremely low temperatures.

The $k - T$ cross sections are shown in figure 9.15, again for both the non-local (a) and the local (b) choice coupling cases, for $J_s = 1$ and $J_t = 0.6$. Besides the breaking down of any stability in the strong coupling regime, here the most remarkable difference is the asymptotic behaviour of the critical curve which is present only in the local model, such that $\lim_{|k|\to\infty} K_b T_c = J_s + J_t$ (=1.6 in this case). In the individual interdependence case, at large enough intensity of this interdependence, the critical social permeability at which the system goes from a completely ordered to a completely disordered state remains approximately constant. This means



(a)



(b)

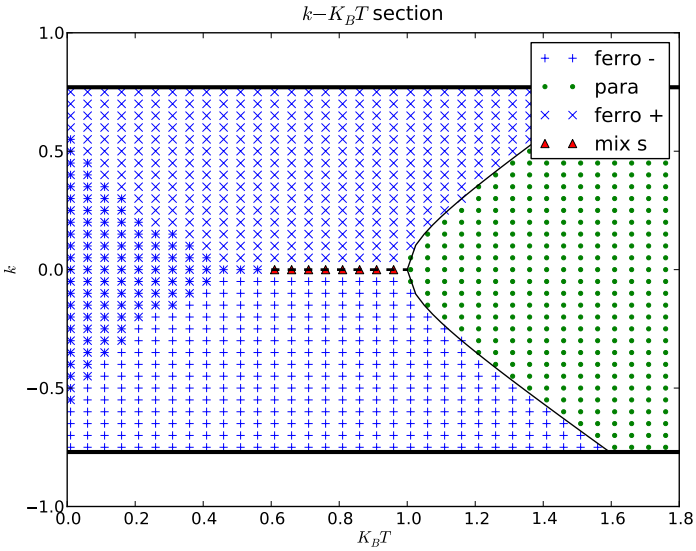
FIGURE 9.14: $J_t - \beta^{-1}$ cross-section for the non-local (a) and local (b) models for $J_s = 1$ and $k = 0.3$. Own elaboration using numerically computed solutions to the equations of state. The images have previously appeared in [71, 74].

that when this choice coupling k is very strong, small changes in its value will never bring about a qualitative change in the system. As has already been discussed multiple times, the non-local model is ill behaved precisely for large values of k, so the non existence of this asymptotic behaviour of the critical curve has no practical implications, as in the regions of strong inter-choice coupling there are no stable states at all (neither ordered or disordered).

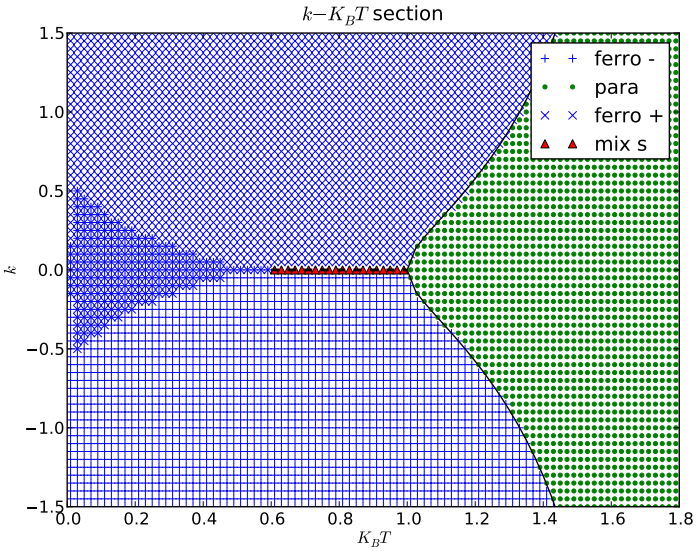
In fact, in the non-local model the degenerate and critical curves intersect at $J_s + J_t$, so whenever there is group interdependence and $K_B T > J_s + J_t$ if there is a stable solution it will be unpolarized. At has been discussed, the individual interdependence will also be unpolarized for $K_B T > J_s + J_t$ due to the asymptotic behaviour of the critical curve. Hence, in both cases, and as was to be expected, for large enough statistical fluctuations (or possibly *irrational* decisions), very large decision couplings will not bring about choice alignment. Note too that this is the region where, once non zero constant opinion fields or IWAs are considered, these will completely determine the alignment and direction of both choices, with no possibility of metastability, hysteresis or the abrupt changes associated with first order phase transitions.

Consider now the warriors and sorcerers are deciding in which of two in-game events to take part (non-local model). Game dynamics are such that, the more other players of the same group are participating, the more experience points they can expect to win. There is an additional payoff ($k > 0$) or penalisation ($k < 0$) for each player of the other group they encounter in the event. Players could also need to decide between two options for two different accessories (local model), in which matching two of these could bring the user an additional protection ($k > 0$), or less protection than a different combination ($k < 0$).

For the case of in-game event participation, stable solutions only exist if the reward/penalisation associated to the fraction of members of the other group taking part, is smaller than the benefit brought about by sharing event with their peers (weak coupling regime). In the case of accessory selection, if the additional protection brought (or prevented by) the item combination is large enough, further changes in this, or in the social temperature, will have no impact in the fraction of players selecting each item (asymptotic behaviour of the critical curve). In both cases, there is always a large region where, despite choice and/or social interaction, statistical fluctuations are large enough to prevent the former from having an impact in the outcome (completely unpolarized or paramagnetic phase).



(a)



(b)

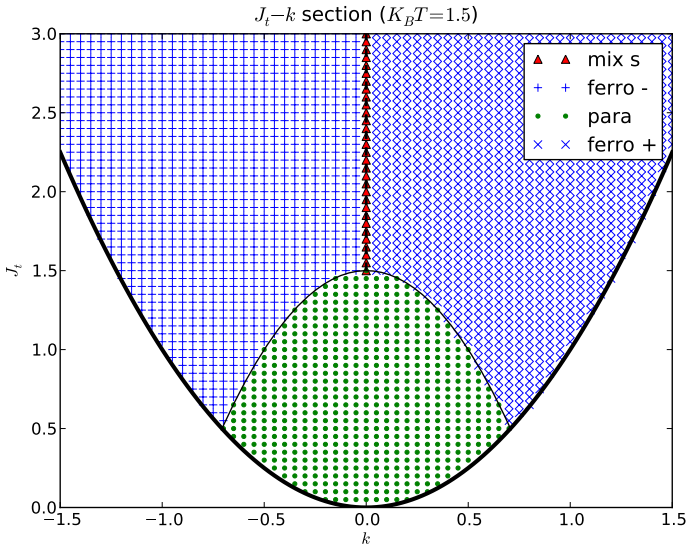
FIGURE 9.15: $k - \beta^{-1}$ cross-section for the non-local (a) and local (b) models for $J_s = 1$ and $J_t = 0.6$. Own elaboration using numerically computed solutions to the equations of state. The images have previously appeared in [71, 74].

This situation can represent player's free will, understood as their possibility to behave against to what seems to be their own interest. It could also codify fluctuating individual preferences that make some players prone to decisions different from those suggested by the couplings.

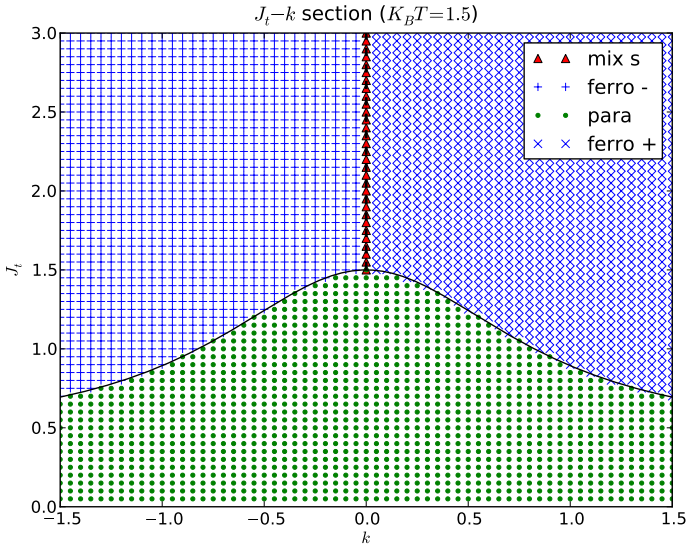
For both the local and non-local model, there is also a region of metastable states associated to the phase transition in inter-choice coupling (both ferromagnetic phases present). This means if, for example, game dynamics were shifting to a situation where sharing events with players of the other group is going from being rewarded to penalised; or when the combination of the two accessories is going from being the best objective option to not; *memory* of the previous situation could prevent the system from going to the best collective state immediately.

The $J_t - k$ (for $J_s = 1$) sections are shown at two different temperatures, for $K_B T = 1.5$ in figure 9.16, and $K_B T = 0.4$ in figure 9.17, as there are two possible qualitatively markedly different situations. In both figures (a) corresponds to the non-local model, and (b) to the local one. In both cases there are mixed segments. In both cases, always for the uncoupled $k = 0$ case, and for $J_t > K_B T = 1.5$ for high temperature ($J_s = 1$ smaller than 1.5) and $J_t < K_B T = 0.4$ for low temperature ($J_s = 1$ larger than 0.4). In both figure 9.16 (a) and figure 9.17 (b), the degenerate curve delimits the parabolic region where stable states exist. For high temperature (figure 9.16), the critical curve separates the paramagnetic region from that where there is a pair of ferromagnetic stable solutions that are equally probable, with relative sign between both components of the average magnetisation vector determined by the sign of k. As was just discussed for the $k - T$ sections, here too is there asymptotic behaviour in the local model ($\lim_{|k|\to\infty} J_t = K_b T_c - J_s$, which is 0.4 in this case) which is not present in the non-local case.

For low enough temperatures, and unlike in the uncoupled case, the unpolarized state is never a stable state, and there will only be partial disorder (in one of the choices) in the mixed segment described above. There is however a region of metastability for low enough $|k|$ and/or high enough J_t. This corresponds to cases in which, in all of the event and item choice examples that have been mentioned throughout this section, there is never the possibility of half of the players opting for each option (except for only one of the groups or choices and only on the mixed segment). Social and choice interactions always brings the system to an ordered state in this case.

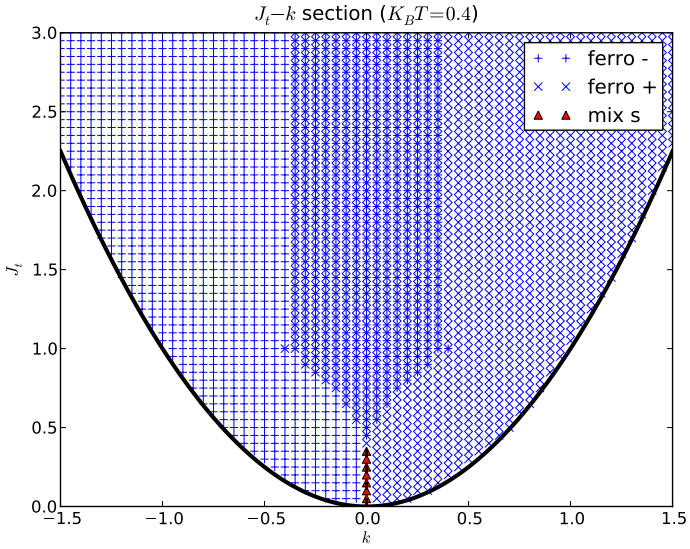


(a)

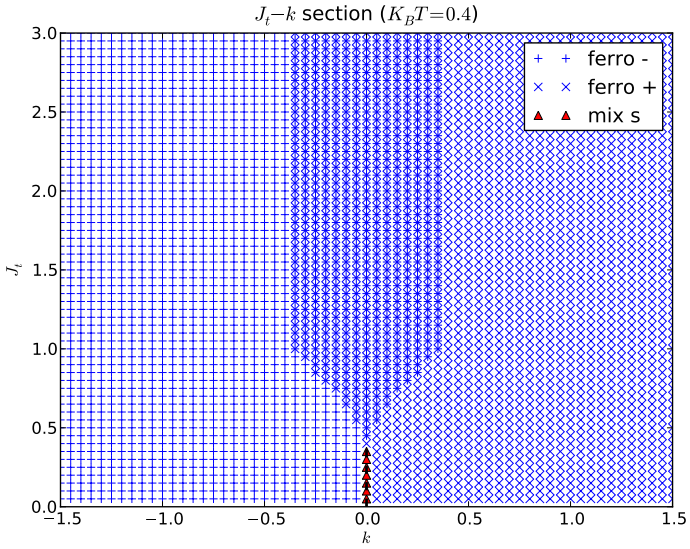


(b)

FIGURE 9.16: $J_t - k$ cross-section for the non-local (a) and local (b) models at high temperature for $J_s = 1$ and $\beta^{-1} = 1.5$. Own elaboration using numerically computed solutions to the equations of state. The images have previously appeared in [71, 74].



(a)



(b)

FIGURE 9.17: $J_t - k$ cross-section for the non-local (a) and local (b) models at low temperature for $J_s = 1$ and $\beta^{-1} = 0.4$. Own elaboration using numerically computed solutions to the equations of state. The images have previously appeared in [71, 74].

9.5 RATIONAL NON HOMOGENEOUS POPULATIONS

This section considers the individual interdependence scenario or local model when the IWAs follow a certain distribution among the population. This is the case of heterogeneous populations (and the general case in regards to choice making), and corresponds (everything else left the same in the Hamiltonian of equation 9.28) to two coupled RFIMs. As studying the system in its full complexity is challenging, the goal here is to draw some conclusions about the deterministic case, i.e., for zero temperature. This, as has been discussed, can be considered (depending on the interpretation given to the uncertainty associated to the social permeability) to represent completely rational individuals. As compared to what precedes, this section is concerned with *quenched disorder* (fixed in time) as opposed to the *annealed disorder* (evolving in time) associated to finite temperature of section 9.3.

The intention here is to outline what will happen to the conclusions described by Gordon et al. [124] for a single decision (which are summarised in chapter 2 section 2.10.3), when choice interdependence is introduced. For this reason, and although it may obscure the comparisons to the homogeneous populations case described till now, the same notation, conventions and terminology that in the aforementioned paper [124] (which was also introduced in chapter 2 section 2.10.3) will be used. Note that, as mentioned in 2 section 2.10.3, the socioeconomic interpretation of equilibrium states is that of Nash equilibria in non-cooperative games.

Let us therefore consider a system made up by N players, in which each individual i is making two simultaneous choices w_i and u_i which can take values 0 or 1 (the latter signifying adoption). Both choices depend on each other with intensity given by the coupling K. Both are also affected by the player's perception of the decisions other users are making with strength J_{ij} for $j \in n_i$ and n_i the set of players with whom player i interacts, also referred to as their *neighbourhood*. The individual preference for the choices are given by each player's IWAs H_i^w and H_i^u, which are IID across the population. A price P_w and P_u for each can be included to facilitate the analysis when this is applied to purchases (and then H_i^w and H_i^u are in fact the IWPs, and the IWAs are given by $H_i^w - P_w$ and $H_i^u - P_u$), and these can be set to zero to recover the pure (non transaction related) decision case.

In a such a system, users can be understood to make choices that will maximise their utility, as given by the expression

$$\begin{aligned} U_i = &\left(H_i^w - P_w + \textstyle\sum_{j\in n_i} J_{ij}^w E_i(w_j)\right) w_i + \\ &+ \left(H_i^u - P_u + \textstyle\sum_{j\in n_i} J_{ij}^u E_i(u_j)\right) u_i + \\ &+ K w_i u_i \end{aligned} \tag{9.49}$$

where $E_i(w_j)$ and $E_i(u_j)$ represent player i's belief on how player j is deciding. This can also be rewritten as

$$U_i = S_i^w w_i + S_i^u u_i + K w_i u_i \tag{9.50}$$

with $S_i^a = H_i^a - P_a + J_a \mu_a$ the individual i's uncoupled surplus in a, and $\mu_a = \frac{1}{N}\sum_i a_i$ the fraction of adopters in a, with $a = \{w, u\}$. Note that as the deterministic case is under consideration, all *randomness* is contained in the IID random variables characterising individual preferences.

Note that, while considering the deterministic case, with no room for any social permeability, is probably an unrealistic approximation, its study can help shed light on general features brought about by choice interaction. In the same way as with the study of completely unbiased populations (section 9.4), which is also an unrealistic approximation for most cases of interest, it will provide insights on what the outcome will look like for situations that are approximately in this regime. It is also a needed first step before moving to study increasingly complex systems.

What follows will therefore be more relevant the more players can be considered to be always making completely rational decisions. Decisions are understood as rational in the sense that they aim at maximising utility, with utility fully determined by the choice random fields, and the social and choice couplings. A good approximation could be that of games with relatively simple dynamics (well defined set of rules on rewards or penalties to be expected depending on the interactions), and where the players are really invested in achieving the best outcome possible (if wins in the game are, for example, associated to rewards in real life).

In the thermodynamic limit $N \to \infty$, the expectation of any choice will be given by its average in the population (which under the current conventions is also the fraction of adopters). In equilibrium, these will be given by the system of fixed point equations:

$$\begin{aligned} \mu_w &= \frac{\mathcal{P}(S_i^w > 0) + \mathcal{P}(S_i^u > 0)\Phi_i^w}{1 - \Phi_i^w \Phi_i^u} \\ \mu_u &= \frac{\mathcal{P}(S_i^u > 0) + \mathcal{P}(S_i^w > 0)\Phi_i^u}{1 - \Phi_i^w \Phi_i^u} \end{aligned} \tag{9.51}$$

where $\Phi_i^a = \mathcal{P}(S_i^a + K > 0) - \mathcal{P}(S_i^a > 0)$.

If H_a is the average IWA in a, and $\frac{H_i^a - H_a}{\sigma_a} \sim f_a(x)$ ($f_a(x)$ has zero mean and variance 1) then

$$\begin{aligned}\mathcal{P}(S_i^a > 0) &= \mathcal{P}(H_i^a - H_a > -S_a) = \mathcal{P}(h_i^a - h_a > -s_a) \\ &= 1 - \int_{-s_a}^{\infty} f_a(x)dx \end{aligned} \tag{9.52}$$

and

$$\begin{aligned}\Phi_i^a &= \begin{cases} \mathcal{P}(-k_a < s_i^a < 0) & \forall K > 0 \\ -\mathcal{P}(0 < s_i^a < -k) & \forall K < 0 \end{cases} \\ &= \begin{cases} \mathcal{P}(-s_a - k_a < h_i^a - h_a < -s_a) & \forall K > 0 \\ -\mathcal{P}(-s_a < h_i^a - h_a < -s_a - k_a) & \forall K < 0 \end{cases} \\ &= \int_{-s_a - k_a}^{-s_a} f_a(x)dx \end{aligned} \tag{9.53}$$

with S_a the average uncoupled surplus in choice a, and $s_i^a = \frac{S_i^a}{\sigma_a}$, $s_a = \frac{S_a}{\sigma_a}$, $k_a = \frac{K}{\sigma_a}$ normalised variables, measuring surpluses and decision coupling in terms of the relevant scales of the a IWA distribution.

Extracting conclusions about generic properties from equations 9.51, 9.52 and 9.53 is challenging, but particular cases of interest can be studied and their phase diagrams explored. To illustrate this, the example where one of the IWAs follows a logistic distribution, and the other a delta one (i.e., is constant across the population) will be analysed. Note the case where both IWAs are delta distributed around zero has actually already been considered, as it is given by the $T = 0$ cross section of the local model's phase diagram described in 9.4.

As was already briefly discussed in chapter 2 section 2.10.2, the logistic distribution appears naturally when discussing binary decision problems. If the difference in random shocks associated to the two different choices is logistically distributed, the system can be understood as being in statistical equilibrium at temperature T (inverse of the logistic parameter). In the deterministic case considered here, this temperature is zero. Logistically distributed preferences thus correspond to players making decisions according to a Boltzmann-Gibbs distribution.

In this case, however, the logistic distribution is used to characterise the quenched (rather than annealed) disorder. This can be a good choice

for describing smoothly continuously distributed random variables. Its similarity to a normal distribution, but with fatter tails, makes it appropriate to describe distributions emerging from some human processes. Examples of its usage in the literature in economic and demographic contexts can be found as early as the first half of the 19th century [283, 284]. It is also the specific example analysed in [124], of which this section intends to be the extension to two coupled choices.

9.5.1 *An example: logistic and delta distributions*

Consider for example a role card game, in which players are aware of the different choices (at least across the population) and scores of the other players. Imagine players can purchase (be that in real money or in-game currency) cards. Consider two such items. One of them provides the same benefit to all players, such as for example an automatic win against any player. The other provides a different relative advantage depending on the other cards the user already has. Acquiring both items together additionally penalises or favours the player. The relative advantage provided by the second card is distributed logistically amongst the population of players.

If H_i^w follows a logistic distribution across the population of players, while H_i^u takes the same value for all users, the joint probability distribution function (PDF) is:

$$(H_i^w - H_w, H_i^u - H_u) \sim f(x,y) = \sigma_w f_w(x) f_u(y) \tag{9.54}$$

where the marginal PDFs are given by

$$\begin{aligned} f_w(x) &= \frac{\pi}{4\sqrt{3}} \frac{1}{\cosh^2\left(\frac{\pi x}{2\sqrt{3}}\right)} \\ f_u(y) &= \delta(y) \end{aligned} \tag{9.55}$$

and the marginal cumulative distribution functions (CDFs) by

$$\begin{aligned} F_w(s) &= \tfrac{1}{2} + \tfrac{1}{2} \tanh\left(\frac{\pi s}{2\sqrt{3}}\right) \\ F_u(t) = \Theta(t) &= \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases} \end{aligned} \tag{9.56}$$

The system of equations of state 9.51 will then take different analytical forms depending on the values of the average uncoupled surplus in u and the sign of the choice interdependence. When $K > 0$, there are three possibilities:

1. *Large average uncoupled constant surplus* $S_u > 0$:

$$\begin{aligned} \mu_w &= \tfrac{1}{2} + \tfrac{1}{2}\tanh\left(\tfrac{\pi(s_w+k_w)}{2\sqrt{3}}\right) \\ \mu_u &= 1 \end{aligned} \tag{9.57}$$

2. *Intermediate average uncoupled constant surplus* $-K < S_u < 0$:

$$\begin{aligned} \mu_w &= \frac{\frac{1}{2}+\frac{1}{2}\tanh\left(\frac{\pi s_w}{2\sqrt{3}}\right)}{1+\frac{1}{2}\tanh\left(\frac{\pi s_w}{2\sqrt{3}}\right)-\frac{1}{2}\tanh\left(\frac{\pi(s_w+k_w)}{2\sqrt{3}}\right)} \\ \mu_u &= \mu_w \end{aligned} \tag{9.58}$$

3. *Low average uncoupled constant surplus* $S_u < -K$:

$$\begin{aligned} \mu_w &= \tfrac{1}{2} + \tfrac{1}{2}\tanh\left(\tfrac{\pi s_w}{2\sqrt{3}}\right) = \mu_w^{(K=0)} \\ \mu_u &= 0 \end{aligned} \tag{9.59}$$

and when $K < 0$:

1. *Large uncoupled constant surplus* $S_u > -K$:

$$\begin{aligned} \mu_w &= \tfrac{1}{2} + \tfrac{1}{2}\tanh\left(\tfrac{\pi(s_w+k_w)}{2\sqrt{3}}\right) \\ \mu_u &= 1 \end{aligned} \tag{9.60}$$

2. *Intermediate uncoupled constant surplus* $0 < S_u < -K$:

$$\begin{aligned} \mu_w &= \frac{\frac{1}{2}+\frac{1}{2}\tanh\left(\frac{\pi(s_w+k_w)}{2\sqrt{3}}\right)}{1-\frac{1}{2}\tanh\left(\frac{\pi s_w}{2\sqrt{3}}\right)+\frac{1}{2}\tanh\left(\frac{\pi(s_w+k_w)}{2\sqrt{3}}\right)} \\ \mu_u &= \frac{\frac{1}{2}-\frac{1}{2}\tanh\left(\frac{\pi s_w}{2\sqrt{3}}\right)}{1-\frac{1}{2}\tanh\left(\frac{\pi s_w}{2\sqrt{3}}\right)+\frac{1}{2}\tanh\left(\frac{\pi(s_w+k_w)}{2\sqrt{3}}\right)} \end{aligned} \tag{9.61}$$

3. *Low uncoupled constant surplus* $S_u < 0$:

$$\begin{aligned} \mu_w &= \tfrac{1}{2} + \tfrac{1}{2}\tanh\left(\tfrac{\pi s_w}{2\sqrt{3}}\right) = \mu_w^{(K=0)} \\ \mu_u &= 0 \end{aligned} \tag{9.62}$$

The large and low constant surplus cases correspond to those where the IWP that is delta distributed (constant across the population) is large

or small (respectively) as compared to the relative strength of the logistically distributed IWP and the coupling between both choices, and the intermediate to those in between.

If the card that has identical effect for all players is not worth its price (at least as compared to other available options), in the case of completely rational agents under consideration, absolutely no player will purchase it. This corresponds to the low uncoupled surplus regime. Note that whether it is worth or not its price depends on the sign and intensity of the inter-choice coupling: if purchasing both cards together gives additional benefit, players will be ready to spend more money on acquiring it than if, on the contrary, users are penalised for the combined purchase. In both cases though, the choice making on the logistically distributed IWP becomes uncoupled, and the exact scenario described in [124] recovered.

On the opposite side of the spectrum, if one of the cards provides an identical but massive playing advantage to all players, the coupled choice problem will operate in the large uncoupled constant surplus regime, in which absolutely all players acquire the card with constant IWP across the population. As was the case for the low uncoupled constant surplus case, what providing a large enough advantage is depends on the size and strength of the inter-choice coupling. In this case, the other choice behaves as it was uncoupled, but with fraction of adopters shifted to larger or smaller values (depending on whether a net benefit or penasilation comes from aligning both choices).

In the intermediate uncoupled constant surplus case, due to the inter-choice coupling, different players will make different choices, even when they are all behaving rationally and the card would provide identical benefit to all of them.

The phase diagrams for the different regimes are described below in some detail. As will be discussed, the are two main takeaways. Remarkably, demand of the card with logistically distributed value will behave qualitatively as in the uncoupled case (with the regions of different behaviour shifted to different values of the order parameters depending on the specific configuration). The card with identical value for all players, does qualitatively change its demand behaviour for some region of the phase space (also determined by the specific values and social strengths considered).

On one hand, as was the case for uncoupled choices, the fraction of adopters for the logistically distributed IWA (the demand for the card with different relative value for different players in the example of choice), will have high demand, low demand, or two coexisting Nash equilibria

corresponding to low and high demand, depending on the exact logistic distribution, the strength of social influence, the strength of the constant opinion field (identical value across players provided by the other card), and the sign of the inter-choice interaction (whether it is beneficial or detrimental to players to buy them simultaneously). This is qualitatively the case in both the low uncoupled surplus case (figure 9.18), of the high surplus one regardless of the sign of k (figures 9.19 and 9.20), and of intermediate surplus, also regardless of the sign and absolute value of k (figures 9.21, 9.22 and 9.23). The only difference between these cases is the exact region of the phase diagram where there is prevalence of high and low demand, or they both coexist.

On the other hand, in regards to the other card, for many regions of the phase diagrams, the situation resembles that of the uncoupled case, with either all or no players adopting. However on an another region with extent depending on its value, social influence strength and inter-coupling, its demand will be linked to that of the logistically distributed value card, and hence it will have partial adoption in the population (figure 9.24).

9.5.1.1 *Phase diagram $h_w - p_w$ vs j_w section for low uncoupled constant surplus*

This region of the parameter space is defined differently depending on the sign of K such that

$$\begin{aligned} H_u - P_u &< -J_u - K \quad \forall K > 0 \\ H_u - P_u &< -J_u \quad \forall K < 0 \end{aligned} \tag{9.63}$$

Here the equations of state then take the form:

$$\begin{cases} \mu_w = \frac{1}{2} + \frac{1}{2} \tanh\left(\frac{\pi s_w}{2\sqrt{3}}\right) = \mu_w^{(K=0)} \\ \mu_u = 0 \end{cases} \tag{9.64}$$

This means when the constant IWA is low enough, the problem has little interest in that choice (all players will be deciding against). The choice associated to the logistically distributed IWA behaves as if there was no decision interdependence. The $h_w - p_w$ vs j_w section of the phase diagram is shown in figure 9.18 [2]. Regions where there will be a high demand in w ($\mu_w > 0.5$) are coloured dark green, and those with low adoption ($\mu_w < 0.5$)

2 Note the similarity to figure 4 in [124], which illustrated the single choice phase diagram section.

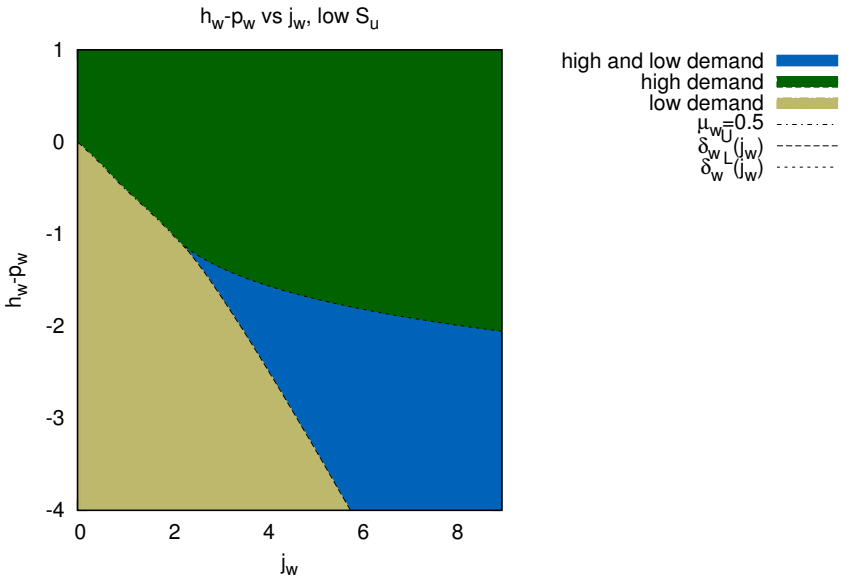


FIGURE 9.18: Phase diagram $h_w - p_w$ vs j_w section for low uncoupled constant surplus S_u. Dark green is used for the region where $\mu_w > 0.5$ (high demand), and dark yellow for those where $\mu_w < 0.5$ (low demand), with the segment dividing both where it is exactly 0.5 plotted as a dashed-dotted line. The blue region corresponds to a region where two possible solutions exist, one of each. A phase transition between both regimes takes place at the dashed lines δ_{w_U} and δ_{w_L}. For all regions $\mu_u = 0$. Own elaboration.

dark yellow. Blue corresponds to a different phase, where there are two Nash equilibria, one corresponding to high and one to low adoption.

9.5.1.2 *Phase diagram $h_w - p_w$ vs j_w section for large uncoupled constant surplus*

This region is defined by the inequalities

$$\begin{aligned} H_u - P_u &> -J_u \quad \forall K > 0 \\ H_u - P_u &> -J_u - K \quad \forall K < 0 \end{aligned} \tag{9.65}$$

and the system of equations of state can be here expressed as

$$\begin{cases} \mu_w = \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{\pi(s_w+k_w)}{2\sqrt{3}}\right) \\ \mu_u = 1 \end{cases} \tag{9.66}$$

This is in many ways qualitatively similar to the situation described above for the low uncoupled constant surplus. Namely, both fractions of adopters become for the most part uncoupled. The choice associated to the logistically distributed IWA does not depend on the other decision, but its expression does contain the strength of the coupling K. The choice associated to the constant opinion field is, as in the low uncoupled constant surplus case, the same for all players, in favour in this case. The situation is depicted in figures 9.19 (positive choice interdependence) and 9.20 (negative one), which show the $h_w - p_w$ vs j_w section of the phase diagram in this case. It is very similar to the low uncoupled constant surplus case, with a shift in the $h_w - p_w$ axis. For $K > 0$, adoption will be higher for the same values of the average opinion field, and lower for $K < 0$.

9.5.1.3 *Phase diagram $h_w - p_w$ vs j_w section for intermediate uncoupled constant surplus*

This is the only region of the parameter space where both average choices remain, in fact, completely coupled. The system will be in this regime whenever

$$\begin{aligned} -K - J_u < H_u - P_u < -J_u \quad \forall K > 0 \\ -J_u < H_u - P_u < -J_u - K \quad \forall K < 0 \end{aligned} \tag{9.67}$$

In this case, the equations of state also depend on the sign of the choice interrelation, such that for $K > 0$:

$$\begin{cases} \mu_w = \frac{\frac{1}{2}+\frac{1}{2}\tanh\left(\frac{\pi s_w}{2\sqrt{3}}\right)}{1+\frac{1}{2}\tanh\left(\frac{\pi s_w}{2\sqrt{3}}\right)-\frac{1}{2}\tanh\left(\frac{\pi(s_w+k_w)}{2\sqrt{3}}\right)} \\ \mu_u = \mu_w \end{cases} \tag{9.68}$$

and for $K < 0$:

$$\begin{cases} \mu_w = \frac{\frac{1}{2}+\frac{1}{2}\tanh\left(\frac{\pi(s_w+k_w)}{2\sqrt{3}}\right)}{1-\frac{1}{2}\tanh\left(\frac{\pi s_w}{2\sqrt{3}}\right)+\frac{1}{2}\tanh\left(\frac{\pi(s_w+k_w)}{2\sqrt{3}}\right)} \\ \mu_u = \frac{\frac{1}{2}-\frac{1}{2}\tanh\left(\frac{\pi s_w}{2\sqrt{3}}\right)}{1-\frac{1}{2}\tanh\left(\frac{\pi s_w}{2\sqrt{3}}\right)+\frac{1}{2}\tanh\left(\frac{\pi(s_w+k_w)}{2\sqrt{3}}\right)} \end{cases} \tag{9.69}$$

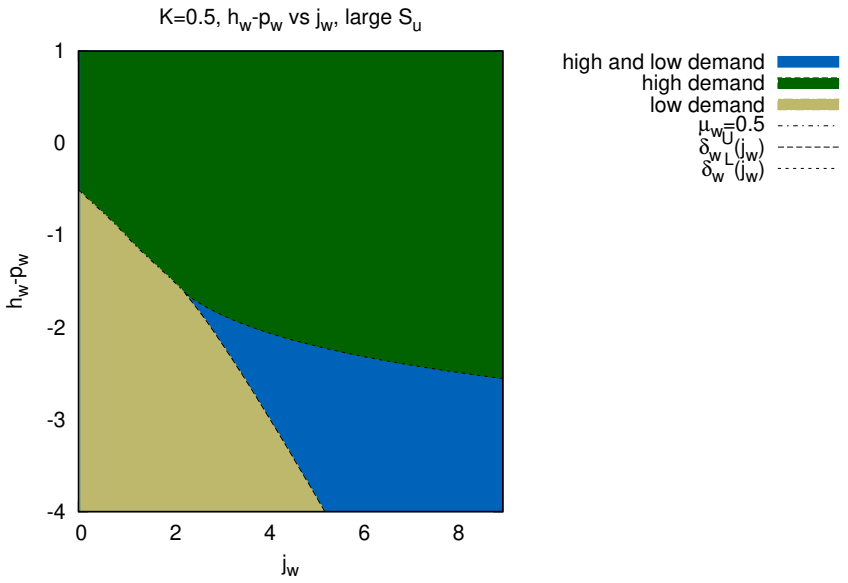


FIGURE 9.19: Phase diagram $h_w - p_w$ vs j_w section for high uncoupled constant surplus S_u for positive interdependence $K = 0.5$. Dark green is used for the region where $\mu_w > 0.5$ (high demand), and dark yellow for those where $\mu_w < 0.5$ (low demand), with the segment dividing both where it is exactly 0.5 plotted as a dashed-dotted line. The blue region corresponds to a region where two possible solutions exist, one of each. A phase transition between both regimes takes place at the dashed lines δ_{w_U} and δ_{w_L}. For all regions $\mu_u = 1$. Own elaboration.

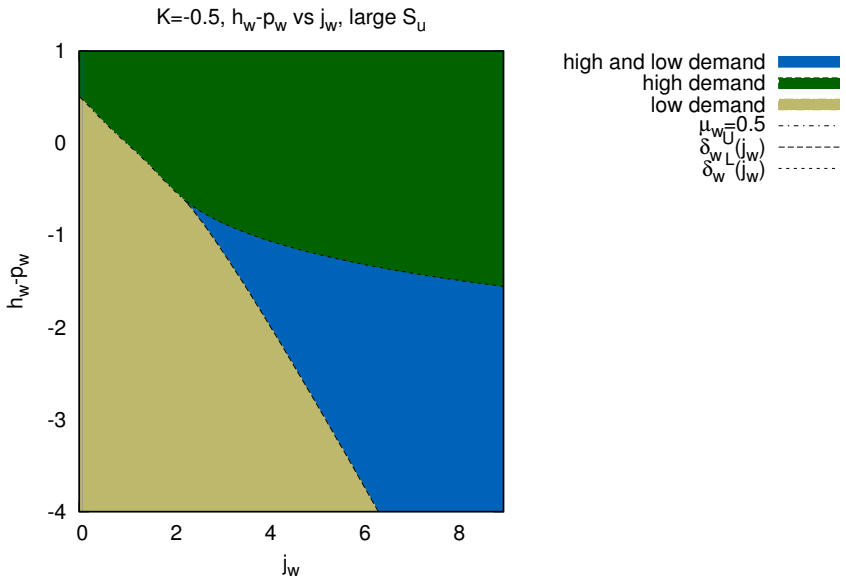


FIGURE 9.20: Phase diagram $h_w - p_w$ vs j_w section for high uncoupled constant surplus S_u for negative interdependence $K = -0.5$. Dark green is used for the region where $\mu_w > 0.5$ (high demand), and dark yellow for those where $\mu_w < 0.5$ (low demand), with the segment dividing both where it is exactly 0.5 plotted as a dashed-dotted line. The blue region corresponds to a region where two possible solutions exist, one of each. A phase transition between both regimes takes place at the dashed lines δ_{w_U} and δ_{w_L}. For all regions $\mu_u = 1$. Own elaboration.

Note in both cases the expression on the average choice w (with a logistic distribution of the opinion field) has a more complex expression than the uncoupled case, while the average choice u (with constant opinion field) completely depends on the uncoupled surplus on the other decision. In particular, for positive interdependence both average choices are always the same.

Using numerically computed solutions (using the Newton-Raphson algorithm) to 9.68, the phase diagram section of figure 9.21 can be plotted. Analogously, for negative inter-coupling and system of equations of state 9.69 one arrives at figure 9.22. Both show that in fact, the behaviour is still qualitatively very similar to the uncoupled case, with regions of interest now shifted, not only on the $h_w - p_w$ axis but also on the j_w one. This is further explored in figure 9.23, which compares the region where multiple solutions exist for different values of the interdependence coupling K.

9.5.1.4 *Phase diagram $h_u - p_u$ vs j_u section*

Considering what has been discussed about the average choice with constant opinion field in the previous subsections, its behaviour in the phase diagram's $h_u - p_u$ vs j_u section is summarised in figure 9.24 for positive K. The low uncoupled constant surplus region is plotted in dark yellow, where all players decide for $u_i = 0$. In the large uncoupled constant surplus region, plotted in dark green, all players decide for $u_i = 1$. In between, for the intermediate case (plotted in white), the fraction of adopters μ_u is identical to that in the other choice.

The behaviour for negative K is very similar but shifted in the $H_u - P_u$ axis, and with μ_u in the intermediate uncoupled constant surplus region not identical to μ_w, although still completely determined by the dynamics in that decision. Regardless of the sign of K, for negative average IWA, there can be full demand, no demand, or something in between, depending on the value of the social coupling J_u. For positive choice interdependence, there will always be full demand (for all values of social pressure). For negative one, the fraction of adopters is always greater than zero, but there is only full adoption for large enough values of the social coupling.

9.6 SUMMARY AND CONCLUSIONS

The object of this chapter was markedly different from the rest of the thesis. It is still an attempt to formalise and systematise what can be learned about human behaviour as is expressed in a video game. While most of the

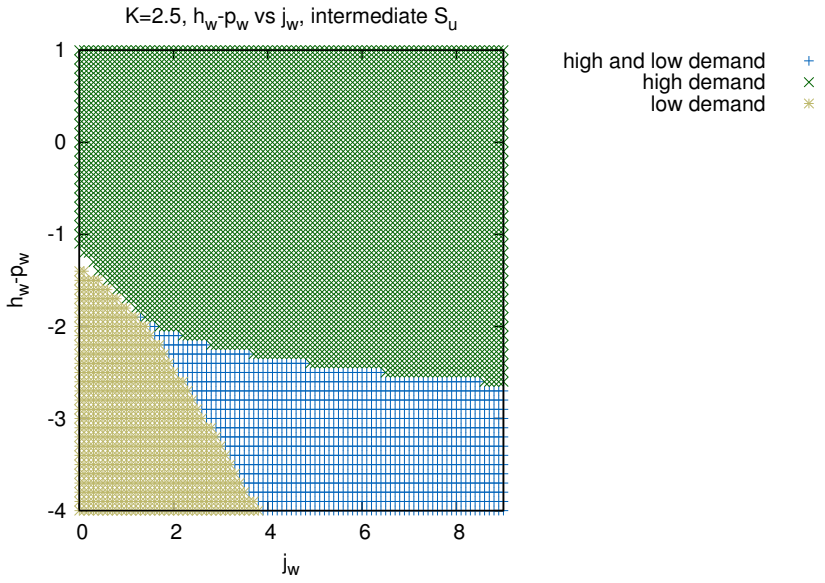


FIGURE 9.21: Numerically computed phase diagram $h_w - p_w$ vs j_w section for intermediate uncoupled constant surplus S_u and positive interdependence $K = 2.5$. Dark green asps (x) are used to represent $\mu_w > 0.5$ (high demand), and dark yellow stars (*) for $\mu_w < 0.5$ (low demand), and blue crosses (+) to two solutions, one of each. For all regions $\mu_u = \mu_w$. Own elaboration.

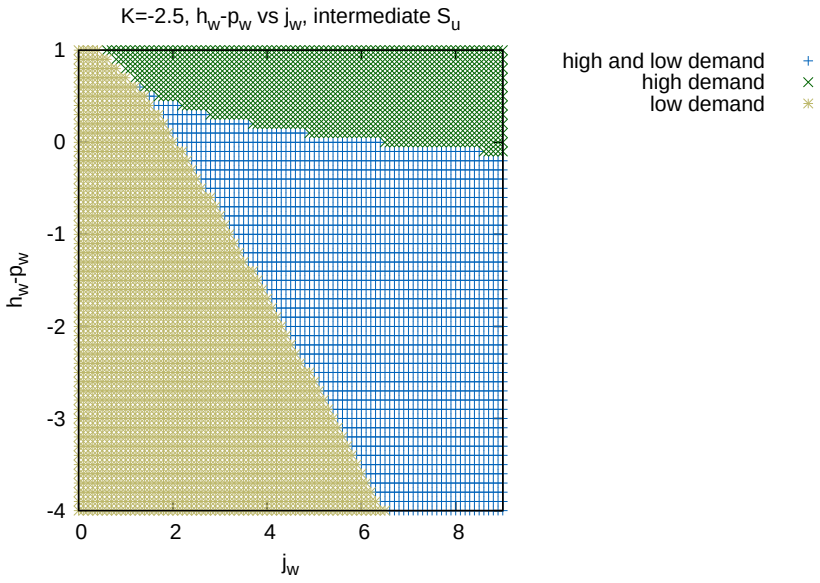


FIGURE 9.22: Numerically computed phase diagram $h_w - p_w$ vs j_w section for intermediate uncoupled constant surplus S_u and negative interdependence $K = -2.5$. Dark green asps (x) are used to represent $\mu_w > 0.5$ (high demand), and dark yellow stars (*) for $\mu_w < 0.5$ (low demand), and blue crosses (+) to two solutions, one of each. Behaviour of μ_u is not contemplated in this plot. Own elaboration.

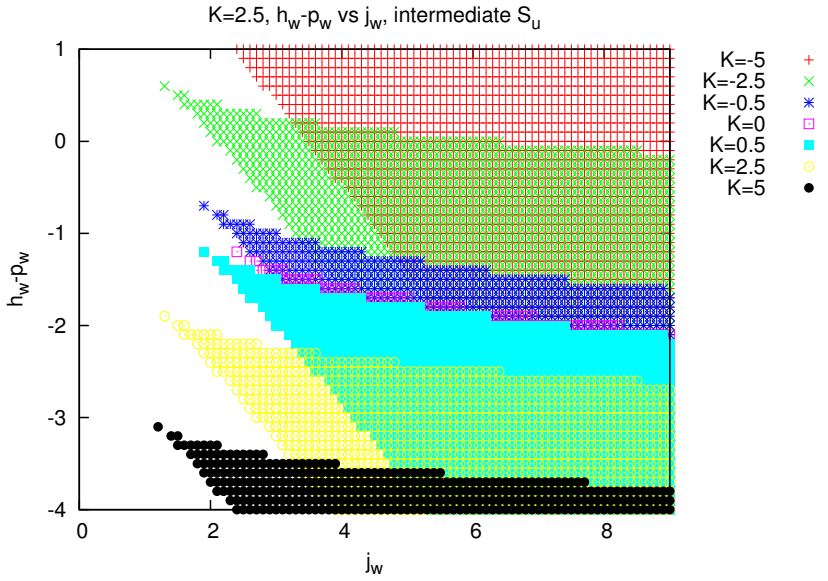


FIGURE 9.23: Phase diagram $h_w - p_w$ vs j_w section: multiple solution region for different values of K for intermediate uncoupled constant surplus. Own elaboration.

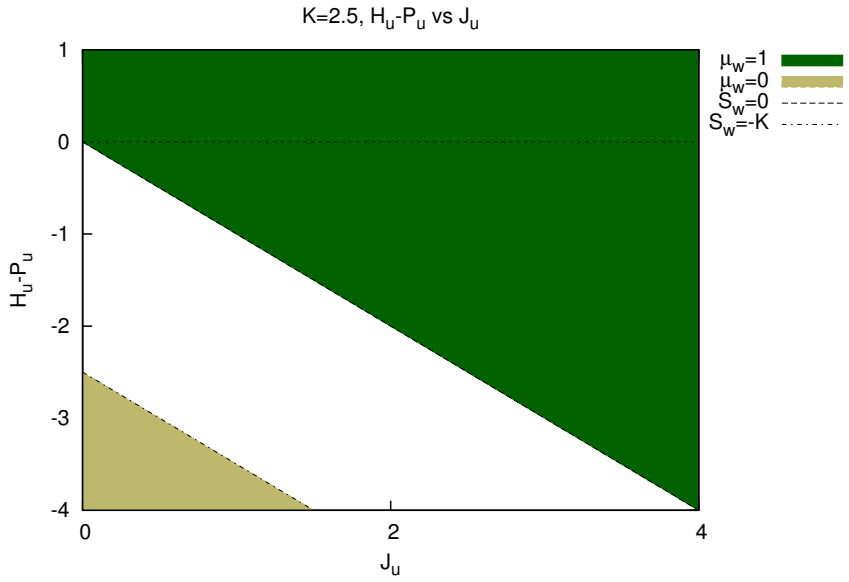


FIGURE 9.24: Phase diagram $H_u - P_u$ vs J_u section for positive interdependence $K = 2.5$. Dark green is used for the region where $\mu_u = 1$ (full demand, corresponding to large uncoupled constant surplus), and dark yellow for those where $\mu_u = 0$ (no demand, corresponding to low uncoupled constant surplus). The white region corresponds to the intermediate uncoupled constant surplus region, where $\mu_u = \mu_w$. Behaviour of μ_w is not contemplated in this plot. Own elaboration.

models described before were applied with the goal in mind of *quantifying* and predicting behaviours, in this case the focus was on understanding *qualitatively* how certain apparently complex collective behaviours can emerge due to the social nature of human systems. This can help explain the global outcome of a concrete decision process, up to discussing the level of success of a video game, in terms of the strength of social interactions, statistical fluctuations, and on how the game dynamics make different player choices depend on each other.

In particular, social games where players make choices which are affected by what they see other players are deciding, have been considered. It has been shown how the interaction between different choices further enriches the picture described for single decisions in chapter 2 section 2.10. The analysis was carried out using two coupled mean field Ising models. When the two choice variables are coupled through infinite range interactions, the model mimics two groups of players (with specific in-group social pressures to conform) whose decision is also impacted by the average choice of the other group. If the choice coupling is done through every individual, then the model resembles a single group of players in which each of them has to decide themselves on two different matters that affect each other. As has been described throughout the chapter, this methodology can be applied to understand the features arising from player interactions in what concerns, from broad groups and generic choices that can help to qualitatively understand churn and purchase dynamics, to very specific ones concerning particular groups and/or reactions to content.

Note that while the methodology used is valid for systems in statistical equilibrium, the equilibria described are not static. The individuals will still typically be changing their minds about one or both of the choices, and what becomes fixed are order parameters which reflect the collective outcome, in terms of the average decisions and thus fractions of adopters. Even though in many cases it is to be expected that the choice making processes in video games will be evolving out of equilibrium, in many others, statistical equilibrium -or slow evolution from a state of dynamic equilibrium to a different one- will be a good approximation. The conclusions drawn in this chapter can be then used to hypothesise about the origin of differences in the average outcomes of different similar games, or of the same game at different moments in time.

A detailed analysis, both analytical and numerical, was carried out for homogeneous unbiased populations, i.e., for cases in which the players have no particular inherent preference guiding their choice making, and

their perception of what is their best interest depends solely on the rewards (abstract or concrete) provided through social imitation and choice interaction. This can be the case for some decisions that can be understood to be governed by trends. For example, if players have to pick between two items of the same characteristics, or on taking part in one of two similar events, the problem could very well be approximated with zero opinion fields. When considering the critical choice of playing or not, if the focus is in players that are very kin in that particular type of game (e.g. puzzle games), it might be a good way of approximating the problem of how they pick between two competing ones. In any case, the study of homogeneous unbiased populations would be the starting point for more complex IWA distributions. Studying the zero field case is enough to know in which regions of the parameter space the social interactions will play a determinant role in the collective outcome once the fields are turned on.

Let us examine, in order to illustrate the discussion, two particular examples. One, that will portray the use of the non-local model, will be that of players of low and high level taking part (or not) in a particular event in the game (instead of another of the same type), in which they can combat other players to gain experience points. For the non local case, the example considered is that of users at a certain stage of the game which need to pick among two swords, and two shields. In this last example, if one of the swords and one of the shields are somehow *matching*, choice interdependence would be favouring to either pick both or none (positive choice interdependence). In the event case, it depends on how the game is designed. It is assumed players will always imitate to some extent users of similar skills, as this would provide opportunities for well balanced combats. If, when there is a significant level mismatch, the inexperienced player still has some chance of winning, and/or if they can learn new skills or earn a significant amount of experience points even when loosing against more advanced players, users will also be attracted by events where there are players of the other group (positive choice interdependence). If the level largely determines combat success, and rewards for losing, or for winning a much less experienced player, are negligible, players will actually avoid events where a large part of their possible opponents don't have approximately matching level (negative interdependence).

Qualitatively, and even quantitatively, both the group and the individual interdependence scenarios generate similar outcomes. As compared to what would happen in the case of unbiased populations if the two choice variables were uncoupled, the interactions will have a significant impact in

the average choice for larger statistical fluctuations (or social temperature) and/or smaller strength of the social pressure or player desire to imitate their peers.

For choice interdependence large enough (in absolute value), the relative alignment between both choices is always that suggested by the variable interrelation. This means, either a majority of players pick both matching items, or none of them, in the individual interdependence example. In the event participation case, a majority of both experienced or inexperienced players will take part, or not, when the interrelation is positive. When it is negative, it will attract a majority of high level players but a minority of low level ones, or the other way round. This of course, unless statistical fluctuations are large enough, or couplings within the same choice low enough, that there is disorder in both decisions. In that case, all players are basically deciding randomly, so half of them end up deciding in favour and half of them against. In regions close to the transition values, small changes in the strength of the social pressure or the choice interdependence could bring about quick and drastic changes in the collective outcome.

When the inter-choice coupling becomes very large however, no stable equilibrium is arrived at for the non-local model. In the example, if players are basing their choice on which event to take part in, more on their perception of what players of a very different level, rather than a similar one, are doing, there are no optimal average values that will maximise global player satisfaction. This could be easily prevented by making sure that the expected incentives to fight an equivalently experienced player are, in average, larger than the possible mean benefits or hindrances of a mismatched encounter.

While this is not the case for the local model, and there is always a state of equilibrium, once the interdependence is strong enough, increasing it more will never bring about a qualitative change (phase transition) in the state of the system. Considering the matching sword and shield example, once this matching is obvious enough to enough players, most of them will only pick them together if at all. Making them even more alike will not change the situation (if social permeability and the drive to imitate their peers don't change very much).

Phenomenologically speaking, probably the most interesting case is that of small relative choice interdependence. In this region, small changes in the strength of this coupling can make the system transition from order to disorder. There can also be additional unstable equilibria where the relative alignment of both choice variables is not the one dictated by their interrela-

tion. Considering the game event example, this means that, particularly if game dynamics are changing and, for example, rewards associated to mismatched combats becoming penalisation's (on average), the system could get stuck in a metastable state where the majority of both inexperienced and experienced players are picking the same event (against their collective and individual best interest). With time (or larger fluctuations) however, the system will eventually decay to real equilibrium, in which most of the users of each group will pick different events. In regards to the item example, when the matching or alikeness of both objects is not obvious, rapid drastic changes should be expected when modifying this item affinity. There could even be situations where a majority of players are picking the specific sword but not its matching shield.

The individual interdependence scenario was also investigated for heterogeneous populations in the deterministic case. In particular, the case where one of the opinion fields is constant, and the other follows a logistic distribution has been studied in some detail. Let us consider how it would apply to the previous combined item example, where each player at a certain stage in the game needs to pick between two swords and two shields, when one of the shields is aesthetically matching one of the swords. Imagine one of the swords is objectively better than the other in the same way for all players (it inflicts more damage upon a successful blow, for example). We can then consider the IWA for picking sword of interest constant, negative if it is worse than the other, positive otherwise. If the problem under study is formulated in terms of picking the sword of interest and the matching shield, the interdependence between both choices is positive. Instead of constant and zero, as in the case of homogeneous unbiased populations discussed above, the IWA in the matching shield is now logistically distributed. This could be the case, for example, in games where a lot of stress is put on the looks of the character. If both shield form which players can choose are objectively the same in regards to the protection offered, the distribution of IWA relates to the user's taste, and how aesthetically pleasing they find that particular shield.

The first remarkable observation is that there is always going to be a range of values for the social imitation strength concerning shields, and the average preference in the group, where there is going to be two possible and equally likely equilibria, one in which a majority of players are picking the shield of interest, and one in which they are not. The value of constant field, will determine what happens to the sword choice, but also where the region of high and low shield adoption, and of both being simultaneously

possible, lie in terms of the strength of the propensity to imitate the shields the others are picking, and the average of the preference for the shield of interest in the group.

If the sword of interest is much worse objectively than its counterpart, nobody will choose it. Players will stick pick the matching shield however, whenever the average associated IWA is positive, and even for negative values if the social imitation strength associated to the shield choice is strong enough. In fact, shield picking behaves in this case exactly as it would do if both choices were uncoupled, indicating the matching looks of both items plays absolutely no role whenever the sword is bad enough.

On the other hand, if the sword of interest is better than the other option, all players will pick it. This favours too the matching shield election. Its adoption will be higher than in the uncoupled case for the same value of the parameters, and the high demand region is shifted to lower values of average preference. Even with no social pressure to conform, and with a negative average IWA, more than half of the players will pick the matching shield.

The most interesting case is when the sword is worse, but by very little. In this region, the dynamics in sword selection become completely coupled to the shield choice, and the fraction of adopters will be the same in both items. The high adoption region is in this cases shifted to lower values of average shield IWA and propensity to imitate other's shield, and the more so the higher the value of the sword-shield interdependence. In a way, the existence of a matching shield compensates for the penalisation of choosing the worse sword, whenever the difference with the better one is small enough.

Note that while the discussion has been made comparing the convenience of both swords (or rather of picking the item of interest or not), this is always going to be measured in terms of the intra- and inter- coupling. It is however the difference between the relative usefulness of both swords what has been considered to determine the value of the constant IWA.

To sum up, this chapter has shown that the use of some models of condensed matter to understand the average outcomes in social games -in particular to study interdependent choice making- is promising. It provides tools to formally understand the impact the interaction between players (and decisions) has, to compare different games or moments in time of the same game depending on the strength of the interactions and how individual player preferences are distributed, and to identify situations were rapid abrupt changes in the collective behaviour are likely.

10

CONCLUSIONS

Rien dans la vie n'est à craindre, tout doit être compris. C'est maintenant le moment de comprendre davantage, afin de craindre moins.

— Marie Curie

This thesis has shown that statistical and mathematical models can be successfully used to understand game dynamics and individual player behaviour, both in a quantitative data-driven fashion, and also from a more formal theoretical perspective to gain insights on the processes at play, particularly social ones.

Although other problems -such as LTV predictions (chapter 6 section 6.5) or content preference prediction (chapter 8) on the data-driven side, or interaction between different events, items or player groups in the quantitative approach (chapter 9)- have been considered, most of the content of this thesis revolves around understanding and predicting the level of login and purchasing activity of players. The evolution of these will determine the degree of success of any free2play game.

Quantitatively speaking, these have been studied through the modelling of conversion (non-PUs becoming PUs), churn (active players becoming inactive) and purchase churn (PUs becoming non-PUs). This work has shown that these quantities can be studied at different levels of aggregation in the game, ranging from how they behave when considering all players of the game, going down to looking at what is going on for individual users.

Hopefully, this thesis has persuaded the reader that it is useful to look at this problem at different scales. Not only do the results complement each other to give a well rounded outlook of the issue. Models at different aggregation levels can actually be used to enhance and enrich each other, as has already been pointed out and will be discussed in some detail below.

The proposed approach is summarised in figure 10.1. At the *macroscale* or game level, i.e., considering all players, conversion and churn can be studied from a time series perspective. In particular (as discussed in chapter 4), classical SSM approaches can be a useful tool to predict the evolution of these, and to quantify the impact of internal and external events on the them. Laying at the other side of the spectrum, it can be checked whether

individual players are still active or not, and whether they are purchasing or not, and models built to predict what will happen in the future, in what can be referred to as the *microscale* or player level. A lot of results have been presented and discussed concerning the use of survival models to this effect (in chapters 5 and 6), showing they can be used to produce accurate predictions on after how many days since first login, accumulated playtime, and at which level, each player will undergo conversion, purchase churn and churn.

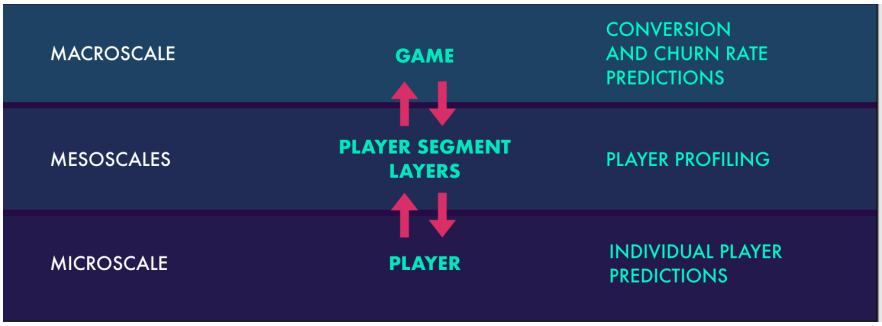


FIGURE 10.1: Schematic representation on how conversion and churn can (and should) be studied at different scales in the game.

In between these scales, player profiling comes in, as one can look at the problem in layers of increasingly complex segmentation landscapes, in the so called *mesoscales*. Closer to the game level, the same SSM approaches can be used (as discussed in chapter 4). This is particularly useful when trying to understand game dynamics, and how different types of players are affected differently by different in-game planning strategies. These are the scales video game studios should be concerned with when trying to answer questions regarding which in-game events to run and when, or what type of marketing and promotion campaigns are more effective. Closer to the microscale, ML models can be run for small cohorts of users, using similar approaches as those discussed at the player level. These are the critical scales when the aim is to develop personalised games in which players can be targeted individually, with content tailored to their specific likes.

Figure 10.2 summaries how, not only do the different scales provide complementary insights depending on the specific goal, they can actually be reconciled and used to enhance each other.

In a top-down approach, when moving from the macroscale to less aggregated levels, it has already been discussed that the same methodologies can

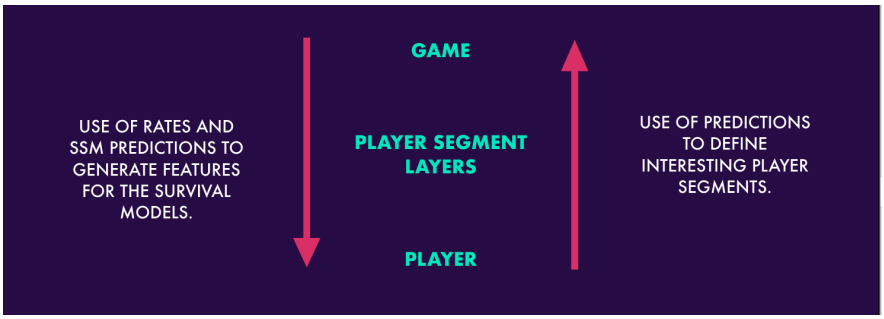


FIGURE 10.2: Schematic representation on how the study of conversion and churn at one scale can be used to complement and enrich the others.

be used for increasingly diverse groupings of players. The need of human expertise and time needed to model these is however quite extensive. In addition to this, the rates, predictions and estimated effects from scales close to the macro, can actually be useful at the player level, by using them to build additional features for the microscale models. For example, when studying non-PU to PU conversion, the time series of past and predicted conversion rate could be used to build features that would already incorporate calendar, campaigns and in-game event effects. Additional features characterising the individual player could then be understood to *correct* the average group behaviour particularising the prediction for each user. The first tests ran in this sense were promising.

When considering how the individual predictions can be used in a bottom-up approach, two possibilities emerge. One can try to generate predictions moving up the mesoscales by aggregating individual predictions, building a hierarchical model that yields consistent results at all levels. Another interesting approach has already been hinted in the thesis (chapter 7), and consists in using the predictions to generate meaningful player segmentations on which to apply the macro-type models. Their predictions can then be used to enrich the individual player models, closing the circle. Note that, while in this thesis the method was only outlined, giving some basic examples using churn predictions, these could be combined with conversion and purchase churn predictions, allowing for an extremely fertile player characterisation.

Besides the use of data driven techniques, the last part of this thesis has illustrated how models coming from statistical mechanics can be used to qualitatively understand how the social nature of some games can affect the

average outcome of player decisions, ranging from the very generic about how often to play or purchase, to very specific ones regarding which events, actions, items... to pick. In particular, the interaction between different choices or groups has been studied in detail for some particular cases.

The main takeaways here follow the same lines as the single choice/group case, that had already been studied in the literature in social science contexts. Namely, that the propensity to imitate others or social pressure, can impact the collective outcome both quantitatively and qualitatively, inducing sometimes part of the population to decide against their best interest. The equivalent to phase transitions can take place, meaning the system can change drastic and quickly with the onset/ending of significant social or choice interactions. They can also account for the existence of multiple equilibria for some regions of the parameter space. This multiplicity of possible states of the system, is often related to metastable states associated to first order phase transitions (in the opinion fields or the choice coupling). This means that, although some of the sudden shifts in the average state that have been described, will be reverted if the changes in the parameters are undone (when they are associated to second order or continuous phase transitions), in the cases were the system leaves a metastable state, the change will be irreversible (hysteresis).

This thesis has been an attempt to use video games as a play field to take steps towards a more scientific body of knowledge concerning human behaviour. The quantitative data driven models play the role of empirical laws, defined *ad hoc* to fit the data. The purely theoretical formal approach is of assistance to conceptually understand the mechanisms that give rise to some of the collective properties observed. It would be desirable, on the one hand, to make the empirical laws richer, capturing the dependence with all relevant variables, thus becoming more explanatory and precise. On the other hand, to move to complex enough formal theories, that would be able to go beyond their toy model character, and make quantitative predictions that can be compared to observations. The end goal would be to bridge the difference between both, providing a sound quantitative and qualitative understanding of the main processes at play, and their associated uncertainties.

10.1 SUMMARY OF MAIN CONTRIBUTIONS

This section provides an schematic summary of the main original contributions of this thesis. The reader is referred to the list of publications 6 for details on where and when the content was published.

10.1.1 *As first author*

1. (Chapter 4) A framework in which state space models are used to understand and predict the evolution of churn, purchase churn and conversion in a game, and in which the impact of game planning and the effect of external events can be assessed.
2. (Chapter 7) The use of predictions for individual players of expected lifetime, playtime, in-game progression and outlay to understand game dynamics and to assist meaningful player profiling.
3. (Chapter 9) Study of collective outcomes of interdependent decision making processes in social video games using statistical physics.

10.1.2 *As second or third author*

1. (Chapter 6, section 6.1) Framework to predict PU potential and days, playtime and level to conversion using survival models.
2. (Chapter 5, section 5.3 & Chapter 6, section 6.4) Analysis of the impact of different churner and purchase churner profiles in the performance of binary classification and survival models for churn and purchase churn prediction.
3. (Chapter 6, section 6.5) Use of multilayer perceptron and convolutional neural networks to predict player lifetime value.

10.1.3 *Unpublished work to which the author contributed*

1. (Chapter 8) An item recommendation system for video games with large dimensional item spaces, using a combination of clustering methods, extremely randomised ensembles and collaborative filtering. The methodology is described but no concrete results presented, as the datasets used to carry out its development are not available for publication.

2. (Chapter 9 section 9.5) The case of collective outcomes of interdependent decision making processes with heterogeneous inclinations in social video games has not been previously published, but was presented in two conferences.

A

INTERDEPENDENT CHOICES: AUXILIARY PLOTS

Here the reader can find additional plots illsutrating how the magnetizations at the critical points of the free energy vary for different values of the parameters. The are intended to further clarify the behavior of the systems (for the non-local and local models with homogeneous unbiased populations) and the phase diagrams described in chapter 9 section 9.4.

A.1 NON-LOCAL MODEL

The follwoing plots give additional information about the numerical analysis for the non-local or group interdependenc model describedn in chapter 9 section 9.2.3.

A.1.1 *Dependence on temperature*

Figure A.1 illustrates how the dependence on the temperature varies when leaving the intra-couplings constant ($J_s = 1$ and $J_t = 0.6$) and varying the inter coupling k, while in figure A.2 $k = \pm 0.3$ together with $J_s = 1$ and J_t is allowed to vary in the different plots.

A.1.2 *Dependence on inter-coupling*

Figure A.3 shows different plots of the magnetization vs k as the temperature is lowered for $J_s = 1$ and $J_t = 0.6$. In figures A.4 and A.5 the temperature is kept constant (at $K_B T = 1.5$ and $K_B T = 0.4$ respectively) for different values of J_t (and $J_s = 1$ for all plots).

A.1.3 *Dependence on intra-couplings*

Figure A.6 shows different plots of the magnetization vs J_t as the temperature is lowered for $J_s = 1$ and $k = \pm 0.3$. In figures A.7 and A.8 the temperature is kept constant (at $K_B T = 1.5$ and $K_B T = 0.4$ respectively) and the graphs for different values of k (and $J_s = 1$ for all plots) plotted.

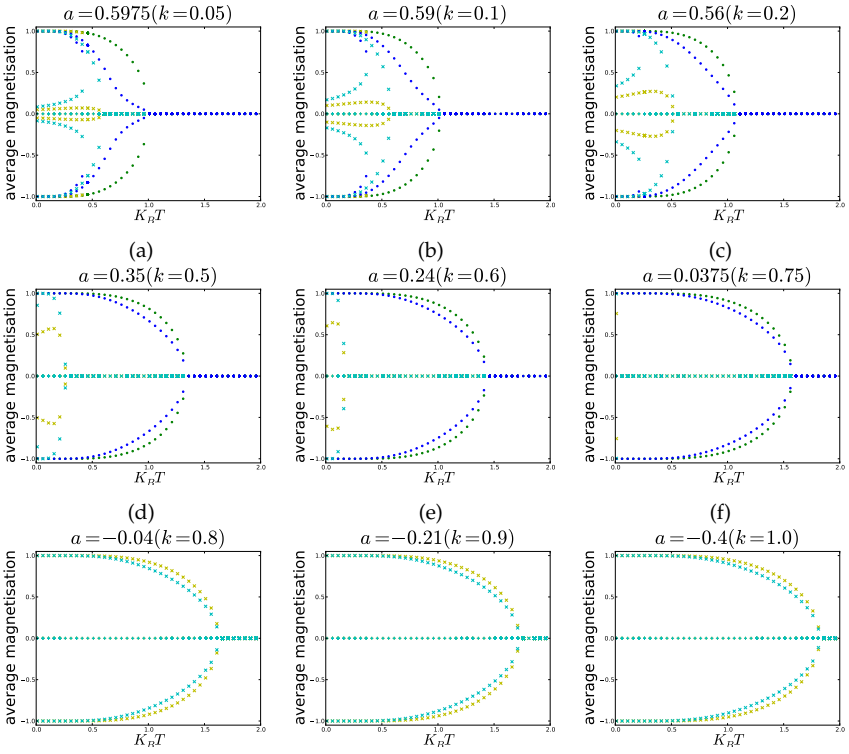


FIGURE A.1: Dependence on temperature of the numerically calculated average magnetization's (s, t) for different values of the inter-coupling k for the non-local model. $J_s = 1$ and $J_t = 0.6$ for all plots. (a) $k = 0.05$, (b) $k = 0.1$, (c) $k = 0.2$, (d) $k = 0.5$, (e) $k = 0.6$, (f) $k = 0.75$, (g) $k = 0.8$, (h) $k = 0.9$ and (i) $k = 1$. In all cases, different solutions are plotted for temperatures between 0.01 and 2 every 0.05 ($K_B T$). Magnetization are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

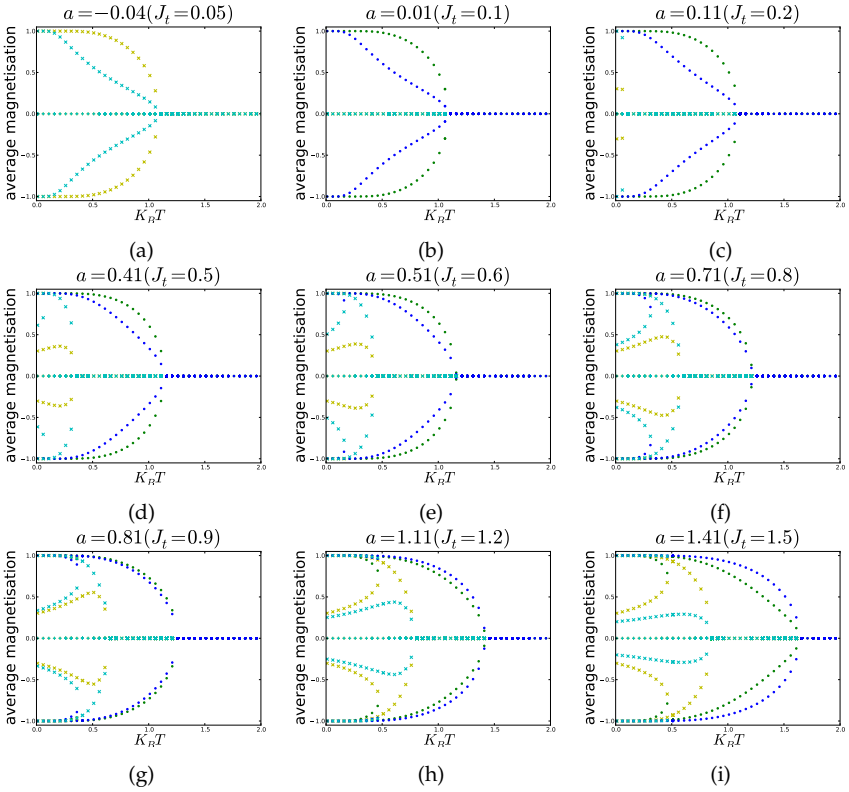


FIGURE A.2: Dependence on temperature of the numerically calculated average magnetization (s, t) for different values of the inter-coupling J_t for the non-local model. $J_s = 1$ and $k = \pm 0.3$ for all plots. (a) $J_t = 0.05$, (b) $J_t = 0.1$, (c) $J_t = 0.2$, (d) $J_t = 0.5$, (e) $J_t = 0.6$, (f) $J_t = 0.8$, (g) $J_t = 0.9$, (h) $J_t = 1.2$ and (i) $J_t = 1.5$. In all cases, different solutions are plotted for temperatures between 0.01 and 2 every 0.05 ($K_B T$). Magnetization are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

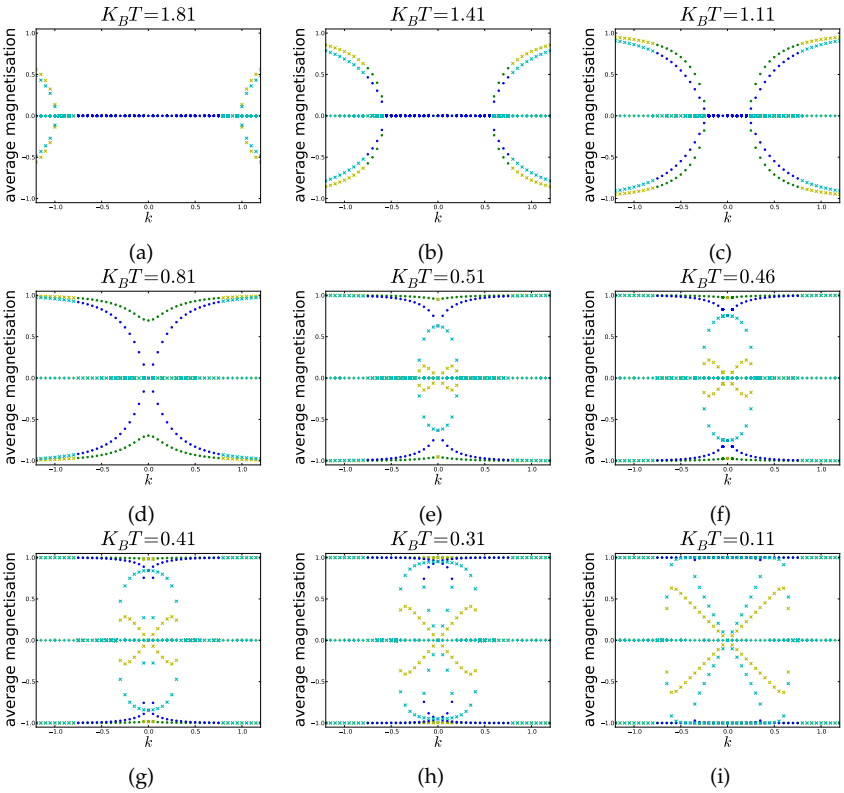


FIGURE A.3: Dependence on inter-coupling of the numerically calculated average magnetization (s, t) for different values of the temperature for the non-local model. $J_s = 1$ and $J_t = 0.6$ for all plots. (a) $K_BT = 1.81$, (b) $K_BT = 1.41$, (c) $K_BT = 1.11$, (d) $K_BT = 0.81$, (e) $K_BT = 0.51$, (f) $K_BT = 0.46$, (g) $K_BT = 0.41$, (h) $K_BT = 0.31$ and (i) $K_BT = 0.11$. In all cases, different solutions are plotted for k between -1.2 and 1.2 every 0.05 (K_BT). Magnetization are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

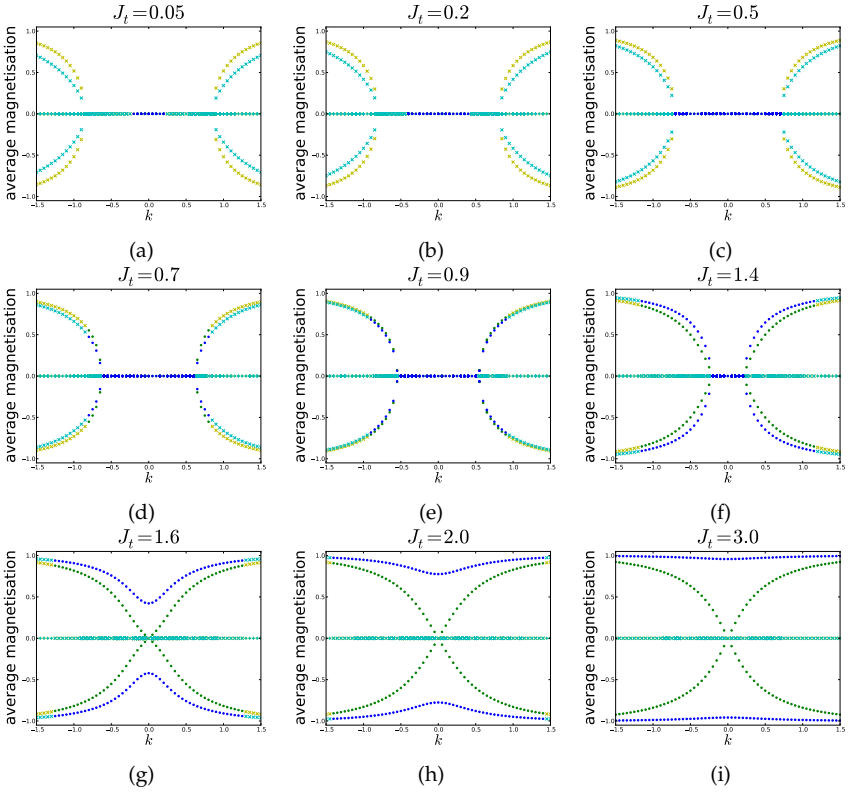


FIGURE A.4: Dependence on inter-coupling of the numerically calculated average magnetization (s, t) for different values of J_t at high temperature for the non-local model. $J_s = 1$ and $K_B T = 1.5$ for all plots. (a) $J_t = 0.05$, (b) $J_t = 0.2$, (c) $J_t = 0.5$, (d) $J_t = 0.7$, (e) $J_t = 0.9$, (f) $J_t = 1.4$, (g) $J_t = 1.6$, (h) $J_t = 2$ and (i) $J_t = 3$. In all cases, different solutions are plotted for k between -1.5 and 1.5 every 0.05 ($K_B T$). Magnetization are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

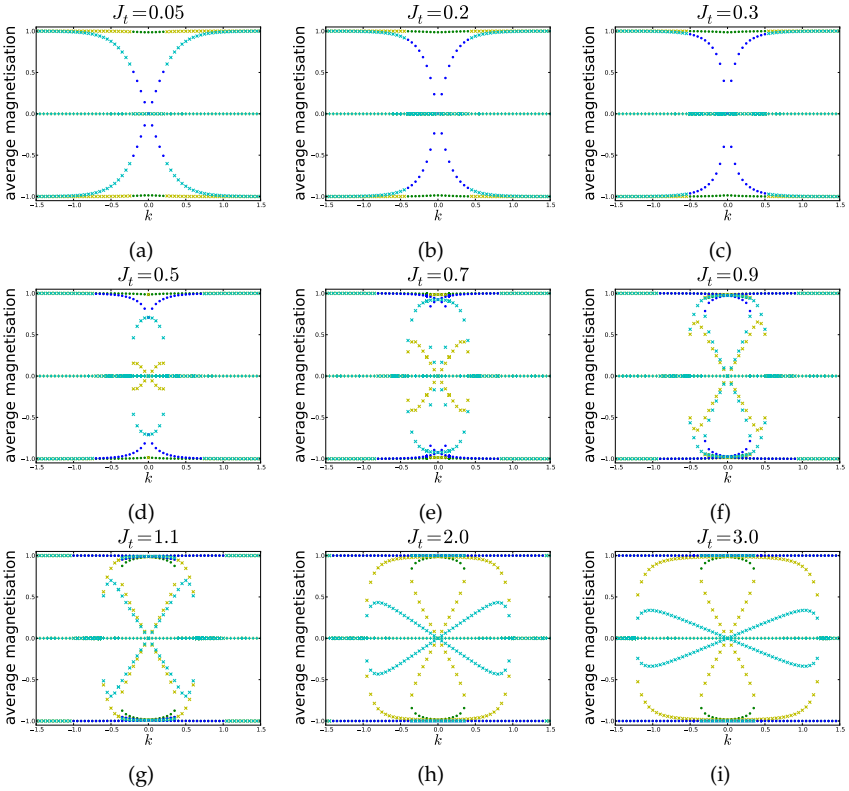


FIGURE A.5: Dependence on inter-coupling of the numerically calculated average magnetization (s, t) for different values of the J_t at low temperature for the non-local model. $J_s = 1$ and $K_BT = 0.4$ for all plots. (a) $J_t = 0.05$, (b) $J_t = 0.2$, (c) $J_t = 0.3$, (d) $J_t = 0.5$, (e) $J_t = 0.7$, (f) $J_t = 0.9$, (g) $J_t = 1.1$, (h) $J_t = 2$ and (i) $J_t = 3$. In all cases, different solutions are plotted for k between -1.5 and 1.5 every 0.05 (K_BT). Magnetization are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

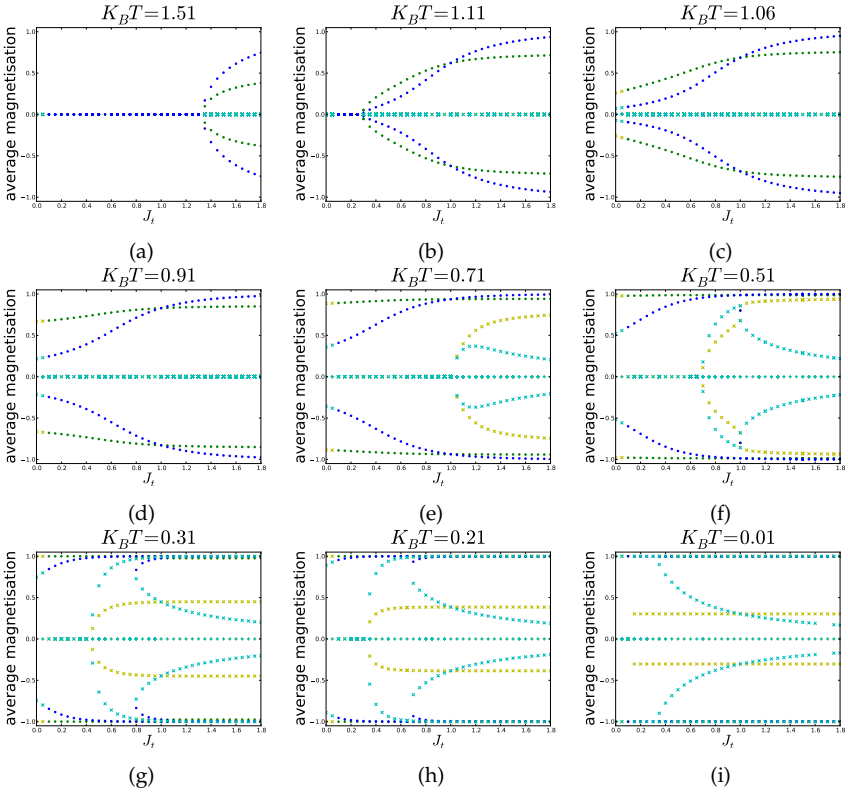


FIGURE A.6: Dependence on intra-coupling J_t of the numerically calculated average magnetization (s, t) for different values of the temperature K_BT for the non-local model. $J_s = 1$ and $k = \pm 0.3$ for all plots. (a) $K_BT = 1.51$, (b) $K_BT = 1.11$, (c) $K_BT = 1.06$, (d) $K_BT = 0.91$, (e) $K_BT = 0.71$, (f) $K_BT = 0.51$, (g) $K_BT = 0.31$, (h) $K_BT = 0.21$ and (i) $K_BT = 0.01$. In all cases, different solutions are plotted for intra-coupling J_t between 0.01 and 1.8 every 0.05. Magnetization are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

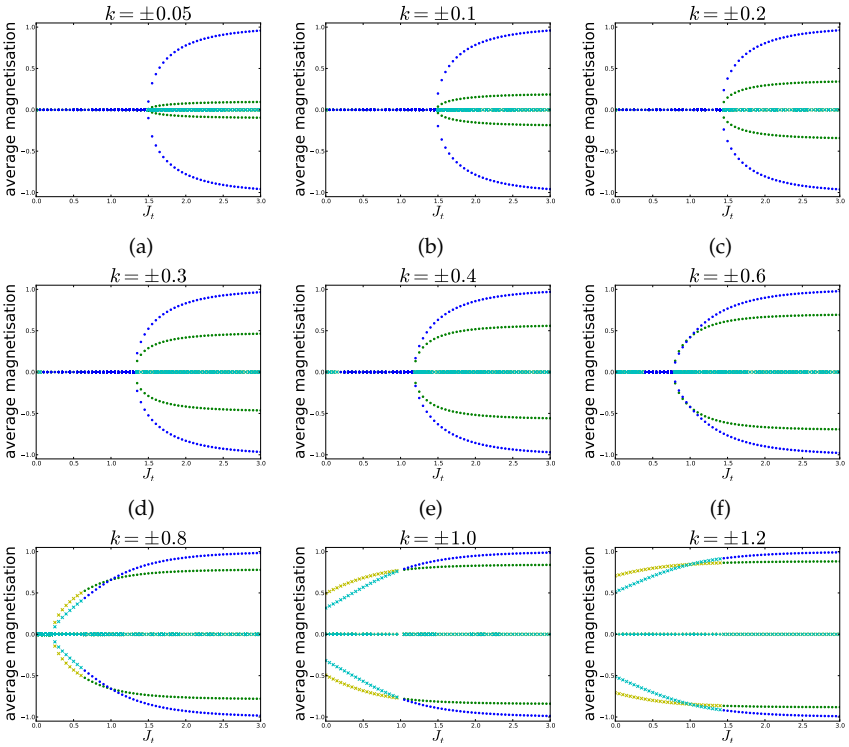


FIGURE A.7: Dependence on intra-coupling J_t of the numerically calculated average magnetization (s, t) for different values of the inter-coupling k at high temperatures for the non-local model. $J_s = 1$ and $K_B T = 1.5$ for all plots. (a) $k = 0.05$, (b) $k = 0.1$, (c) $k = 0.2$, (d) $k = 0.3$, (e) $k = 0.4$, (f) $k = 0.6$, (g) $k = 0.8$, (h) $k = 1$ and (i) $k = 1.2$. In all cases, different solutions are plotted for intra-coupling J_t between 0.01 and 3 every 0.05. Magnetization are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

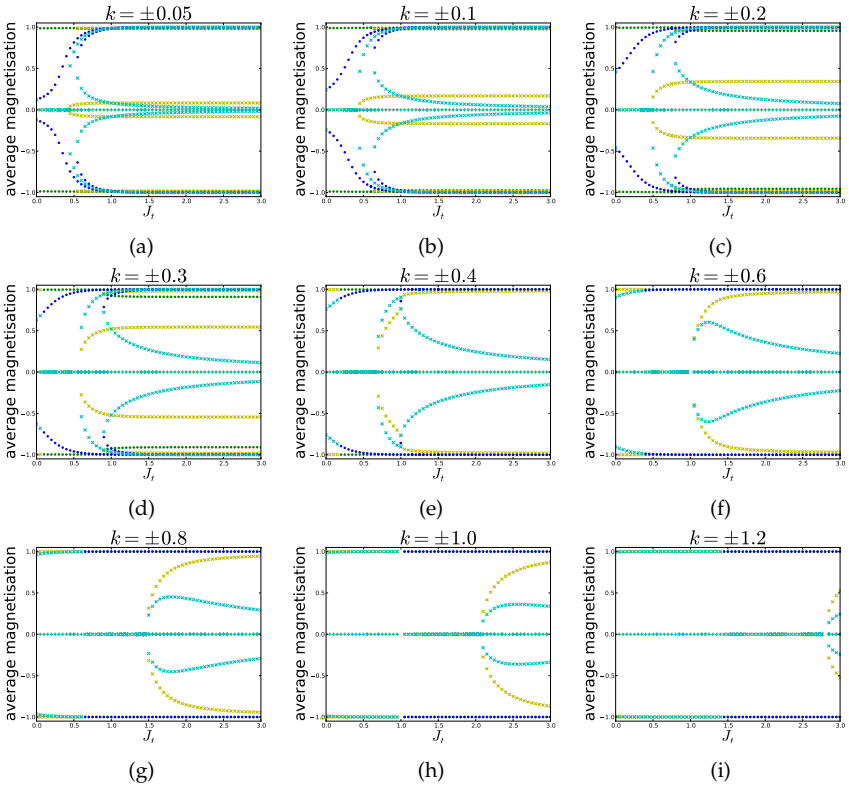


FIGURE A.8: Dependence on intra-coupling J_t of the numerically calculated average magnetization (s, t) for different values of the inter-coupling k at low temperatures for the non-local model. $J_s = 1$ and $K_B T = 0.4$ for all plots. (a) $k = 0.05$, (b) $k = 0.1$, (c) $k = 0.2$, (d) $k = 0.3$, (e) $k = 0.4$, (f) $k = 0.6$, (g) $k = 0.8$, (h) $k = 1$ and (i) $k = 1.2$. In all cases, different solutions are plotted for intra-coupling J_t between 0.01 and 3 every 0.05. Magnetization are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

A.2 LOCAL MODEL

The follwoing plots give additional information about the numerical analysis for the local or individual interdependence model described in chapter 9 section 9.3.3.

A.2.1 *Dependence on temperature*

Figure A.9 illustrates how the dependence on the temperature varies when leaving the intra-couplings constant ($J_s = 1$ and $J_t = 0.6$) and varying the inter coupling k, while in figure A.10 $k = \pm 0.3$ together with $J_s = 1$ and J_t is allowed to vary in the different plots.

A.2.2 *Dependence on inter-coupling*

Figure A.11 shows different plots of the magnetization vs k as the temperature is lowered for $J_s = 1$ and $J_t = 0.6$. In figures A.12 and A.13 the temperature is kept constant (at $K_B T = 1.5$ and $K_B T = 0.4$ respectively) for different values of J_t (and $J_s = 1$ for all plots).

A.2.3 *Dependence on intra-couplings*

Figure A.14 shows different plots of the magnetization vs J_t as the temperature is lowered for $J_s = 1$ and $k = \pm 0.3$. In figures A.15 and A.16 the temperature is kept constant (at $K_B T = 1.5$ and $K_B T = 0.4$ respectively) and the graphs for different values of k (and $J_s = 1$ for all plots) plotted.

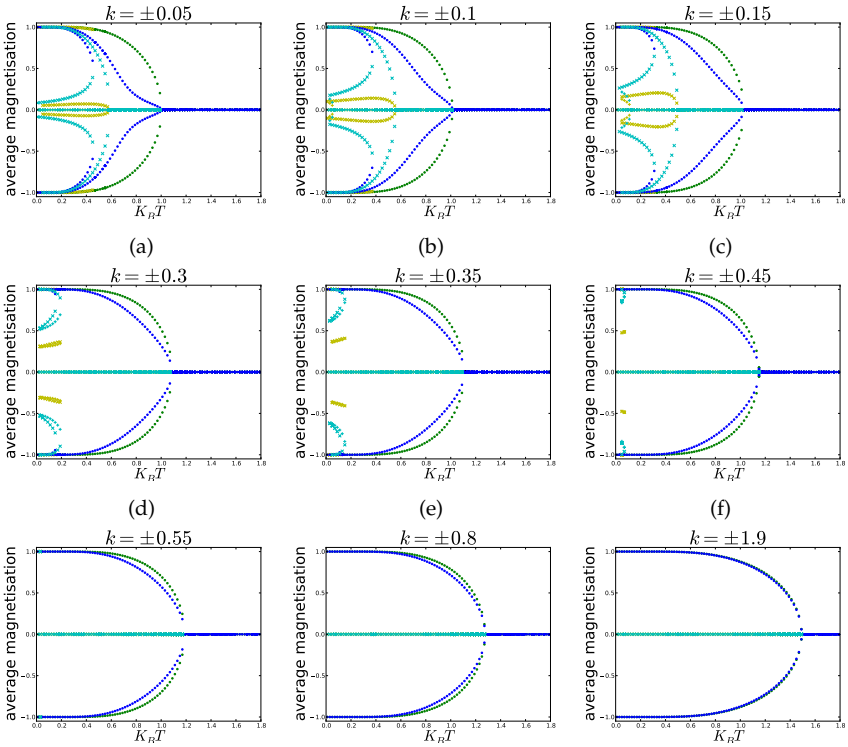


FIGURE A.9: Dependence on temperature of the numerically calculated average magnetisations (s, t) for different values of the inter-coupling k for the local model. $J_s = 1$ and $J_t = 0.6$ for all plots. (a) $k = \pm 0.05$, (b) $k = \pm 0.1$, (c) $k = \pm 0.15$, (d) $k = \pm 0.3$, (e) $k = \pm 0.35$, (f) $k = \pm 0.45$, (g) $k = \pm 0.55$, (h) $k = \pm 0.8$ and (i) $k = \pm 1.9$. In all cases, different solutions are plotted for temperatures between 0.01 and 1.8 every 0.02 ($K_B T$). Magnetisations are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

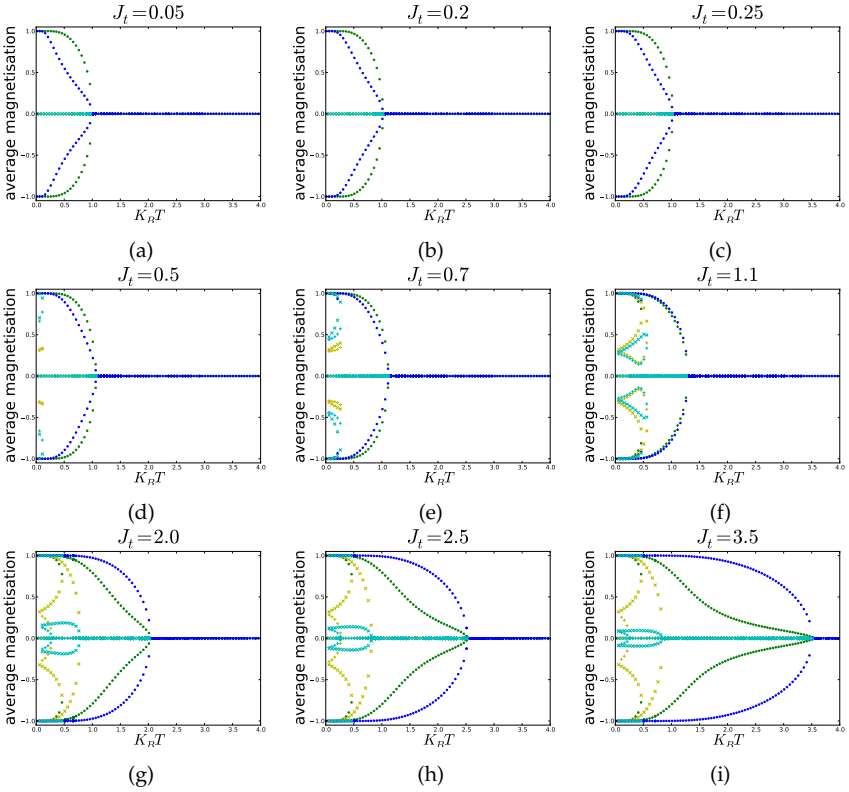


FIGURE A.10: Dependence on temperature of the numerically calculated average magnetisations (s, t) for different values of the intra-coupling J_t for the local model. $J_s = 1$ and $k = 0.15$ for all plots. (a) $J_t = 0.05$, (b) $J_t = 0.2$, (c) $J_t = 0.25$, (d) $J_t = 0.5$, (e) $J_t = 0.7$, (f) $J_t = 1.1$, (g) $J_t = 2$, (h) $J_t = 2.5$ and (i) $J_t = 3.5$. In all cases, different solutions are plotted for temperatures between 0.01 and 4 every 0.05 (K_BT). Magnetisations are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

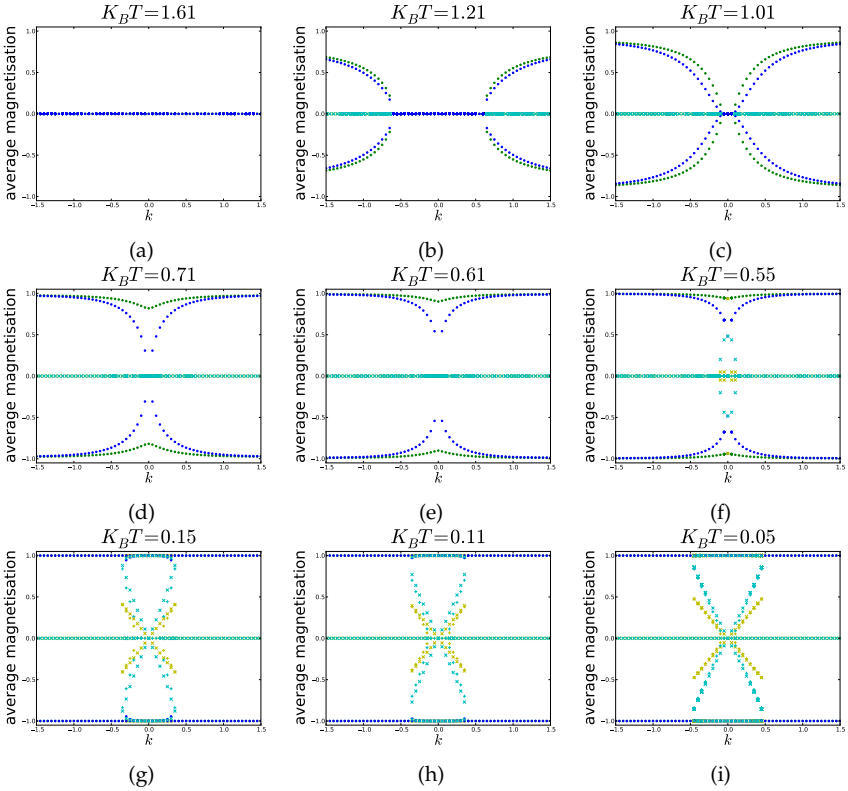


FIGURE A.11: Dependence on inter-coupling k of the numerically calculated average magnetisations (s, t) for different values of the temperature K_BT for the local model. $J_s = 1$ and $k = 0.15$ for all plots. (a) $K_BT = 1.61$, (b) $K_BT = 1.21$, (c) $K_BT = 1.01$, (d) $K_BT = 0.71$, (e) $K_BT = 0.61$, (f) $K_BT = 0.55$, (g) $K_BT = 0.15$, (h) $K_BT = 0.11$ and (i) $K_BT = 0.05$. In all cases, different solutions are plotted for inter-coupling between -1.5 and 1.5 every 0.05 (K_BT). Magnetisations are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

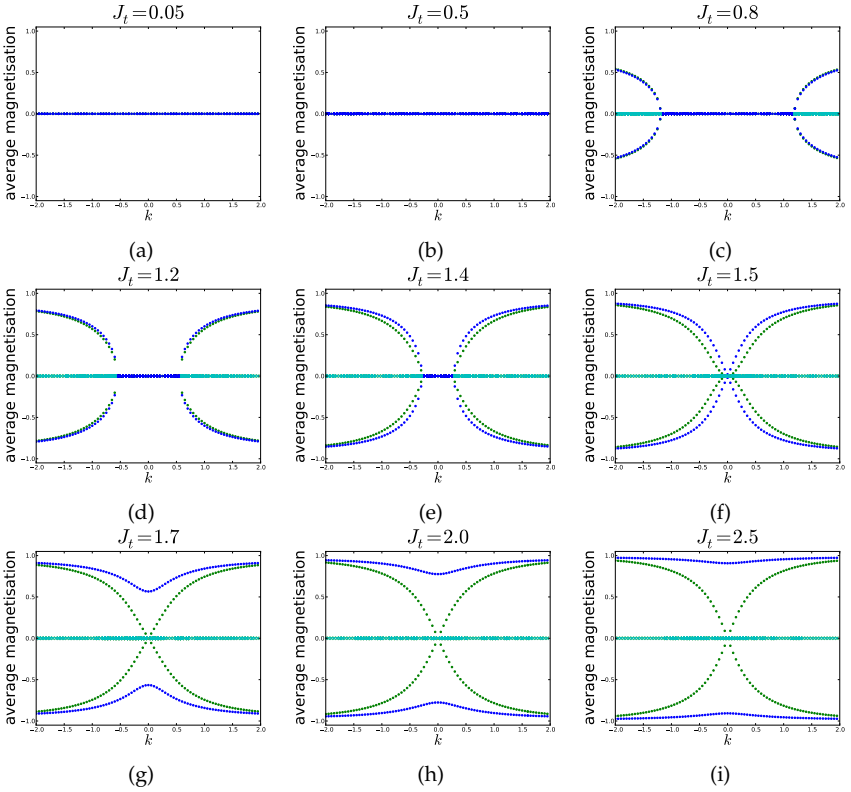


FIGURE A.12: Dependence on inter-coupling k of the numerically calculated average magnetisations (s, t) for different values of the intra-coupling J_t at high temperatures for the local model. $J_s = 1$ and $K_B T = 1.5$ for all plots. (a) $J_t = 0.05$, (b) $J_t = 0.5$, (c) $J_t = 0.8$, (d) $J_t = 1.2$, (e) $J_t = 1.4$, (f) $J_t = 1.5$, (g) $J_t = 1.7$, (h) $J_t = 2$ and (i) $J_t = 2.5$. In all cases, different solutions are plotted for inter-coupling k between -2 and 2 every 0.05. Magnetisations are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

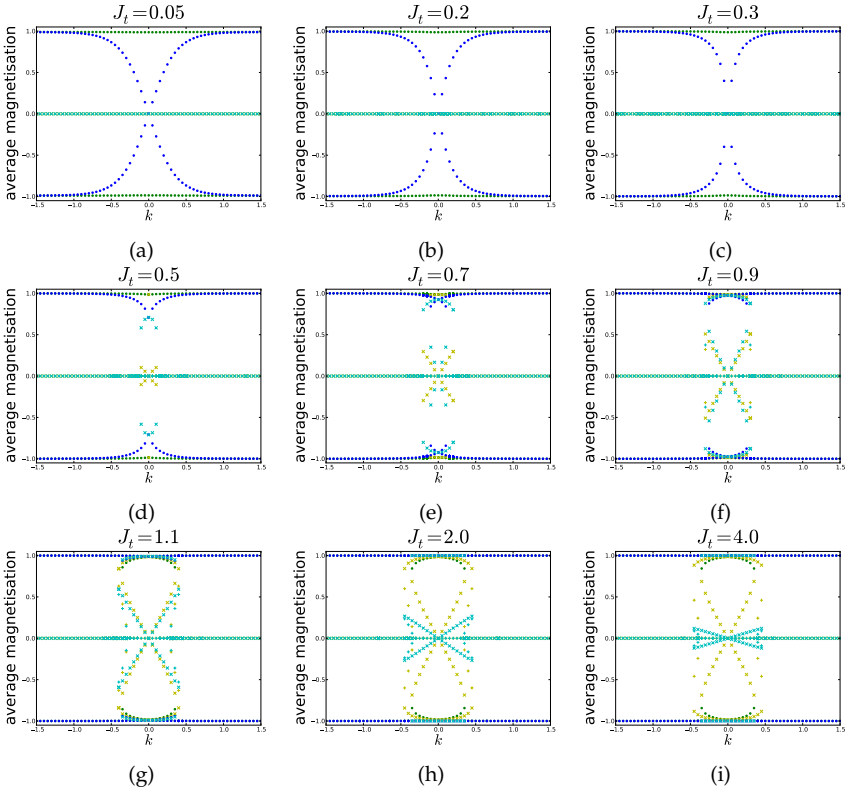


FIGURE A.13: Dependence on inter-coupling k of the numerically calculated average magnetisations (s, t) for different values of the intra-coupling J_t at low temperatures for the local model. $J_s = 1$ and $K_B T = 0.4$ for all plots. (a) $J_t = 0.05$, (b) $J_t = 0.2$, (c) $J_t = 0.3$, (d) $J_t = 0.5$, (e) $J_t = 0.7$, (f) $J_t = 0.9$, (g) $J_t = 1.1$, (h) $J_t = 2$ and (i) $J_t = 4$. In all cases, different solutions are plotted for inter-coupling k between -1.5 and 1.5 every 0.05. Magnetisations are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

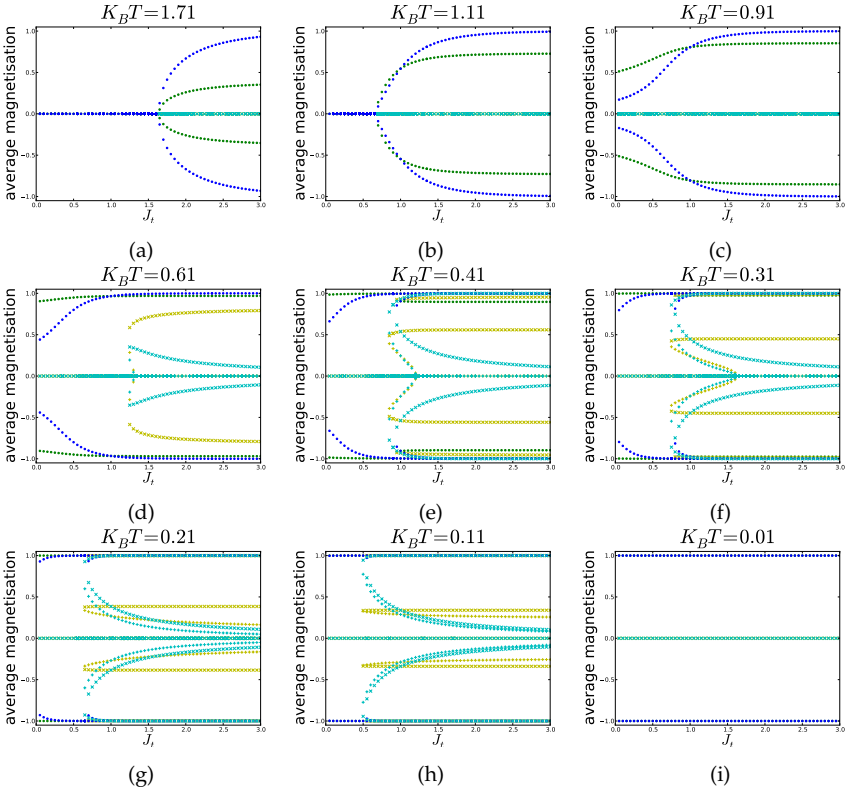


FIGURE A.14: Dependence on intra-coupling J_t of the numerically calculated average magnetisations (s, t) for different values of the temperature K_BT for the local model. $J_s = 1$ and $k = 0.15$ for all plots. (a) $K_BT = 1.71$, (b) $K_BT = 1.11$, (c) $K_BT = 0.91$, (d) $K_BT = 0.61$, (e) $K_BT = 0.41$, (f) $K_BT = 0.31$, (g) $K_BT = 0.21$, (h) $K_BT = 0.11$ and (i) $K_BT = 0.01$. In all cases, different solutions are plotted for intra-coupling J_t between 0.01 and 3 every 0.05. Magnetisations are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

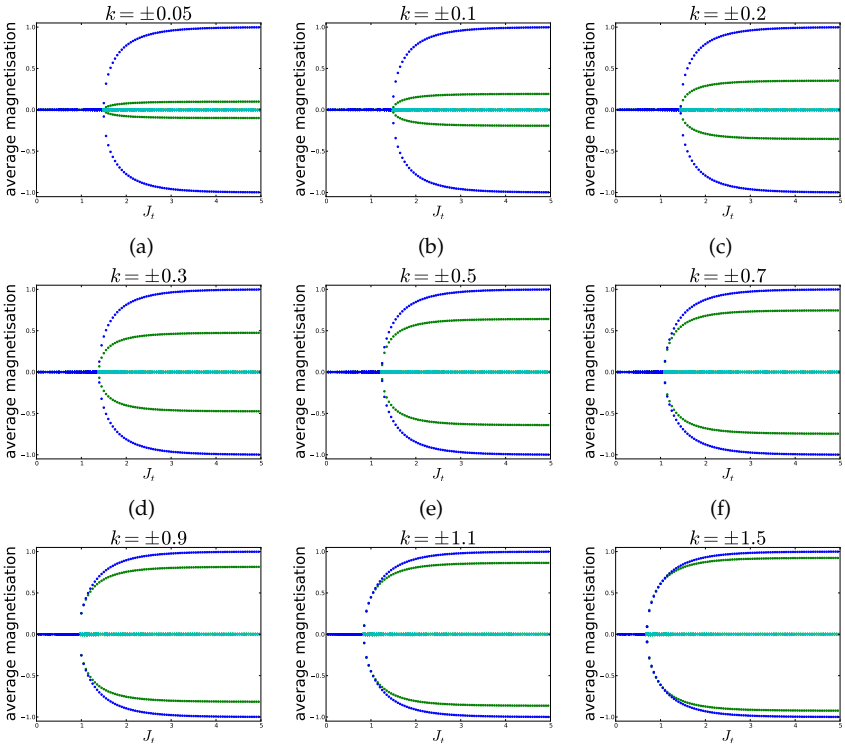


FIGURE A.15: Dependence on intra-coupling J_t of the numerically calculated average magnetisations (s, t) for different values of the inter-coupling k at high temperatures for the local model. $J_s = 1$ and $K_B T = 1.5$ for all plots. (a) $k = 0.05$, (b) $k = 0.1$, (c) $k = 0.2$, (d) $k = 0.3$, (e) $k = 0.5$, (f) $k = 0.7$, (g) $k = 0.9$, (h) $k = 1.1$ and (i) $k = 1.5$. In all cases, different solutions are plotted for intra-coupling J_t between 0.01 and 5 every 0.05. Magnetisations are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

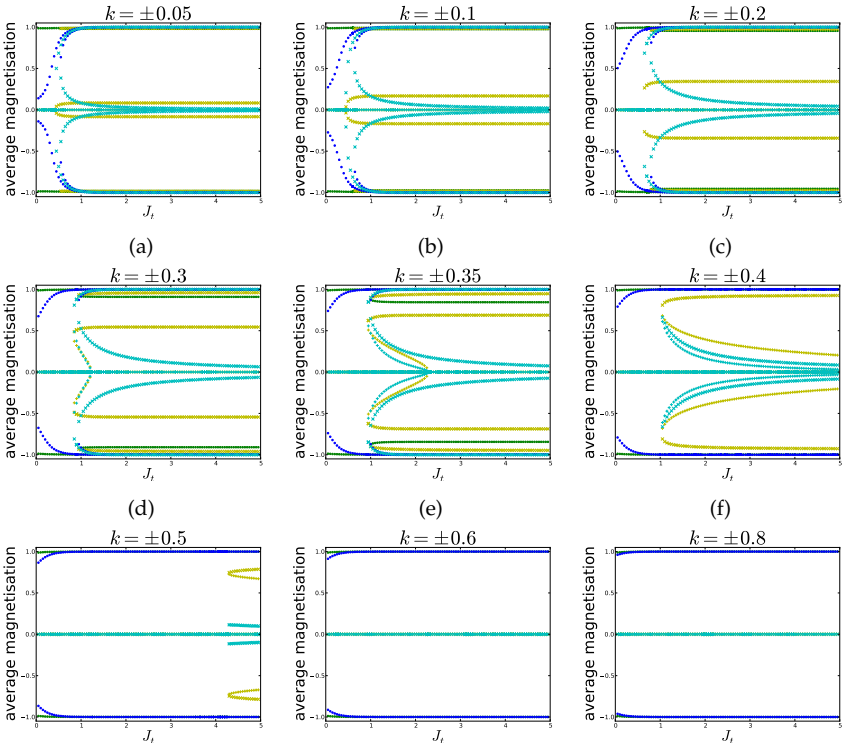


FIGURE A.16: Dependence on intra-coupling J_t of the numerically calculated average magnetisations (s, t) for different values of the inter-coupling k at low temperatures for the local model. $J_s = 1$ and $K_B T = 0.4$ for all plots. (a) $k = 0.05$, (b) $k = 0.1$, (c) $k = 0.2$, (d) $k = 0.3$, (e) $k = 0.35$, (f) $k = 0.4$, (g) $k = 0.5$, (h) $k = 0.6$ and (i) $k = 0.8$. In all cases, different solutions are plotted for intra-coupling J_t between 0.01 and 5 every 0.05. Magnetisations are plotted in green for s and blue for t. Dark points are used for stable solutions and lighter asp ($\times$, for saddle points) or cross ($+$, for maxima) for non stable solutions. Own elaboration using numerically computed solutions to the equations of state. The image had already appeared in [71].

BIBLIOGRAPHY

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. & Zheng, X. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems* http://tensorflow.org/. 2015.
2. Adebiyi, A., Adewumi, A. & Ayo, C. *Stock price prediction using the ARIMA model* in *Proceedings - UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, UKSim 2014* (2014).
3. Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716 (1974).
4. Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L. & Savarese, S. *Social LSTM: Human Trajectory Prediction in Crowded Spaces* in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 961.
5. Alizadeh, M., Weber, I., Cioffi-Revilla, C., Fortunato, S. & Macy, M. *Psychological and Personality Profiles of Political Extremists* 2017.
6. Andrews, R. L. Forecasting Performance of Structural Time Series Models. *Journal of Business and Economic Statistics* **12**, 129 (1994).
7. Axelrod, R. The dissemination of culture - A model with local convergence and global polarization. *Journal of Conflict Resolution* **41**, 203 (1997).
8. Axelrod, R. & Bennett, D. S. A Landscape Theory of Aggregation. *British Journal of Political Science* **23**, 211 (1993).
9. Axelrod, R. & Hamilton, W. D. The Evolution of Cooperation. *Science* **211**, 1390 (1981).
10. Axelrod, R., Mitchell, W., Thomas, R. E., Bennett, D. S. & Bruderer, E. Coalition Formation in Standard-Setting Alliances. *Management Science* **41**, 1493 (1995).

11. Axtell, R. *The Emergence of Firms in a Population of Agents: Local Increasing Returns, Unstable Nash Equilibria, and Power Law Size Distributions* Working paper No.3. Washington, Brookings Institution. 1999.
12. Ball, P. The physical modelling of society: a historical perspective. *Physica A* **314**, 1 (2002).
13. Ball, P. *Critical mass. How one things lead to another* (arrow books, 2004).
14. Ball, P. The Physical Modelling of Human Social Systems. *ComPlexUs* **1**, 190 (2004).
15. Bansal, S., Kamper, H., Livescu, K., Lopez, A. & Goldwater, S. *Low-Resource Speech-to-Text Translation* arXiv:1803.09164. 2018.
16. Batislam, E. P., Denizel, M. & Filiztekin, A. Empirical validation and comparison of models for customer base analysis. *International Journal of Research in Marketing* **24**, 201 (2007).
17. Bauckhage, C., Drachen, A. & Sifa, R. Clustering game behavior data. *IEEE Transactions on Computational Intelligence and AI in Games* **7**, 266 (2014).
18. Baumann, F., Lorenz-Spreen, P., Sokolov, I. M. & Starnini, M. Modeling Echo Chambers and Polarization Dynamics in Social Networks. *Phys. Rev. Lett.* **124**, 048301 (4 2020).
19. Baxter, R. J. *Exactly Solved Models in Statistical Mechanics* (Academic Press, 1982).
20. Beane, T. & Ennis, D. Market Segmentation: A Review. *European Journal of Marketing* **21**, 20 (1987).
21. Becker, G. S. *A Treatise on the Family* (Harvard University Press, 1981).
22. Becker, G. S. A note on restaurant pricing and other examples of social influences on price. *Journal of Political Economy* **99**, 1109 (1991).
23. Belgiu, M. & Dragut, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* **114**, 24 (2016).
24. Bengio, Y. Learning deep architectures for AI. *Foundations and Trends® in Machine Learning* **2**, 1 (2009).
25. Berger, P. D. & Nasr, N. I. Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing* **12**, 17 (1998).
26. Bernardes, A. T., Stauffer, D. & Kertesz, J. Election results and the Sznajd model on Barabasi network. *European Physical Journal B* **25**, 123 (2002).

27. Bertens, P., Guitart, A., Chen, P. P. & Periáñez, Á. *A Machine-Learning Item Recommendation System for Videogames* in *IEEE Conference on Computational Intelligence and Games (CIG 2018)* Aug. 14–17, 2018 (IEEE, Maastricht, The Netherlands, 2018), 1.

28. Bertens, P., Guitart, A. & Periáñez, Á. *Games and Big Data: A Scalable Multi-Dimensional Churn Prediction Model* in *IEEE Conference on Computational Intelligence and Games (CIG 2017)* Aug. 22–25, 2017 (IEEE, New Yor, USA, 2017), 33.

29. Binder, K. & Young, A. P. Spin glasses: Experimental facts, theoretical concepts, and open questions. *Reviews of Modern Physics* **58**, 801 (1986).

30. Bjørnstad, O. N. & Grenfell, B. T. Noisy Clockwork: Time Series Analysis of Population Fluctuations in Animals. *Science* **293**, 638 (2001).

31. Blume, L. The Statistical Mechanics of Strategic Interaction. *Games and Economic Behavior* **5**, 387 (1993).

32. Borghesi, C. & Bouchaud, J.-P. Of songs and men: a model for multiple choice with herding. *Quality & Quantity* **41**. 10.1007/s11135-007-9074-6, 557 (4 2007).

33. Box, G. & Jenkins, G. M. *Time Series Analysis: Forecasting and Control* (Holden-Day, 1976).

34. Bradley, A. P. The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recogn.* **30**, 1145 (1997).

35. Breiman, L. Random forests. *Machine learning* **45**, 5 (2001).

36. Breiman, L., Friedman, J., Olshen, R. & Stone, C. Classification and regression trees. Wadsworth Int. *Group* **37**, 237 (1984).

37. Brier, G. W. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **78**, 1 (1950).

38. Brock, W. A. & Durlauf, S. N. Discrete choice with social interactions. *Review of Economic Studies* **68**, 235 (2001).

39. Brockwell, P. & A Davis, R. *An Introduction to Time Series and Forecasting* (Springer, 2002).

40. Brush, S. G. History of the Lenz-Ising Model. *Reviews of Modern Physics* **39**, 883 (1967).

41. Burda, Z., Jurkiewicz, J. & Nowak, M. A. Is econophysics a solid science? *Acta Physica Polonica B* **34**, 87 (2003).

42. Callen, E. & Shapero, D. A theory of social imitation. *Physics Today* **27**, 23 (1974).

43. Campbell, M. & Ormerod, P. *Social interaction and the dynamics of crime* Preprint. 2000.

44. Campello, R., Moulavi, D. & Sander, J. *Density-Based Clustering Based on Hierarchical Density Estimates* in *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining* **7819** (2013), 160.

45. Carbone, G. & Giannoccaro, I. Model of human collective decision-making in complex environments. *The European Physical Journal B* **88** (2015).

46. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Reviews of Modern Physics* **81**, 591 (2009).

47. Castellano, C., Marsili, M. & Vespignani, A. Nonequilibrium Phase Transition in a Model for Social Influence. *Physical Review Letters* **85**, 3536 (2000).

48. Chamberlain, B. P., Cardoso, A., Liu, C. H., Pagliari, R. & Deisenroth, M. P. *Customer Life Time Value Prediction Using Embeddings* in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)* (2017).

49. Chatfield, C. & Goodhardt, G. J. A Consumer Purchasing Model with Erlang Inter-Purchase Time. *Journal of the American Statistical Association* **68**, 828 (1973).

50. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* **16**, 321 (2002).

51. Chen, P. P., Guitart, A., del Río, A. F. & Periáñez, Á. *Customer Lifetime Value in Video Games Using Deep Learning anf Parametric Models* in *IEEE International Conference on Big Data (Big Data 2018)* Dec. 10–13, 2018 (IEEE, Seattle, USA, 2018), 2134.

52. Chen, P. P., Guitart, A. & Periáñez, Á. The winning solution to the IEEE CIG 2017 Game Data Mining competition. *Machine Learning and Knowledge Extraction* **1**, 252 (2019).

53. Chen, P., Yuan, H. & Shu, X. *Forecasting crime using the ARIMA model* in *Proceedings - 5th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2008* **5** (2008), 627.

54. Chen, R., Hua, Q., Chang, Y., Wang, B., Zhang, L. & Kong, X. A Survey of Collaborative Filtering-Based Recommender Systems: From Traditional Methods to Hybrid Methods Based on Social Networks. *IEEE Access* **6**, 64301 (2018).

55. Chen, T. & Guestrin, C. *XGBoost: A Scalable Tree Boosting System* in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, San Francisco, California, USA, 2016), 785.

56. Cheung, Y.-W. & Lai, K. S. Lag order and critical values of the augmented Dickey–Fuller test. *Journal of Business & Economic Statistics* **13**, 277 (1995).

57. Ciodaro, T., Deva, D., de Seixas, J. M. & Damazio, D. Online particle detection with Neural Networks based on topological calorimetry information. *Journal of Physics: Conference Series* **368**, 012030 (2012).

58. Clark, T., Bradburn, M., Love, S. & Altman, D. Survival Analysis Part I: Basic concepts and first analyses. *British Journal of Cancer* **89(2)**, 232 (2003).

59. *Fundamental Problems in Statistical Mechanics* (ed Cohen, E.) (Elsevier, 1985).

60. Collet, F., Pra, P. D. & Sartori, E. A Simple Mean Field Model for Social Interactions: Dynamics, Fluctuations, Criticality. *Journal of Statistical Physics* **139**, 820 (2010).

61. Contreras, J., Espinola, R., Nogales, F. J. & Conejo, A. J. ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems* **18**, 1014 (2003).

62. Cowley, B. *Player Profiling and Modelling in Computer and Video Games* PhD thesis (2009), 299.

63. Cox, D. R. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)* **34**, 187 (1972).

64. Cox, D. R. & Oakes, D. *Analysis of survival data* (CRC Press, 1984).

65. Cricenti, A. L., Branch, P. A. & Armitage, G. J. *Time-series modelling of server to client IP packet length in first person shooter games* in *2007 15th IEEE International Conference on Networks* (2007), 507.

66. Cui, Y., Tobossi, R. & Vigouroux, O. Modelling customer online behaviours with neural networks: applications to conversion prediction and advertising retargeting. *arXiv preprint arXiv:1804.07669* (2018).

67. Dall'Asta, L., Castellano, C. & Marsili, M. Statistical physics of the Schelling model of segregation. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, L07002 (2008).

68. Davidovici-Nora, M. Innovation in business models in the video game industry: Free-To-Play or the gaming experience as a service. *The Computer Games Journal* **2**, 22 (2013).

69. Davidson-Pilon, C., Kalderstam, J., Jacobson, N., Reed, S., Kuhn, B., Zivich, P., Williamson, M., AbdealiJK, Datta, D., Fiore-Gartland, A., Parij, A., WIlson, D., Gabriel, Moneda, L., Moncada-Torres, A., Stark, K., Gadgil, H., Jona, Singaravelan, K., Besson, L., Peña, M. S., Anton, S., Klintberg, A., GrowthJeff, Noorbakhsh, J., Begun, M., Kumar, R., Hussey, S., Seabold, S. & Golland, D. *CamDavidsonPilon/lifelines: v0.25.11* version v0.25.11-2. 2021.

70. Deffuant, G., Neau, D., Amblard, F. & Weisbuch, G. Mixing beliefs among interacting agents. *Advances in Complex Systems* **3**, 87 (2000).

71. Del Río, A. F. *Coupled Ising models and interdependent discrete choices under social influence in homogeneous populations* MA thesis (Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain, 2011).

72. Del Río, A. F., Chen, P. P. & Periáñez, Á. *Profiling Players with Engagement Predictions* in *2019 IEE Conference on Games (CoG)* Aug. 20–23, 2019 (IEE, London, England, 2019).

73. Del Río, A. F., Guitart, A. & Periáñez, Á. A Time Series Approach To Player Churn and Conversion in Videogames. *Intelligent Data Analysis* **25**, 177 (2021).

74. Del Río, A. F., Korutcheva, E. & de la Rubia, J. Interdependent binary choices under social influence: phase diagram for homogeneous unbiased populations. *Complexity* **17**, 31 (2012).

75. Diaz-Uriarte, R. & de Andres, S. A. *Variable selection from random forests: application to gene expression data* arXiv:q-bio/0503025. 2005.

76. Drachen, A., Pastor, M., Liu, A., Fontaine, D. J., Chang, Y., Runge, J., Sifa, R. & Klabjan, D. *To Be or Not to Be... Social: Incorporating Simple Social Features in Mobile Game Customer Lifetime Value Predictions* in *Proceedings of the Australasian Computer Science Week Multiconference (ACSW)* Article no. 40 (2018).

77. Drachen, A., Sifa, R., Bauckhage, C. & Thurau, C. *Guns, swords and data: Clustering of player behavior in computer games in the wild* in *2012 IEEE conference on Computational Intelligence and Games (CIG)* (2012), 163.

78. Drachen, A., Thurau, C., Sifa, R. & Bauckhage, C. A comparison of methods for player clustering via behavioral telemetry. *arXiv preprint arXiv:1407.3950* (2014).

79. Durlauf, S. N. in *The Economy as an Evolving Complex System II* (eds B. Arthur, S. N. D. & Lane, D.) (Addison-Wesley Pub. Co., 1997).

80. Durlauf, S. N. How can statistical mechanics contribute to social science? *Proceedings of the National Academy of Science* **96**, 10582 (1999).

81. Dwyer, F. R. Customer lifetime valuation to support marketing decision making. *Journal of Direct Marketing* **11**, 6 (1997).

82. Ehrenberg, A. S. C. The Pattern of Consumer Purchases. *Journal of the Royal Statistical Society Series C* **8**, 26 (1959).

83. Ehrenberg, A. *Repeat-buying: Facts, Theory and Applications* (Griffin, 1988).

84. Elghafghug, a., Vanderstichel, R., St-Hilaire, S. & Stryhn, H. Using state-space models to predict the abundance of juvenile and adult sea lice on Atlantic salmon. *Epidemics* **24**, 76 (2018).

85. Epstein, J. M. Learning to Be Thoughtless: Social Norms and Individual Computation. *Computational Economics* **18**, 9 (1 2001).

86. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise* in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (AAAI Press, Portland, Oregon, 1996), 226.

87. Esteva Andre andKuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M. & Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115 (2017).

88. Fader, P. S. & Hardie, B. G. *A note on deriving the Pareto/NBD model and related expressions* `http://www.brucehardie.com/notes/009/pareto_nbd_derivations_2005-11-05.pdf`. 2005.

89. Fader, P. S., Hardie, B. G. & Lee, K. L. RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research* **42**, 415 (2005).

90. Fader, P. S., Hardie, B. G. S. & Lee, K. L. "Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model. *Marketing Science* **24**, 275 (2005).

91. Farris, P. W., Bendle, N., Pfeifer, P. & Reibstein, D. *Marketing metrics: The definitive guide to measuring marketing performance* (Pearson Education, 2010).

92. Faruk, D. Ö. An hybrid neural network and ARIMA model for water quality time series prediction. *Engineering Applications of Artificial Intelligence* **23**, 586 (2010).

93. Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters* **27**, 861 (2006).

94. Feigenbaum, J. Financial physics. *Reports on Progress in Physics* **66**, 1611 (2003).

95. Fernandez, M. A., Korutcheva, E. & de la Rubia, F. J. A 3-states magnetic model of binary decisions in sociophysics. *Physica A: Statistical Mechanics and its Applications* **462**, 603 (2016).

96. Fields, T. *Mobile and Social Game Design: Monetization Methods and Mechanics* 2nd ed., 2 (CRC Press, 2014).

97. Föllmer, H. Random economies with many interacting agents. *Journal of Mathematical Economics* **1**, 51 (1974).

98. Fortunato, S. & Castellano, C. Scaling and Universality in Proportional Elections. *Physical Review Letters* **99**, 138701 (2007).

99. Fotso, S. *et al. PySurvival: Open source package for Survival Analysis modeling* https://www.pysurvival.io/. 2019.

100. Frederickson, B. *Implicit: Fast Python Collaborative Filtering for Implicit Datasets* https://implicit.readthedocs.io.

101. Friedman, J. Stochastic Gradient Boosting. *Computational Statistics and Data Analysis* **38**, 367 (2002).

102. Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics* **29**, 1189 (2000).

103. Galam, S. Reorientations, freezing, and plastic phase. *Journal of Applied Physics* **63**, 3760 (1988).

104. Galam, S. Rational group decision making: A random field Ising model at T = 0. *Physica A: Statistical and Theoretical Physics* **238**, 66 (1997).

105. Galam, S. Modelling rumors: the no plane Pentagon French hoax case. *Physica A - Statistical Mechanics and Applications* **320**, 571 (2003).

106. Galam, S. & Gabay, M. Coupled Spin Systems and Plastic Crystals. *EPL (Europhysics Letters)* **8**, 167 (1989).

107. Galam, S., Gefen, Y. & Shapir, Y. Sociophysics: A new approach of sociological collective behaviour. I. mean-behaviour description of a strike. *The Journal of Mathematical Sociology* **9**, 1 (1982).

108. Galam, S., Henriques, V. B. & Salinas, S. R. Compressible spin models for plastic crystals. *Physical Review B* **42**, 6720 (1990).

109. Galam, S. & Moscovici, S. Towards a theory of collective phenomena: Consensus and attitude changes in groups. *European Journal of Social Psychology* **21**, 49 (1991).

110. Galam, S., Salinas, S. R. & Shapir, Y. Randomly coupled Ising models. *Physical Review B* **51**, 2864 (1995).

111. Galam, S. Sociophysics: a review of Galam models. *International Journal of Modern Physics C* **19**, 409 (2008).

112. Galam, S. *Sociophysics: A Physicist's Modeling of Psycho-political Phenomena* (Springer, 2012).

113. Galam, S. The Trump phenomenon: An explanation from sociophysics. *International Journal of Modern Physics B* **31** (2017).

114. Galam, S. & Cheon, T. Asymmetric Contrarians in Opinion Dynamics. *Entropy* **22**, 25 (2019).

115. Galam, S. & Javarone, M. A. Modeling Radicalization Phenomena in Heterogeneous Populations. *PLOS ONE* **11** (ed Braunstein, L. A.) e0155407 (2016).

116. Gauvin, L., Vannimenus, J. & Nadal, J.-P. Phase diagram of a Schelling segregation model. *European Physical Journal B* **70**, 293 (2009).

117. Geurts, P., Ernst, D. & Wehenkel, L. Random forest in remote sensing: A review of applications and future directions. *Machine Learning* **63**, 3 (2006).

118. Gintis, H. *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Interaction* (Princeton University Press, 2000).

119. Glaeser, E. L., Sacerdote, B. & Scheinkman, J. A. Crime and social interactions. *Quaterly Journal of Economics* **111**, 507 (1996).

120. Glorot, X. & Bengio, Y. *Understanding the difficulty of training deep feedforward neural networks* in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (eds Teh, Y. W. & Titterington, M.) **9** (PMLR, Chia Laguna Resort, Sardinia, Italy, 2010), 249.

121. Goldenfeld, N. *Lectures on Phase Transitions and the Renormalization Group* (Westview Press, 1992).

122. Golo, N. & Galam, S. Conspiratorial Beliefs Observed through Entropy Principles. *Entropy* **17**, 5611 (2015).

123. González-Avella, J., Eguíluz, V. M., Marsili, M., Vega-Redondo, F. & San Miguel, M. Threshold Learning Dynamics in Social Networks. *PLoS ONE* **6**, e20207 (2011).

124. Gordon, M. B., Nadal, J.-P., Phan, D. & Semeshenko, V. Discrete Choices under Social Influence: Generic Properties. *Mathematical Models and Methods in Applied Sciences* **19(1)**, 1441 (2008).

125. Gordon, M. B., Nadal, J.-P., Phan, D. & Vannimenus, J. Seller's dilemma due to social interactions between customers. *Physica A - Statistical Mechanics and Applications* **356**, 628 (2005).

126. Gower, J. C. A general coefficient of similarity and some of its properties. *Biometrics* **27**, 857 (1971).

127. Goyat, S. The basis of market segmentation: a critical review of literature. *European Journal of Business and Management* **3**, 45 (2011).

128. Graf, E., Schmoor, C., Sauerbrei, W. & Schumacher, M. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine* **18**, 2529 (1999).

129. Granovetter, M. Threshold models of collective behavior. *The American Journal of Sociology* **83**, 1420 (1978).

130. Graves, A., Mohamed, A.-r. & Hinton, G. *Speech Recognition with Deep Recurrent Neural Networks* arXiv:1303.5778. 2013.

131. Guitart, A., Chen, P. P., Bertens, P. & Periáñez, Á. Forecasting Player Behavioral Data and Simulating in-Game Events. *Advances in Intelligent Systems and Computing* **886**, 274 (2018).

132. Guitart, A., del Río, A. F. & Periáñez, Á. *Understanding Player Engagement and In-Game Purchasing Behavior with Ensemble Learning* in *2019 GAME-ON Conference on Simulation and AI in Computer Games (GAME-ON'2019)* Sept. 18–20, 2019 (EUROSIS, Breda University of Applied Sciences, Breda, The Netherlands, 2019).

133. Guitart, A., Tan, S. H., del Río, A. F. & Periáñez, Á. *From Non-Paying to Premium: Predicting User Conversion in Video Games with Ensemble Learning* in *ACM Foundations of Digital Games (FDG'2019)* Aug. 26–30, 2019. **97** (San Luis Obispo, USA, 2019).

134. Gupta, S. Impact of Sales Promotions on When, What, and How Much to Buy. *Journal of Marketing Research* **25**, 342 (1988).

135. Hadiji, F., Sifa, R., Drachen, A., Thurau, C., Kersting, K. & Bauckhage, C. *Predicting player churn in the wild* in *Computational Intelligence and Games (CIG), 2014 IEEE Conference on* (2014), 1.

136. Hamilton, J. D. *Time Series Analysis* 1st ed. (Princeton University Press, 1994).

137. Hannan, E. & Quinn, B. The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **41**, 190 (1979).

138. Harvey, A. *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge University Press, 1991).

139. Harvey, A. in *Handbook of Economic Forecasting* (eds Elliott, G., Granger, C. & Timmermann, A.) 1st ed., 327 (Elsevier, 2006).

140. Hegselmann, R. & Krause, U. Opinion dynamics and bounded confidence: models, analysis and simulation. *Journal of Artificial Societies and Social Simulation* **5** (2002).

141. Helbing, D., Farkas, I. & Vicsek, T. Simulating dynamical features of escape panic. *Nature* **407**, 487 (2000).

142. Helbing, D., Hennecke, A. & Treiber, M. Phase Diagram of Traffic States in the Presence of Inhomogeneities. *Physical Review Letters* **82**, 4360 (1999).

143. Herniter, J. A Probablistic Market Model of Purchase Timing and Brand Selection. *Management Science* **18**, P102 (1971).

144. Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **9**, 1735 (1997).

145. Hoekstra, J. C. & Huizingh, E. K. The lifetime value concept in customer-based marketing. *Journal of Market-Focused Management* **3**, 257 (1999).

146. Holyst, J. A., Kacperski, K. & Schweitzer, F. Phase transitions in social impact models of opinion formation. *Physica A* **285**, 199 (2000).

147. Hothorn, T., Hornik, K. & Zeileis, A. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* **15**, 651 (2006).

148. Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A. & Van Der Laan, M. Survival Ensembles. *Biostatistics* **7**, 355 (2006).

149. Hothorn, T., Lausen, B., Benner, A. & Radespiel-Tröger, M. Bagging survival trees. *Statistics in medicine* **23**, 77 (2004).

150. Hougaard, P. Fundamentals of survival data. *Biometrics* **55**, 13 (1999).

151. Hu, Y., Koren, Y. & Volinsky, C. *Collaborative Filtering for Implicit Feedback Datasets* in *2008 Eighth IEEE International Conference on Data Mining* (2008), 263.

152. Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing In Science & Engineering* **9**, 90 (2007).

153. Hwang, H., Jung, T. & Suh, E. An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert systems with applications* **26**, 181 (2004).

154. Hyndman, R. J. & Koehler, A. B. Another look at measures of forecast accuracy. *International Journal of Forecasting* **22**, 679 (2006).

155. Imry, Y. On the statistical mechanics of coupled order parameters. *Journal of Physics C: Solid State Physics* **8**, 567 (1975).

156. Ishwaran, H., Gerds, T. A., Kogalur, U. B., Moore, R. D., Gange, S. J. & Lau, B. M. Random survival forests for competing risks. *Biostatistics* **15**, 757 (2014).

157. Ishwaran, H., Kogalur, U. B., Blackstone, E. H., Lauer, M. S., *et al.* Random survival forests. *The annals of applied statistics* **2**, 841 (2008).

158. Ivakhnenko, A., Ivakhnenko, A., Lapa, V., Lapa, V., LAPA, V. & McDonough, R. *Cybernetics and Forecasting Techniques* (American Elsevier Publishing Company, 1967).

159. James, G., Witten, D., Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R* (Springer Publishing Company, Incorporated, 2014).

160. Jarque, C. M. & Bera, A. K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* **6**, 255 (1980).

161. Ji, W., Wang, X. & Zhu, F. Time-aware conversion prediction. *Frontiers of Computer Science* **11**, 702 (2017).

162. Johnson, N., Velasquez, N., Restrepo, N., Leahy, R., Gabriel, N., Oud, S., Zheng, M., Manrique, P., Wuchty, S. & Lupu, Y. The online competition between pro- and anti-vaccination views. *Nature* **582** (2020).

163. Jun Ding, D. G. & Chen, X. Alone in the Game: Dynamic Spread of Churn Behavior in a Large Social Network a Longitudinal Study in MMORPG. *falta* **24**, 123 (1996).

164. Kalman, R. E. A new approach to linear filtering and prediction problems. *Journal of basic Engineering* **82**, 35 (1960).

165. Kaufman, L. & Rousseeuw, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis* (Wiley, 1990).

166. Kawale, J., Pal, A. & Srivastava, J. *Churn prediction in MMORPGs: A social influence based approach* in *Proceedings of the International COnference on Computational Science and Engineering, 2009. CSE'09.* **4** (2009), 423.

167. Kim, K.-J., Yoon, D., Jeon, J., Yang, S.-i., Lee, S.-K., Lee, E., Jang, Y., Kim, D.-W., Chen, P. P., Guitart, A., Bertens, P., Periáñez, Á., Hadiji, F., Müller, M., Joo, Y., Lee, J. & Hwang, I. Game Data Mining Competition on Churn Prediction and Survival Analysis using Commercial Game Log Data. *IEEE Transactions on Games*, 1 (2018).

168. Kim, S., Choi, D., Lee, E. & Rhee, W. Churn prediction of mobile and online casual games using play log data. *PLoS ONE* **12**, e0180735 (2017).

169. Kingma, D. P. & Ba, J. *Adam: A method for stochastic optimization* in *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)* arXiv:1412.6980 (2014).

170. Kirman, A. in *Complex Systems* (eds Jean-Philippe Bouchaud, M. M. & Dalibard, J.) 217 (Elsevier, 2007).

171. Korutcheva, E. R. & Uzunov, D. I. Ising models and coupled order parameters. *Preprint Joint Inst. Nuclear Research Dubna, E-17-88-467.* (1988).

172. Kreyszig, E., Kreyszig, H. & Norminton, E. J. *Advanced Engineering Mathematics* Tenth (Wiley, Hoboken, NJ, 2011).

173. Krupakar, H., Rajvel, K., Bharathi, B., Deborah, S. A. & Krishnamurthy, V. A survey of voice translation methodologies — Acoustic dialect decoder. *2016 International Conference on Information Communication and Embedded Systems (ICICES)* (2016).

174. Kumar, V., Ramani, G. & Bohling, T. Customer lifetime value approaches and best practice applications. *Journal of Interactive Marketing* **18**, 60 (2004).

175. Laver, M. & Sergenti, E. *Party Competition: An Agent-Based Model* (Princeton University Press, 2012).

176. Lawrence, R. J. The Lognormal Distribution of Buying Frequency Rates. *Journal of Marketing Research* **17**, 212 (1980).

177. LeCun, Y., Bengio, Y., *et al.* Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* **3361**, 1995 (1995).

178. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436 (2015).

179. LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. & Jackel, L. D. Backpropagation applied to handwritten zip code recognition. *Neural computation* **1**, 541 (1989).

180. Lin, H. & Tegmark, M. Critical Behavior in Physics and Probabilistic Formal Languages. *Entropy* **19**, 299 (2017).

181. Ljung, G. M. & Box, G. E. P. On a measure of lack of fit in time series models. *Biometrika* **65**, 297 (1978).

182. Lorenz, J. Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C* **18**, 1819 (2007).

183. Luton, W. *Free-to-play: Making money from games you give away* (New Riders, 2013).

184. Ma, Y. P., Goncalves, S., Mignot, S., Nadal, J. P. & Gordon, M. B. Cycles of cooperation and free-riding in social systems. *European Physical Journal B* **71**, 597 (2009).

185. Mantegna, R. & Stanley, H. *An Introduction to Econophysics* (Cambridge University Press, 1999).

186. Mantegna, R. & Kertész, J. Focus on Statistical Physics Modeling in Economics and Finance. *New Journal of Physics* **13**, 025011 (2011).

187. Marsan, G. A., Bellomo, N. & Gibelli, L. Stochastic evolutionary differential games toward a systems theory of behavioral social dynamics. *Mathematical Models and Methods in Applied Sciences* **26**, 1051 (2016).

188. Martino, A. D. & Marsili, M. Statistical mechanics of socio-economic systems with heterogeneous agents. *Journal of Physics A: Mathematical and General* **39**, R465 (2006).

189. Martins, A. C. R. Mobility and social network effects on extremist opinions. *Phys. Rev. E* **78**, 036104 (3 2008).

190. Marzo, G. D., Zaccaria, A. & Castellano, C. *Emergence of polarization in a voter model with personalized information* arXiv:2007.04903. 2020.

191. Mason, L., Baxter, J., Bartlett, P. & Frean, M. *Boosting Algorithms as Gradient Descent* in *Proceedings of the 12th International Conference on Neural Information Processing Systems* (MIT Press, Denver, CO, 1999), 512.

192. Mason, L., Baxter, J., Bartlett, P. & Frean, M. *Boosting Algorithms as Gradient Descent in Function Space* 1999.

193. McCarty, J. A. & Hastak, M. Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of Business Research* **60**, 656 (2007).

194. McCulloch, W. S. & Pitts, W. H. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* **5**, 115 (1943).

195. McFadden, D. *Conditional Logit Analysis of Qualitative Choice Behavior* (ed Zarembka, P.) 105 (Academic Press: New York, 1974).

196. McFadden, D. *Econometric analysis of qualitative response models* (eds Griliches, Z. & Intriligator, M.) (Amsterdam: North-Holland, 1984).

197. McInnes, L., Healy, J. & Astels, S. hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software* **2**, 205 (2017).

198. McKinney, W. *Data Structures for Statistical Computing in Python* in *Proceedings of the 9th Python in Science Conference* (eds van der Walt, S. & Millman, J.) (2010), 51.

199. Mermin, N. D. & Wagner, H. Absence of Ferromagnetism or Antiferromagnetism in One- or Two-Dimensional Isotropic Heisenberg Models. *Phys. Rev. Lett.* **17**, 1133 (22 1966).

200. Mezard, M., Parisi, G. & Virasoro, M. A. *Spin Glass Theory and Beyond* (World Scientific, 1987).

201. Michard, Q. & Bouchaud, J.-P. Theory of collective opinion shifts: from smooth trends to abrupt swings. *The European Physical Journal B - Condensed Matter and Complex Systems* **47**. 10.1140/epjb/e2005-00307-0, 151 (1 2005).

202. Mittal, S., Madigan, D., Burd, R. S. & Suchard, M. A. High-dimensional, massive sample-size Cox proportional hazards regression for survival analysis. *Biostatistics* **15**, 207 (2013).

203. Mogensen, U. B., Ishwaran, H. & Gerds, T. A. Evaluating random forests for survival analysis using prediction error curves. *Journal of statistical software* **50**, 1 (2012).

204. Mondal, P., Shit, L. & Goswami, S. Study of Effectiveness of Time Series Modeling (ARIMA) in Forecasting Stock Prices. *International Journal of Computer Science, Engineering and Applications* **4** (2014).

205. Monechi, B., Ibáñez-Berganza, M. & Loreto, V. *Hamiltonian Modeling of Macro-Economic Urban Dynamics* arXiv:2001.05725. 2020.

206. Mozer, M. C., Wolniewicz, R. H., Grimes, D. B., Johnson, E. & Kaushansky, H. *Churn reduction in the wireless industry* in *Advances in Neural Information Processing Systems* (2000), 935.

207. Mutanen, T., Ahola, J. & Nousiainen, S. *Customer churn prediction-a case study in retail banking* in *Proc. of ECML/PKDD Workshop on Practical Data Mining* (2006), 13.

208. Nadal, J. P., Weisbuch, G., Chenevez, O. & Kirman, A. *A formal approach to market organisation: choice functions, mean field approximation and the maximum entropy principle* (eds Lesourne, J. & Orléan, A.) (Economica, London, 1998).

209. Nakayama, S. & Nakamura, Y. A fashion model with social interaction. *Physica A - Statistical Mechanics and Applications* **337**, 625 (2004).

210. Nash, J. Non-Cooperative Games. *Annals of Mathematics* **54**, 286 (1951).

211. Nash, J. F. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences* **36**, 48 (1950).

212. Newman, M. E. J. Models of the small world. *Journal of Statistical Physics* **101**, 819 (2000).

213. Newman, M. E. J. The structure and function of complex networks. *Siam Review* **45**, 167 (2003).

214. Notarmuzi, D. & Castellano, C. Analytical study of quality-biased competition dynamics for memes in social media. *EPL (Europhysics Letters)* **122**, 28002 (2018).

215. Novikov, A. PyClustering: Data Mining Library. *Journal of Open Source Software* **4**, 1230 (2019).

216. Oliphant, T. *NumPy: A guide to NumPy*. USA: Trelgol Publishing. http://www.numpy.org/. 2006.

217. Orléan, A. Bayesian interactions and collective dynamics of opinion: Herd behavior and mimetic contagion. *Journal of Economic Behavior and Organization* **28**. Workshop on the Emergence and Stability of Institutions, Louvain, Belgium, Dec 1992, 257 (1995).

218. Ormerod, P. & Campbell, M. *The evolution of family structures in a social context*. Preprint. 1998.

219. Ortega, D., Rodríguez-Laguna, J. & Korutcheva, E. A Schelling model with a variable threshold in a closed city segregation model. Analysis of the universality classes. **574**, 1260110 (2021).

220. Ortega, D., Rodríguez-Laguna, J. & Korutcheva, E. Avalanches in an extended Schelling model: an explanation of urban gentrification. **574**, 125943 (2021).

221. Palchykov, V., von Ferber, C., Folk, R. & Holovatch, Y. Coupled order-parameter system on a scale-free network. *Phys. Rev. E* **80**, 011108 (2009).

222. Parisi, G. *Field Theory, Disorder and Simulation* (World Scientific, 1992).

223. Parzen, E. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* **33**, 1065 (1962).

224. Periáñez, Á., Guitart, A., Chen, P. P. & del Río, A. F. in *Data Analitycs Applications in Gaming and Entertainment* (ed Wallner, G.) 119 (Auerbach Publications, USA, 2019).

225. Periáñez, Á., Saas, A., Guitart, A. & Magne, C. *Churn Prediction in Mobile Social Games: Towards a Complete Assessment Using Survival Ensembles* in *IEEE Conference on Data Science and Advanced Analytics (DSAA 2016)* Oct. 17–19, 2016 (IEEE, Montreal, Canada, 2016), 564.

226. Permatasari, C. I., Sutopo, W. & Hisjam, M. *Sales forecasting newspaper with ARIMA: A case study* in. **1931** (2018).

227. Platzer, M. Customer Base Analysis with BTYDplus (2016).

228. Platzer, M. & Reutterer, T. Ticking away the moments: Timing regularity helps to better predict customer activity. *Marketing Science* **35**, 779 (2016).

229. Prechelt, L. in *Neural Networks: Tricks of the Trade Lecture Notes in Computer Science* 1524, 55 (Springer, 1998).

230. Prentice, R. L., Kalbfleisch, J. D. & Peterson, A. V. The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541 (1978).

231. Pyle, D. *Data preparation for data mining* (Morgan Kaufmann Publishers, 1999).

232. Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G. R., Duggan, G. E., Irvine, J., Le, Q. V., Litsch, K., Mossin, A., Tansuwan, J., Wang, D., Wexler, J., Wilson, J., Ludwig, D., Volchenboum, S. L., Chou, K. J., Pearson, M. A., Madabushi, S., Shah, N. H., Butte, A. J., Howell, M. D., Cui, C., Corrado, G. S. & Dean, J. Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine* **1** (2018).

233. Rendle, S., Freudenthaler, C., Gantner, Z. & Schmidt-Thieme, L. *BPR: Bayesian Personalized Ranking from Implicit Feedback* arXiv:1205.2618. cite arxiv:1205.2618Comment: Appears in Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI2009). 2012.

234. Roondiwala, M., Patel, H. & Varma, S. Predicting Stock Prices Using LSTM. *International Journal of Science and Research (IJSR)* **6** (2017).

235. Rosenblatt, M. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics* **27**, 832 (1956).

236. Rothenbuehler, P., Runge, J., Garcin, F. & Faltings, B. *Hidden markov models for churn prediction* in *Proccedings of SAI Intelligent Systems Conference (IntelliSys), 2015* (2015), 723.

237. Ruder, S. *An overview of gradient descent optimization algorithms* arXiv:1609.04747. 2017.

238. Runge, J., Gao, P., Garcin, F. & Faltings, B. *Churn Prediction for High-value Players in Casual Social Games* in *Computational Intelligence and Games (CIG), 2014 IEEE Conference on* (2014), 1.

239. Saas, A., Guitart, A. & Periáñez, Á. *Discovering playing patterns: Time series clustering of free-to-play game data* in *IEEE Conference on Computational Intelligence and Games (CIG 2016)* Sept. 20–23, 2016 (IEEE, Santorini, Greece, 2016), 1.

240. Santalla, S. N., Koroutchev, K., Korutcheva, E. & Rodríguez-Laguna, J. Power accretion in social systems. *Physical Review E* **100**, 012143 (2019).

241. Sateesh Babu, G., Zhao, P. & Li, X.-L. *Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life* in *Database Systems for Advanced Applications* (eds Navathe, S. B., Wu, W., Shekhar, S., Du, X., Wang, X. S. & Xiong, H.) (Springer International Publishing, Cham, 2016), 214.

242. Schelling, T. Dynamic models of segregation. *Journal of Mathematical Sociology* **1**, 143 (1971).

243. Schelling, T. Hockey Helmets, Concealed Weapons, and Daylight saving: A Study of Binary Choices with Externalities. *Journal of Conflict Resolution* **17**, 381 (1973).

244. Schelling, T. *Micromotives and macrobehaviour* (Norton, 1978).

245. Scherer, D., Müller, A. & Behnke, S. in *Artificial Neural Networks–ICANN 2010* 92 (Springer, 2010).

246. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Networks* **61**, 85 (2015).

247. Schmittlein, D. C., Morrison, D. G. & Colombo, R. Counting your customers: Who-are they and what will they do next? *Management Science* **33**, 1 (1987).

248. Schmittlein, D. C. & Morrison, D. G. Prediction of Future Random Events with the Condensed Negative Binomial Distribution. *Journal of the American Statistical Association* **78**, 449 (1983).

249. Schwarz, G. Estimating the dimension of a model. *The Annals of Statistics* **6**, 461 (1978).

250. Seabold, S. & Perktold, J. *Statsmodels: Econometric and statistical modeling with python* in *9th Python in Science Conference* (2010).

251. Semeshenko, V., Gordon, M. B. & Nadal, J.-P. Collective states in social systems with interacting learning agents. *Physica A: Statistical Mechanics and its Applications* **387**, 4903 (2007).

252. Shakti, S., Hassan, M., Zhenning, Y., Caytiles, R. & Iyenger, N. C. S. N. Annual Automobile Sales Prediction Using ARIMA Model. *International Journal of Hybrid Information Technology* **10**, 13 (2017).

253. Shaw, R. & Stone, M. Database marketing. *Gower* (1988).

254. Sherstinsky, A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena* **404**, 132306 (2020).

255. Shumway, R. & Stoffer, D. *Time Series Analysis and Its Applications: With R Examples* (Springer New York, 2010).

256. Sifa, R., Hadiji, F., Runge, J., Drachen, A., Kersting, K. & Bauckhage, C. *Predicting purchase decisions in mobile free-to-play games* in *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference* (2015).

257. Sifa, R., Runge, J., Bauckhage, C. & Klapper, D. *Customer Lifetime Value Prediction in Non-Contractual Freemium Settings: Chasing High-Value Users Using Deep Neural Networks and SMOTE* in *HICSS* (2018).

258. Simm, J., de Abril, I. M. & Sugiyama, M. *Tree-Based Ensemble Multi-Task Learning Method for Classification and Regression* 6 (The Institute of Electronics, Information and Communication Engineers, 2014), 1677.

259. Simon, P. & Ricci-Tersenghi, F. Coupled Ising models with disorder. *Journal of Physics A: Mathematical and General* **33**, 5985 (2000).

260. Sîrbu, A., Loreto, V., Servedio, V. D. P. & Tria, F. Opinion Dynamics: Models, Extensions and External Effects. *Participatory Sensing, Opinions and Collective Awareness*, 363 (2016).

261. Smith, W. R. Product Differentiation and Market Segmentation as Alternative Marketing Strategies. *Journal of Marketing* **21**, 3 (1956).

262. Stanley, H. E., Plerou, V. & Gabaix, X. A statistical physics view of financial fluctuations: Evidence for scaling and universality. *Physica A-Statistical Mechanics and its Applications* **387**. 6th International Conference on Applications of Physics in Financial Analysis, Lisbon, 2007, 3967 (2008).

263. Stanley, M. H. R., Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Leschhorn, H., Maass, P., Salinger, M. A. & Stanley, H. E. Scaling behaviour in the growth of companies. *Nature* **379**, 804 (1996).

264. Stauffer, D. Econophysics - A new area for computational statistical physics? *International Journal of Modern Physics C* **11**, 1081 (2000).

265. Stauffer, D. Sociophysics: the Sznajd model and its applications. *Computer Physics Communications* **146**, 93 (2002).

266. Stauffer, D. Introduction to statistical physics outside physics. *Physica A: Statistical and Theoretical Physics* **336**. Proceedings of the XVIII Max Born Symposium 'Statistical Physics outside Physics', 1 (2004).

267. Su, X. & Khoshgoftaar, T. A Survey of Collaborative Filtering Techniques. *Adv. Artif. Intell.* **2009**, 421425:1 (2009).

268. Sülo, I., Keskin, S. R., Dogan, G. & Brown, T. *Energy Efficient Smart Buildings: LSTM Neural Networks for Time Series Prediction* in *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)* (2019), 18.

269. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. *Going deeper with convolutions* in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), 1.

270. Sznajd-Weron, K. & Sznajd, J. Opinion evolution in closed community. *International Journal of Modern Physics C* **11**, 1157 (2000).

271. Sznajd-Weron, K. & Sznajd, J. Who is left, who is right? *Physica A: Statistical Mechanics and its Applications* **351**, 593 (2005).

272. Takács, G., Pilászy, I. & Tikk, D. *Applications of the conjugate gradient method for implicit feedback collaborative filtering* in *RecSys '11* (2011).

273. Tektas, M. Weather Forecasting Using ANFIS and ARIMA MODELS. *Environmental Research, Engineering and Management* **51** (2010).

274. Tessone, C., Toral, R., Amengual, P., Wio, H. & Miguel, M. S. Neighborhood models of minority opinion spreading. *European Physical Journal B* **39**, 535 (2004).

275. Tkachenko, Y. *Autonomous CRM control via CLV approximation with deep reinforcement learning in discrete and continuous action space* arXiv:1504.01840. 2015.

276. Tolimieri, N., Holmes, E. E., Williams, g. D., Pacunski, R. & Lowry, D. Population assessment using multivariate time-series analysis: A case study of rockfishes in Puget Sound. *Ecology and Evolution* **7**, 2846 (8 2017).

277. Tria, F., Loreto, V. & Servedio, V. Zipf's, Heaps' and Taylor's Laws are Determined by the Expansion into the Adjacent Possible. *Entropy* **20**, 752 (2018).

278. Tsantekidis, A., Passalis, N., Tefas, A., Kanniainen, J., Gabbouj, M. & Iosifidis, A. *Forecasting Stock Prices from the Limit Order Book Using Convolutional Neural Networks* in *2017 IEEE 19th Conference on Business Informatics (CBI)* **01** (2017), 7.

279. Turaga, S. C., Murray, J. F., Jain, V., Roth, F., Helmstaedter, M., Briggman, K., Denk, W. & Seung, H. S. Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural computation* **22**, 511 (2010).

280. Ubaldi, E., Burioni, R., Loreto, V. & Tria, F. *The exploration of the Adjacent Possible explains the emergence and evolution of social networks* arXiv:2003.00989. 2020.

281. Van Ham, M., Uesugi, M., Tammaru, T., Manley, D. & Janssen, H. Changing occupational structures and residential segregation in New York, London and Tokyo. *Nature Human Behaviour* (2020).

282. Vapnik, V. N. *The Nature of Statistical Learning Theory* (Springer, 1999).

283. Verhulst, P.-F. Recherches mathématiques sur la loi d'accroissement de la population. *Journal des économistes* **12**, 276 (1845).

284. Verhulst, P.-F. Notice sur la loi que la population suit dans son accroissement. *Corresp. Math. Phys.* **10**, 113 (1838).

285. Voigt, S. & Hinz, O. Making digital freemium business models a success: Predicting customers' lifetime value via initial purchase information. *Business & Information Systems Engineering* **58**, 107 (2016).

286. Wang, J., Zhang, Y., Posse, C. & Bhasin, A. *Is it time for a career switch?* in *Proceedings of the 22nd international conference on World Wide Web* (2013), 1377.

287. Weisbuch, G., Kirman, A. & Herreiner, D. Market Organisation and trading relationships. *The Economic Journal* **110**, 411 (2000).

288. Weisbuch, G. & Stauffer, D. Adjustment and social choice. *Physica A* **323**, 651 (2003).

289. Weisbuch, G. From Anti-Conformism to Extremism. *Journal of Artificial Societies and Social Simulation* **18**, 1 (2015).

290. Wheat, R. D. & Morrison, D. G. Estimating purchase regularity with two interpurchase times. *Journal of Marketing Research* **27**, 87 (1990).

291. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).

292. Williams, T., Kelley, C. & many others. *Gnuplot 4.6: an interactive plotting program* `http://gnuplot.sourceforge.net/`. 2013.

293. Wu, C., Guo, S., Hong, Y., Xiao, B., Wu, Y., Zhang, Q., Initiative, A. D. N., *et al.* Discrimination and conversion prediction of mild cognitive impairment using convolutional neural networks. *Quantitative Imaging in Medicine and Surgery* **8**, 992 (2018).

294. Wu, F. Y. The Potts model. *Rev. Mod. Phys.* **54**, 235 (1 1982).

295. Wu, F. & Huberman, B. A. *Social Structure and Opinion Formation* Computational Economics (EconWPA, 2004).

296. Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M. & Dean, J. *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation* arXiv:1609.08144. 2016.

297. Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K. C., Hua, Y., Gueroussov, S., Najafabadi, H. S., Hughes, T. R., Morris, Q., Barash, Y., Krainer, A. R., Jojic, N., Scherer, S. W., Blencowe, B. J. & Frey, B. J. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347** (2015).

298. Yang, B., Sun, S., Li, J., Lin, X. & Tian, Y. Traffic Flow Prediction Using LSTM with Feature Enhancement. *Neurocomputing* **332**, 320 (2018).

299. Yang, J. B., Nguyen, M. N., San, P. P., Li, X. L. & Krishnaswamy, S. *Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition* in *Proceedings of the 24th International Conference on Artificial Intelligence* (AAAI Press, Buenos Aires, Argentina, 2015), 3995.

300. Yin, W., Kann, K., Yu, M. & Schütze, H. *Comparative Study of CNN and RNN for Natural Language Processing* arXiv:1702.01923. 2017.

301. Young, A. *Spin Glasses and Random Fields* (World Scientific, 1997).

302. Zeng, Q., Li, D., Huang, G., Wang, X., Zhang, Y., Tang, W. & Zhou, H. Time series analysis of temporal trends in the pertussis incidence in Mainland China from 2005 to 2016. *Nature Scientific Reports* **6** (2016).

303. Zhang, X., Zhang, T., Young, A. A. & Li, X. Applications and Comparisons of Four Time Series Models in Epidemiological Surveillance Data. *PLOS ONE* **9**, 1 (2014).

304. Zhou, C., Fang, Z., Xu, X., Zhang, X., Ding, Y., Jiang, X. & ji, Y. Using long short-term memory networks to predict energy consumption of air-conditioning systems. *Sustainable Cities and Society* **55** (2020).

PUBLICATIONS

Articles in peer-reviewed journals:

1. Del Río, A. F., Guitart, A. & Periáñez, Á. A Time Series Approach To Player Churn and Conversion in Videogames. *Intelligent Data Analysis* **25** (2021).
2. Del Río, A. F., Korutcheva, E. & de la Rubia, J. Interdependent binary choices under social influence: phase diagram for homogeneous unbiased populations. *Complexity* **17**, 31 (2012).

Conference papers:

1. Guitart, A., del Río, A. F. & Periáñez, Á. *Understanding Player Engagement and In-Game Purchasing Behavior with Ensemble Learning* in *2019 GAME-ON Conference on Simulation and AI in Computer Games (GAME-ON'2019)* Sept. 18–20, 2019 (EUROSIS, Breda University of Applied Sciences, Breda, The Netherlands, 2019).
2. Del Río, A. F., Chen, P. P. & Periáñez, Á. *Profiling Players with Engagement Predictions* in *2019 IEE Conference on Games (CoG)* Aug. 20–23, 2019 (IEE, London, England, 2019).
3. Guitart, A., Tan, S. H., del Río, A. F. & Periáñez, Á. *From Non-Paying to Premium: Predicting User Conversion in Video Games with Ensemble Learning* in *ACM Foundations of Digital Games (FDG'2019)* Aug. 26–30, 2019. **97** (San Luis Obispo, USA, 2019).
4. Chen, P. P., Guitart, A., del Río, A. F. & Periáñez, Á. *Customer Lifetime Value in Video Games Using Deep Learning anf Parametric Models* in *IEEE International Conference on Big Data (Big Data 2018)* Dec. 10–13, 2018 (IEEE, Seattle, USA, 2018), 2134.

Book chapter:

1. Periáñez, Á., Guitart, A., Chen, P. P. & del Río, A. F. in *Data Analitycs Applications in Gaming and Entertainment* (ed Wallner, G.) 119 (Auerbach Publications, USA, 2019).

Masters thesis:

1. Del Río, A. F. *Coupled Ising models and interdependent discrete choices under social influence in homogeneous populations* MA thesis (Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain, 2011).

All publications except the book chapter are available through arXiv at:

`https://arxiv.org/a/fernandezdelrio_a_1.html`

Additionally, some of the content of this thesis was presented by the author at the following workshops and confrences:

1. *Understanding player conversion and churn: from game dynamics to individual player predictions.* IEE Conference on Games (CoG). Online. 2020.
2. *What can AI and big data do for video games?* Quo Vadis (gamesweekberlin). Berlin (Germany). 2019.
3. *Interdependent choices under social influence.* DPG. Berlin (Germany). 2013.
4. *Interdependent choices, individual preferences and social influence.* FISES. Barcelona (Spain). 2012.
5. *Interdependent choices under social influence.* The unexpected conference: new trends in sociophysics. Paris (France). 2011.