
ARTÍCULOS / ARTICLES

ANÁLISIS DE LOS PATRONES ESPACIO-TEMPORALES DE EVENTOS A PARTIR DE DATOS DE TWITTER: EL CASO DE LA *WORLD PRIDE* 2017 EN MADRID

Joaquín Osorio Arjona

Departamento de Geografía Universidad Complutense de Madrid
joaquoso@ucm.es
ORCID iD: <https://orcid.org/0000-0002-0102-8756>

Recibido: 30/04/2019; Aceptado: 27/11/2019.

Cómo citar este artículo/Citation: Osorio Arjona, Joaquín (2020). Análisis de los patrones espacio-temporales de eventos a partir de datos de Twitter: el caso de la *World Pride* 2017 en Madrid. *Estudios Geográficos*, 81 (288), e032. <https://doi.org/10.3989/estgeogr.202047.027>

RESUMEN: Este trabajo analiza los patrones espaciotemporales de un macroevento en una ciudad a partir de nuevas fuentes de datos, partiendo como hipótesis que las multitudes registran una alta actividad en las redes sociales durante los programas del evento. Identificando usuarios que han publicado tweets geolocalizados en el centro de Madrid durante la *World Pride* 2017, se puede localizar su procedencia, y evaluar el impacto del evento a nivel espaciotemporal a partir de la comparación con los resultados observados durante una semana habitual. Los resultados obtenidos muestran un crecimiento del número de usuarios extranjeros y un fuerte aumento de la actividad en las principales zonas del evento mientras que la actividad de las zonas más alejadas disminuye.

PALABRAS CLAVE: mega eventos; redes sociales; puntos calientes; SIG; huella digital; huella espacio-temporal

ANALYSIS OF THE SPACE-TEMPORAL PATTERNS OF EVENTS FROM TWITTER DATA: THE CASE OF MADRID 2017 *WORLD PRIDE*

ABSTRACT: This work analyses the spatio-temporal patterns of a mass event in a city from new data sources, starting from the hypothesis that crowds register high activity in social networks during the event programs. Identifying users who have posted geolocated tweets in the centre of Madrid during the 2017 *World Pride*, their origin cities and countries can be located, and the impact of the event at a space-time level can be evaluated from the comparison with the observed results during a regular week. The obtained results show a growth in the number of foreign users and a strong increase in activity in the main areas of the event, while activity in the more remote areas decreases.

KEYWORDS: mega-events; social network; hot spots; GIS; digital footprint; space-time footprint

INTRODUCCIÓN

Los eventos de masas son fenómenos con alta repercusión y seguimiento por parte de la población debido a su magnitud, importancia en la sociedad y publicidad. Los megaeventos se pueden definir como eventos a gran escala que están orientados especialmente al mercado del turismo internacional. Destacan por su tamaño en términos de asistencia, financiación pública, cobertura televisiva, construcción de infraestructuras e impacto en los residentes de la ciudad o país organizador (Knott, *et al.*, 2015). Eventos a gran escala como los Juegos Olímpicos, exposiciones mundiales, conciertos, etc. ocurren cada año en todo el mundo, atrayendo a un gran número de participantes y turistas que viajan a un destino concreto, principalmente una gran ciudad (Xu y González, 2017).

Los festivales y grandes eventos tienen una gran importancia para la administración y organización de una ciudad. El principal motivo de su celebración es el impacto económico que generan los turistas a partir de su consumo (Batista *et al.*, 2018). Los efectos de la celebración de estos eventos se dan en varias áreas (económica, comercial, infraestructuras, política, medioambiental, etc.). Estos efectos pueden ser positivos (aumento del turismo, de las oportunidades de empleo, reconocimiento y visibilidad a nivel internacional, atracción de capital y empresas, desarrollo de infraestructuras, rejuvenecimiento de áreas urbanas, aumento de la seguridad, oportunidades de participación, atención al medioambiente, etc.), pero también negativos (centralización de recursos, aumento de los precios, de la especulación inmobiliaria, importante generación de desechos, problemas de congestión y contaminación, pérdida de control de recursos en detrimento de las empresas privadas, contratos basura, etc.), (Knott *et al.*, 2015).

Para facilitar la organización y gestión de los grandes eventos, y poder controlar los efectos que tienen sobre la ciudad, las empresas y ayuntamientos necesitan información para realizar diagnósticos eficaces del impacto que tienen sobre la ciudad. Esta información se obtiene a partir de datos. Sin embargo, adquirir una alta cantidad de datos en un área geográfica usando métodos convencionales consume tiempo y es caro (Cáceres, 2007). Hasta hace unos años, la principal fuente de información provenía de encuestas, formularios o conteos, métodos que necesitan tiempo y trabajo para elaborarse, son costosos, y normalmente carecen de infor-

mación espacial. Además, las fuentes de datos tradicionales como los censos ofrecen información en la distribución espacial de la población durante la noche (lugar de residencia) pero no en su localización durante el día (Moya *et al.*, 2017; Osorio y García-Palomares, 2019). Sin embargo, el desarrollo de las Tecnologías de la Información y Comunicación (TIC) y la irrupción del *Big Data* están proporcionando desde hace pocos años fuentes de información con una alta escala espacial y temporal, poco costosas (en muchos casos gratuita) y de libre acceso en general (Osorio y García-Palomares, 2017).

Recientemente, empresas públicas y privadas han empezado a usar el *Big Data* como fuente de información para la administración y gestión de las ciudades gracias a estas ventajas. Los datos masivos han alcanzado tanto valor que se comenta que son el petróleo de la economía del siglo XXI (Gutiérrez, 2018). Se puede tomar las ciudades como armazones físicos cuyo sistema nervioso son los datos (Tascón y Coullaut, 2016). El uso de los datos generados por sensores y otras nuevas tecnologías para la gestión de la ciudad ha acuñado en los últimos años conceptos nuevos como las *Smart Cities*, o *data driven cities*. La participación ciudadana como generadores de datos se antoja valiosa para los organismos públicos y privados, ya que la huella digital generada puede ayudar al diseño y funcionalidad de la ciudad (Masala y Pallares-Barbera, 2016). Los usuarios de Internet han pasado de tener una actitud pasiva como receptor de información a adquirir una actitud activa como generador de contenidos cuando interactúan en Internet, pudiendo entender el *Big Data* como un gran repositorio con datos de la actividad de la gente (Gutiérrez, 2018; Gutiérrez y García-Palomares, 2016; Miller, 2010). Un ejemplo del uso de fuentes de datos masivos para la administración y gestión de eventos se halla en el uso del *Big Data* por parte del Ayuntamiento de Sevilla para contabilizar el número de desplazamientos a pie en la ciudad durante la Semana Santa.¹

Las redes sociales son una de las herramientas con las que se puede conseguir datos masivos y de alta riqueza espaciotemporal. Definidas como “formas de interacción entre gente que crean, comparten, e intercambian información e ideas en redes y comunidades virtuales”, estas herramientas son canales de comunicación entre personas y enlaces que forman una comunidad virtual, pero que afecta a los comportamientos de la gente y la vida real (Zeng y Gerritsen, 2014). Entre las redes sociales destaca Twitter por su alto uso, su facilidad de descarga de datos y

su posibilidad de contar con datos geolocalizados (Osorio y García-Palomares, 2017).

Los eventos, sobre todo los más importantes a nivel mundial (pero también los festivales de menor entidad), cuentan con un seguimiento que se magnifica en las redes sociales. En España en el año 2016, las dos principales búsquedas realizadas en *Google* fueron sobre los Juegos Olímpicos de Río de Janeiro y la Eurocopa celebrada en Francia. Mientras, en Twitter, el *trending topic* o tema más comentado a lo largo del año ha sido el Festival de Eurovisión celebrado en Kiev, seguido de los Juegos Olímpicos y la Eurocopa en cuarto lugar. Además, las organizaciones usan las redes sociales como una base fundamental para publicitar los eventos de masas con el fin de conseguir una mayor repercusión (Leung *et al.*, 2013).

La *World Pride* es uno de los eventos que más relevancia a nivel internacional ha tenido en redes sociales en estos últimos años, por la magnitud de visitantes que la ciudad organizadora recibe de todo el mundo, y por la polémica que conlleva su propia temática. Este evento tiene como objetivo la promoción a nivel global de aspectos relacionados con el colectivo de lesbianas, gays, bisexuales, y transexuales (LGBT) mediante festivales, conciertos, desfiles y otros tipos de actividades. Redes sociales como Twitter son vehículos de comunicación bastante utilizados por este colectivo para ganar visibilidad gracias a elementos como las tendencias o *trending topics*.

Se pueden observar dos tipos de eventos a partir de sus temas y características en Twitter. Una primera categoría englobaría a los eventos globales, destacando su gran volumen de mensajes, y la mayor importancia del tema del evento respecto a la ubicación (un ejemplo sería la *World Pride*, evento a nivel mundial cuya importancia radica en su temática). El segundo tipo de eventos se sitúan a nivel local, con una cantidad menor de mensajes, y donde la ubicación es tan importante como la temática (los festivales anuales del Orgullo de Madrid destacan tanto en su temática como en el hecho de que es una festividad centrada en la propia ciudad de Madrid) (Liu *et al.*, 2018).

El objetivo principal de este trabajo es estudiar el impacto que ha tenido la *World Pride* 2017 en la ciudad de Madrid en comparación con una semana habitual a partir de datos geolocalizados de Twitter. Estos datos son fáciles de explotar en un Sistema de Información Geográfica (SIG), donde se tratan y depuran para su análisis. A partir de análisis geoestadísticos y visualización cartográfica se pueden com-

parar los resultados obtenidos y evaluar el impacto del evento, partiendo de la hipótesis de que cuanto mayor es la densidad de la gente en un sitio, mayor es la posibilidad de usar una herramienta de comunicación como las redes sociales para compartir la actividad, opinión, o posición de una persona (Masala y Pallares-Barbera, 2016). Además, el análisis de este evento supone una puesta en valor de los datos obtenidos por Twitter, convirtiéndolos de información a conocimiento.

La originalidad del presente trabajo radica en el uso de datos geolocalizados de Twitter para investigar en profundidad el impacto de un mega evento como la *World Pride* sobre una ciudad. Mientras que ha habido varios trabajos basados en la detección de eventos a partir del contenido, hay muy pocos estudios realizados sobre la huella espacio-temporal de eventos a partir de redes sociales. Para ello se propone el uso de análisis de estadística espacial y cartografía para ahondar en la impronta espacial del evento, en contraposición con otros trabajos que han empleado Twitter, pero se han limitado a visualizar el conteo de tweets. Además, este estudio otorga una metodología de identificación del país o provincia de origen de cada visitante a partir de la descarga de los 3.200 últimos mensajes de cada usuario detectado en el caso de estudio, un aspecto que no se ha tratado en trabajos previos y que multiplica el valor de la aportación de este trabajo. Al utilizar identificadores numéricos y trabajar en una escala de provincias y países se mitigan los problemas de seguridad y privacidad que conlleva el uso de datos de Twitter.

El siguiente trabajo está dividido en cinco secciones. A continuación de esta introducción, en la que se ha presentado el tema de la investigación, el objetivo y justificación del trabajo, la sección 2 desarrollará un marco teórico sobre la oportunidad del uso de nuevas fuentes de datos en la investigación, y realizará una revisión de investigaciones relacionadas con estas fuentes de datos y el impacto de los eventos de masas. La sección 3 revisará el caso de estudio, datos y metodología empleados en la investigación, mientras que en la sección 4 se analizarán y discutirán los resultados obtenidos. Finalmente, se establecerán una serie de conclusiones en la sección 5.

ESTADO DEL ARTE

La revolución tecnológica de las TIC ha resuelto limitaciones relacionadas con la generación, explota-

ción y difusión de datos basados en las dimensiones del *Big Data* (volumen, variedad, veracidad, velocidad, valor y visualización), ofreciendo nuevas fuentes de información que mejoran y complementan las fuentes estadísticas. El continuo desarrollo de Internet y de la telefonía móvil permite crear, almacenar y compartir información de forma instantánea y accesible a cualquier lector desde cualquier punto del planeta (Gutiérrez-Puebla y García-Palomares, 2016; Kitchin, 2013). Los dispositivos móviles tienen una penetración extensa de mercado en muchos países y pueden ser utilizados para identificar localizaciones de forma precisa gracias a su sistema GPS, de forma rápida y accesible y sin las limitaciones tradicionales de coste y cobertura (Bar-Gera, 2007; Cáceres, 2012). Se estima que el 80% de los datos provenientes del *Big Data* están geolocalizados (Leszczynski y Crampton, 2016). Estos datos constituyen una materia prima muy valiosa para el estudio del comportamiento humano, ya que permiten el análisis de pautas espaciales y procesos que no pueden ser estudiados con estadísticas oficiales o encuestas (Gutiérrez-Puebla, 2018). En el campo teórico de la Geografía, las nuevas fuentes de datos cuentan con un gran potencial en forma de información abundante y por la posibilidad de promover su valor académico a una mayor audiencia (Kitchin, 2013; Kitchin, 2014).

Al igual que estas fuentes de datos presentan una serie de fortalezas, también poseen retos y limitaciones como el acceso, la calidad y propósito de los datos, el sesgo demográfico, la cobertura espacial y temporal, y la posible falta de estructura de los datos (Lansley *et al.* 2018; Miller y Goodchild, 2014). Hay que tener también en cuenta que los datos descargados por fuentes como redes sociales no están creados para analizar la movilidad urbana por lo que han de ser tratados por una serie de herramientas y habilidades para poder sacar información y convertirla en conocimiento (Osorio y García-Palomares, 2019).

Twitter es la red social más utilizada en estudios urbanos debido a su considerable alcance y a la descarga gratuita de los datos en Internet (Moya *et al.*, 2017). Los usuarios pueden permitir que Twitter registre las coordenadas GPS de longitud y latitud de sus móviles. Por tanto, cuando un tweet es producido, Twitter graba la información geográfica del usuario en ese momento temporal, junto a una variedad de metadatos, generando datos de grano fino que permiten la monitorización de procesos espacio-temporales a tiempo casi real (Frías-Martínez, Soto, y Frías-Martínez, 2012; Gutiérrez-Puebla, 2018). El uso de

datos de Twitter es muy interesante para analizar el uso de los espacios urbanos y los patrones geográficos de movilidad, siempre teniendo en cuenta que los usuarios de Twitter son una pequeña parte de toda la población, y que los datos georeferenciados son una porción de la cantidad total de mensajes útiles disponibles (Masala y Pallares-Barbera, 2016). También hay que tener en cuenta las cuestiones de privacidad que surgen al usar este tipo de datos.

Los datos masivos aportan información sobre multitud de procesos, como el crecimiento de las ciudades o la movilidad de la población (Gutiérrez-Puebla, 2018), con lo que surge la oportunidad de estudiar temas que no habían sido tratados o que habían sido dejados de lado ante la falta de información que ofrecían las fuentes tradicionales (Gutiérrez-Puebla y García-Palomares, 2016). En el campo del turismo, el desarrollo de las TIC está modificando métodos de investigación y herramientas de administración y marketing (Raun, Ahas y Tiru, 2016). Las redes sociales son usadas como fuentes para generar contenido como reseñas, opiniones, valoraciones, fotos, etc., con el que se puede estudiar patrones de movilidad, estimar ratios de visita de atracciones específicas, o identificar puntos calientes turísticos en ciudades (Battista *et al.*, 2018). Gracias a las nuevas fuentes de datos como Twitter es posible analizar el comportamiento de los turistas extranjeros, observar los sitios visitados por los turistas y medir la composición turística en el tiempo y en el espacio con mayor precisión (Raun, Ahas y Tiru, 2016).

El impacto de los eventos de masas es un tema que presenta una interesante oportunidad para su análisis a partir de las TIC en general, y de las redes sociales como Twitter en particular. Tradicionalmente se han empleado cuestionarios y encuestas sobre la población turística, cómo es el caso del estudio realizado por Knott *et al.* (2015) sobre cómo el Mundial de Fútbol de Sudáfrica del año 2010 afectó a la población local. Sin embargo, en los últimos años, ha aumentado el interés por detectar el impacto de eventos a partir de las nuevas fuentes de datos, ya que no solo permiten identificar puntos de interés, sino también pueden sacar datos sobre la ubicación o el evento (Li *et al.*, 2015). Así, por ejemplo, Versichele *et al.* (2012) han utilizado *Bluetooth* para estudiar el comportamiento y la densidad de la población en el festival de Gent; Xu y González (2017) emplearon varias fuentes de *Big Data* (*Airbnb*, *Waze*, y CDRs de telefonía móvil) para estimar el impacto de los megaeventos en el tráfico de la ciudad de Río de Janeiro durante los

Juegos Olímpicos de 2016; o Marine-Roig y Anton Clavé (2015) estudiaron la imagen de Barcelona como destino turístico mediante análisis de *blogs* de viajes y páginas webs. En España, *Vodafone* y *CARTO* estudiaron la *MTV Music Week Bizkaia* combinando una aplicación móvil desarrollada por ellos y sensores IoT ubicados en el evento².

En el caso concreto de los estudios del impacto de manifestaciones LGBT, hay una primera aproximación realizada por BBVA analizando la semana del desfile nacional del Orgullo de Madrid de los años 2011 y 2012. Usando los datos de transacciones realizadas con las tarjetas bancarias, se detectó un aumento de gasto del 24% en la ciudad respecto a semanas anteriores o posteriores al evento. También observaron una tendencia creciente del 9% en el consumo de 2012 respecto al año 2011, y una distribución espacial importante sobre todo en el distrito Centro³.

La detección de eventos mediante datos de Twitter ha sido recientemente un tema recurrente. Lee y Sumiya (2010) diseñaron un sistema de detección e identificación de eventos en tiempo real para identificar festivales locales en Japón. Weng *et al.* (2011) propusieron un método de detección de eventos basado en la agrupación a partir de palabras. Abdelhaq *et al.* (2013) elaboraron un sistema de detección de eventos a partir de la extracción de mensajes con determinadas palabras claves en un determinado ámbito temporal, su identificación en un área espacial, y la formación de grupos para su análisis. Popescu y Pennacchiotti (2011) clasificaron los eventos relacionados con famosos en diferentes categorías, estudiaron la reacción de las personas mediante sus tweets a esos eventos mediante análisis de sentimientos, y realizaron índices de intensidad y controversia. Kim *et al.* (2016) propusieron un sistema de identificación de eventos en Estados Unidos a partir de la captura de palabras claves que permitan la identificación de comunidades locales. Huang *et al.* (2018) se basaron en el análisis de texto de los mensajes mediante el algoritmo *Latent Dirichlet Allocation* para detectar eventos pequeños y determinar sus patrones espacio-temporales.

Sobre el estudio del contenido del texto en eventos ya localizados, Kirilenko y Stepchenkova (2017) han estudiado la distribución tanto en Rusia como a nivel global de los tweets sobre los Juegos Olímpicos de Sochi, han comparado los principales idiomas empleados, y han analizado el nivel a partir de un análisis de sentimientos. Para ello, han separado los mensajes en ruso, clasificados como la opinión de los habitan-

tes del país organizador, y los tweets en inglés, vistos cómo la opinión de los invitados o visitantes.

En cuanto al análisis de lo que pasa en eventos a partir de datos de Twitter, Steiger *et al.* (2015) trataron el impacto de las series mundiales de béisbol en la ciudad de Boston. Para ello buscaron patrones semánticos en los *tweets* a partir de su texto, teniendo en cuenta que la información semántica de un evento puede corresponder a eventos pasados o futuros, y que puede ser vaga y ambigua, dificultando la identificación de un evento. Como resultado, obtuvieron clústeres a partir de un número de tweets próximos tanto espaciotemporalmente como semánticamente. Zhou y Xu (2017) emplearon datos de Twitter para comparar la evolución espacial y temporal en la región Nueva York – Washington DC de un evento natural (tormenta torrencial) y un evento social (visita del Papa Francisco). En su metodología, eliminaron los mensajes no escritos en inglés y separaron los tweets restantes por intervalos de una hora en capas. Con estas capas, buscaron identificar eventos locales y regionales, y emplearon las palabras claves para poder hallar los espectros temporal y espacial de los eventos, y realizar análisis de sentimientos a partir del contenido de los mensajes.

CASO DE ESTUDIO, DATOS, Y METODOLOGÍA

Caso de estudio

El área de estudio es la ciudad central de Madrid, con una población estimada en casi 1 millón de habitantes en el año 2017. Esta zona está comprendida por los distritos dentro de la carretera de circunvalación M-30 (Figura 1). Se trata del área de la ciudad de mayor concentración de actividad y servicios. En este espacio, se ha trabajado a niveles de agregación espacial en barrios y en secciones censales.

El barrio de Chueca, ubicado al norte de la Gran Vía, es de especial interés para este trabajo debido a su papel como área central de actividades de la *World Pride* (destacando las plazas de Chueca, del Rey, y de Pedro Zerolo). Este barrio, de ambiente multicultural, es el epicentro de la ciudad para la comunidad LGBT desde los años 1980, década en la que personas pertenecientes a la comunidad LGBT madrileña se hicieron con muchas viviendas y locales, floreciendo las discotecas, pubs, bares, y tiendas orientadas a dicho colectivo. Desde 1986 se celebra anualmente en el barrio de Chueca las fiestas de

celebración del Orgullo Gay. En la década de los 90 se desarrolla la manifestación LGBT como elemento reivindicativo unido a celebraciones lúdicas⁴. Los festivales se organizan el fin de semana posterior al Día del Orgullo LGBT, el 28 de junio.

La *World Pride* es el mayor evento del colectivo LGTB del mundo. Se celebra a nivel mundial desde el año 2000. Su quinta edición se ha celebrado en Madrid en el año 2017, desde el 23 de junio hasta el 2 de julio. El programa más destacado fue la manifestación celebrada el 1 de julio en los Paseos del Prado y de Recoletos. El evento contó además con diversas actividades, como conciertos o carreras. El Ayuntamiento de Madrid ha calculado una asistencia de 2.200.000 personas y un beneficio económico para la ciudad de 115 millones de euros⁵.

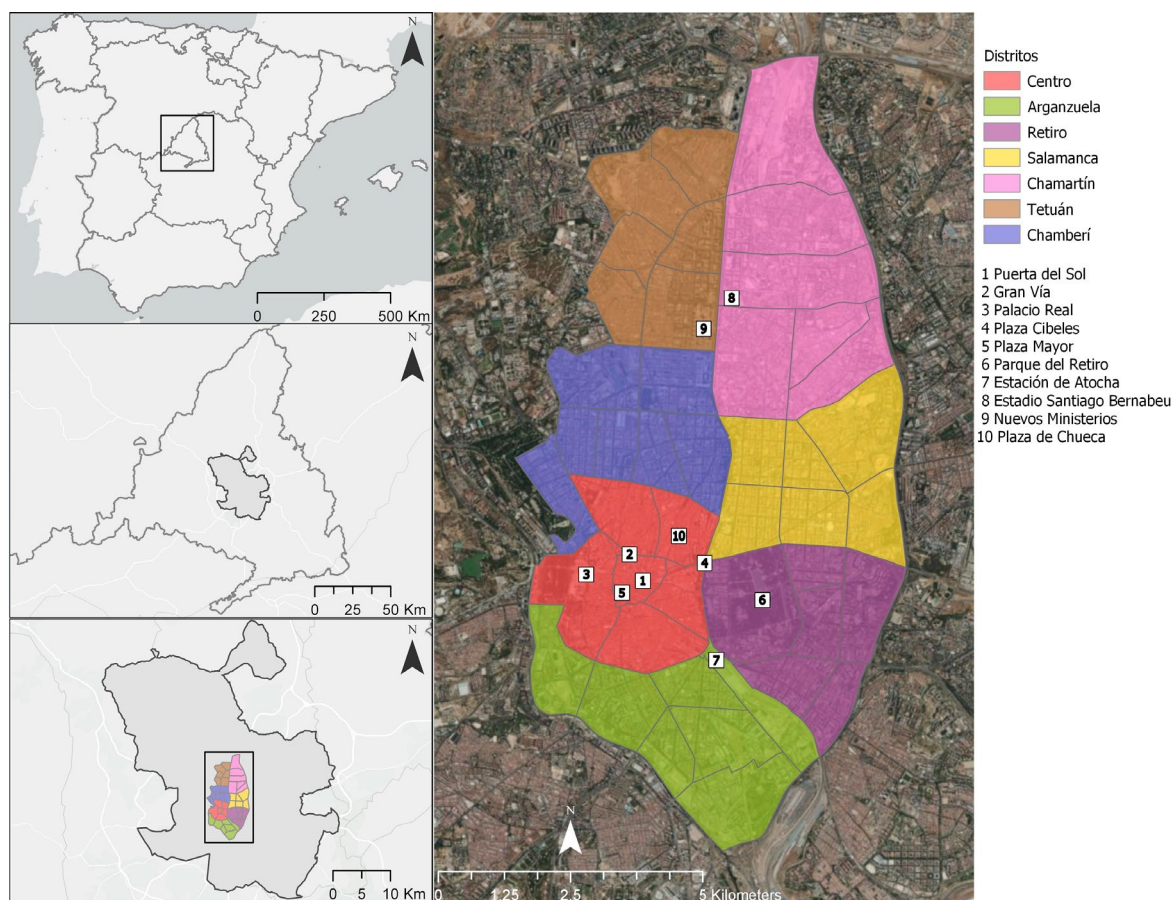
El marco temporal del trabajo ha sido el periodo de tiempo en el que se celebró la *World Pride*. Para

comparar los comportamientos del evento con los de una semana de actividad habitual en la ciudad, se ha seleccionado la semana del 12 de junio al 18 de junio.

Datos

La base de datos empleada consiste en mensajes de Twitter, georreferenciados y ubicados en la almendra central de Madrid en el periodo comprendido entre el lunes 12 de junio y el domingo 16 de julio. Dicha base de datos inicial consta de 48.175 tweets de 14.353 usuarios. Cada tweet cuenta con información relativa al identificador de usuario, el *nick* del usuario, la latitud y longitud, fecha y hora, idioma y los *hashtags* que incluye el tweet. Esta información ha sido enriquecida con la referencia del distrito, barrio y sección censal desde donde fue publicado cada tweet.

FIGURA 1
DISTRITOS Y BARRIOS DE LA ALMENDRA CENTRAL DE MADRID



Fuente: Elaboración propia.

Recogida, limpieza, preparación y enriquecimiento de las bases de datos de Twitter

El primer paso ha consistido en la recopilación de *tweets* mediante un código *Python* que ha descargado los datos en *streaming* de la API de Twitter. Una vez recopilados en una base de datos *noSQL*, estos mensajes han sido incorporados al Sistema de Información Geográfica *ArcGIS Pro* y localizados según sus coordenadas geográficas. En el periodo de la *World Pride*, en la almendra central, se publicaron 18.439 *tweets* compartidos por 7.406 usuarios. Para comparar los resultados de la *World Pride* con la actividad habitual de una semana próxima, se han tomado los *tweets* escritos en la semana del 12 al 18 de junio, donde se generaron un total de 8.896 mensajes publicados por 4.053 usuarios.

Finalmente, se han identificado y extraído las identificaciones de los 7.406 usuarios que se detectaron en la *World Pride* para obtener los 3.200 últimos *tweets* de cada uno de ellos, e intentar obtener así más mensajes que no se hubiesen captado en *streaming* (Huang & Wong, 2016). Además, esta búsqueda permitió crear una nueva base de datos de *tweets* a escala global donde identificar sus procedencias. Como resultado, se ha obtenido una tercera base de datos formada por 1.215.994 *tweets* con información enriquecida del lugar desde donde fueron publicados a partir de la unión espacial en SIG con capas de países y de provincias de España. Al haber enriquecido la base con el nombre del país desde donde fue publicado cada *tweet*, se pudo sacar el número de *tweets* que había escrito cada usuario desde los distintos países encontrados a partir de Twitter, y se estableció como país de origen aquel con un mayor número de mensajes. Para los usuarios cuyo país de origen era España se repitió el proceso a nivel provincial, con el fin de conocer la provincia de origen.

Análisis de los datos

El análisis temporal de los datos se efectuó sobre la distribución diaria del número de *tweets* y usuarios agregados a partir de sus identificadores durante el periodo de estudio. Respecto a la identificación de idiomas, se ha seleccionado por usuario el idioma con mayor número de mensajes.

La impronta espacial del evento se analizó a nivel de barrios y de secciones censales mediante el conteo de usuarios localizados por cada unidad y el cálculo del

incremento del porcentaje respecto al número de usuarios encontrados en el mismo polígono en la base de datos de la semana habitual. Estos resultados se complementaron con un análisis de autocorrelación espacial (Moran I) para evaluar el grado de agrupación de usuarios en el evento, y con análisis LISA (*Local Indicators of Spatial Association*) para la visualización de concentraciones de valores altos, bajos, y *outliers* espaciales en la ciudad durante la *World Pride* (Anselin, 1995).

La distribución de la población en la ciudad varía a diferentes horas del día y puede ser analizada a partir de la huella digital de la población a cada hora del día en cada punto de la ciudad (Moya *et al.*, 2017). El análisis espaciotemporal del caso se ha realizado a nivel de secciones censales, calculando el número de usuarios que ha *twitteado* en cada sección por franjas horarias. La visualización se ha realizado por mapas de símbolos proporcionales a partir de dos animaciones que muestran por distintos momentos del día los cambios en el número de usuarios detectados en el centro de Madrid durante la semana del evento y una semana habitual. Cada franja es de 6 horas, dando como resultado visualizaciones en cuatro momentos del día (mañana, mediodía, tarde, y noche).

RESULTADOS

Análisis temporal

Los primeros resultados mostraron un mayor volumen de mensajes y de usuarios en la semana de la *World Pride* que el periodo comprendido por las dos semanas previas o las dos semanas posteriores al evento. Se puede observar un paulatino crecimiento conforme avanza el festival (Figura 2a). El momento cumbre de este crecimiento se da el fin de semana del 1 y 2 de julio, día de la clausura del evento, con un incremento del 88% de *tweets* publicados y del 71% de usuarios respecto al promedio del periodo analizado. Una vez terminada la *World Pride*, se aprecia un descenso de la actividad en Twitter, indicativo de que los visitantes se van de la ciudad tras haber terminado el evento, pero también de que los propios madrileños dejaron un mayor número de mensajes en la fiesta y su actividad es menor en una semana habitual.

Si reducimos el ámbito espacial al barrio de Chueca, se puede observar con mayor claridad un incremento de la actividad de Twitter en la franja con las fechas de la *World Pride* (Figura 2b). En este caso hay un aumento brusco de la actividad durante todo el even-

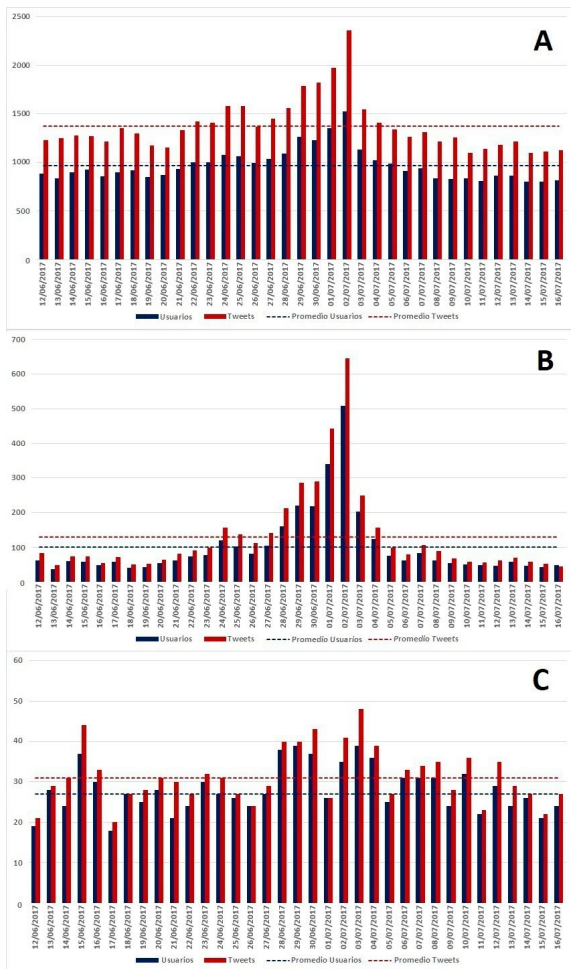
to en comparación con el bajo volumen de tweets en los días anteriores o posteriores. Así, en los días 1 y 2 de julio el incremento respecto al promedio sube al 706% en cuanto a tweets, y al 695% de usuarios. Además, el lunes 3 de julio (el día después del final de la *World Pride*) fue el día con mayor número de mensajes enviados en el barrio del Aeropuerto de Barajas (Figura 2c), con un aumento del 77% de tweets y del 54% de usuarios respecto a la media.

Si analizamos la actividad por número de usuarios según la distribución horaria de una semana habitual y la comparamos con la semana de la *World Pride*, en la almendra central se puede observar como la curva de actividad tiene a grandes rasgos el mismo perfil, con la diferencia de que en la *World Pride* hubo

más usuarios (Figura 3a). En ambas situaciones, los momentos de mayor actividad se dieron durante la tarde (de 12 a 16) y la noche (de 20 a 23), situándose el pico a las 22 de la noche. A esta hora, el número de usuarios durante el festival tuvo un 28% de mayor actividad respecto a la semana habitual.

Sin embargo, si analizamos la actividad por horas en el barrio de Chueca, las curvas entre una semana habitual y el *World Pride* son muy diferentes. En la recta de la *World Pride* destacan dos grandes picos de actividad: de 13 a 15 de la tarde y de 21 a 23 de la noche, causando una gran diferencia de actividad respecto a una semana habitual donde se ve un patrón más lineal y con pocos crecimientos. Durante la hora cumbre, las 22 horas, el incremento fue del 350% de usuarios respecto a la semana habitual (Figura 3b).

FIGURA 2
VOLUMEN DE ACTIVIDAD POR DÍAS EN LA ALMENDRA CENTRAL (A), BARRIO DE CHUECA (B), Y AEROPUERTO (C)

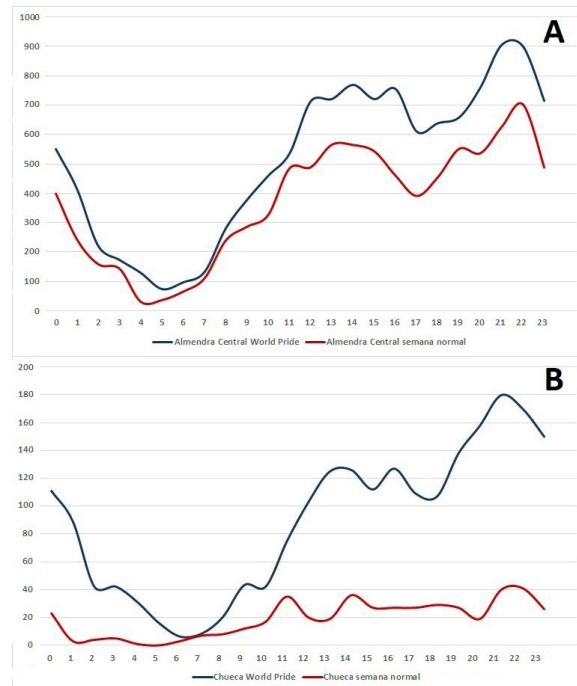


Fuente: Elaboración propia a partir de datos de Twitter.

Carácter multicultural del evento: las lenguas durante la *World Pride*

A partir de las bases de datos de tweets se ha podido extraer el idioma principal de cada usuario.

FIGURA 3
COMPARACIÓN DE NÚMERO DE TWEETS POR HORAS EN LA ALMENDRA CENTRAL (A) Y EN EL BARRIO DE CHUECA (B)



Fuente: Elaboración propia a partir de datos de Twitter.

Mientras en una semana habitual en la ciudad central aparecen 26 idiomas entre los tweets recogidos, durante la *World Pride* esta cifra aumenta hasta 32 idiomas diferentes. El cambio drástico sucede en el barrio de Chueca, donde los 9 idiomas utilizados en la semana habitual contrastan con las 21 lenguas registradas durante la *World Pride*. Los idiomas más encontrados han sido el castellano y el inglés, seguidos por el portugués, el francés, y el italiano (Tabla 1).

La distribución de idiomas en la almendra central durante la *World Pride* y durante una semana habitual permite apreciar el aumento del porcentaje de lenguas extranjeras empleadas durante la *World Pride*, destacando el idioma inglés (25% de presencia durante el evento frente al 20% durante una semana habitual). Este fenómeno es más visible en el barrio Chueca donde en la *World Pride* hay un 36% de mensajes escritos en inglés, frente a un 17% en una semana habitual. El caso contrario sucede con el idioma español, con registros muy similares en las dos semanas de estudio en la almendra central (en torno al 70%), pero con un porcentaje ligeramente menor en la *World Pride*. De nuevo, esta situación se visualiza mejor en el barrio de Chueca; durante la *World Pride* (un 56%) el uso del español es mucho menor respecto a su uso en una semana habitual (un 71%).

Lugares de procedencia de los visitantes

En la procedencia de los visitantes de la *World Pride* destaca la propia Comunidad de Madrid como principal lugar de origen, coincidiendo con la información del propio Ayuntamiento de Madrid⁶. En cuanto a otras provincias, en segundo lugar, se sitúa Barcelona, seguida de las provincias más pobladas del litoral

mediterráneo (Valencia, Alicante), Andalucía (Sevilla, Cádiz, Málaga), otras cercanas como Toledo, y las Islas Canarias. (Figura 4).

A nivel internacional los principales países de origen son de América, destacando Estados Unidos, México, Brasil, Colombia y Argentina. La segunda área de visitantes es Europa occidental, donde destaca Reino Unido, seguida de Italia, Francia, y Portugal (Figura 3). Estos resultados coinciden con los obtenidos por el buscador de viajes *GoEuro*. Esta empresa permite comprar o reservar viajes a partir de la selección de un lugar de destino y una fecha de desplazamiento. Según sus datos, los países con mayor número de visitantes durante la *World Pride* fueron también los países de América y Europa occidental (destacando Argentina, Colombia, Estados Unidos, Francia y Reino Unido). En cuanto a rutas nacionales destacaron las conexiones con Barcelona, Valencia, y Sevilla (las provincias con mayor registro de usuarios locales según la Figura 5), y referente a los trayectos internacionales los más importantes fueron con las ciudades de París, Roma, y Lisboa⁷.

La impronta espacial del evento

Las zonas con mayor número de usuarios localizados durante la celebración de la *World Pride* coinciden con los principales lugares que acogieron el evento. A nivel de barrios se pueden apreciar cómo los barrios con un mayor número de usuarios detectados durante el evento fueron Sol (principal centro de la ciudad), Chueca y Atocha (Figura 6a).

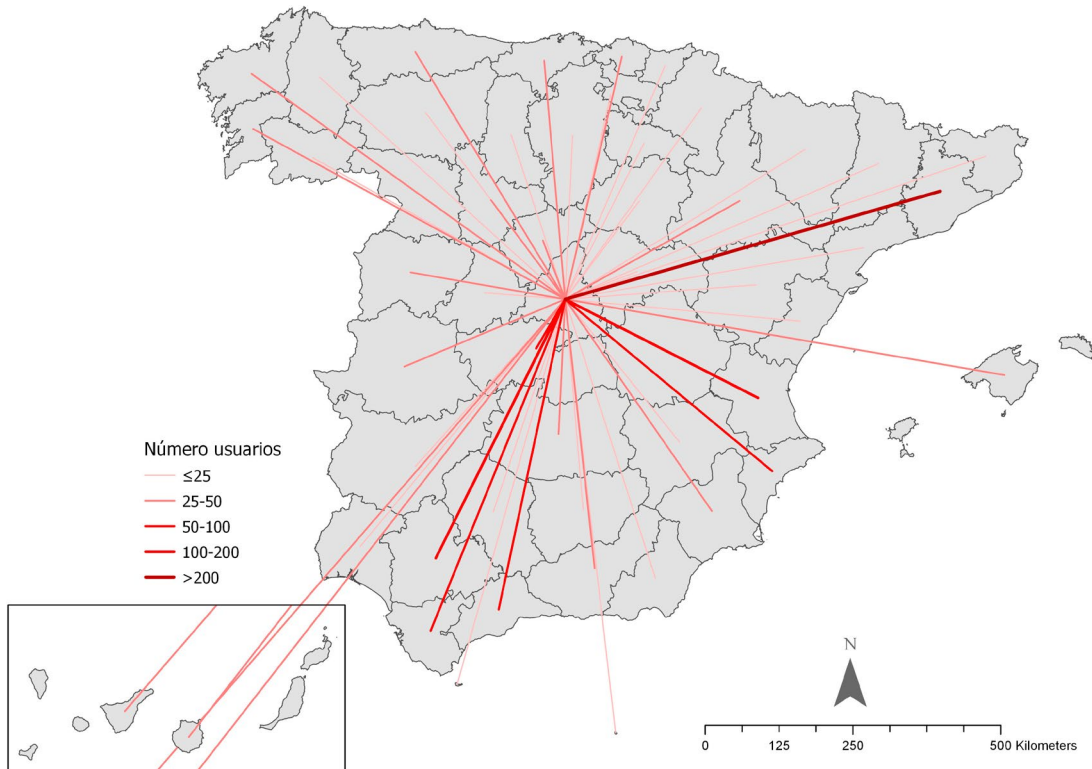
Comparando el porcentaje de crecimiento de usuarios respecto al número de usuarios que se de-

TABLA 1
NÚMERO DE USUARIOS SEGÚN IDIOMA CONFIGURADO EN TWITTER

Idioma	Almendra Central World Pride	%	Almendra Central semana normal	%	Chueca World Pride	%	Chueca semana normal	%
Castellano	5073,00	68,50	2856,00	70,47	882,00	55,82	206,00	71,28
Inglés	1860,00	25,11	836,00	20,63	574,00	36,33	50,00	17,30
Portugués	149,00	2,01	83,00	2,05	18,00	1,14	5,00	1,73
Francés	60,00	0,81	29,00	0,72	9,00	0,57	3,00	1,04
Italiano	50,00	0,68	16,00	0,39	9,00	0,57	0,00	0,00
Otros	214,00	2,89	233,00	5,75	88,00	5,57	25,00	8,65
Totales	7406,00	100,00	4053,00	100,00	1580,00	100,00	289,00	100,00

Fuente: Elaboración propia a partir de datos de Twitter.

FIGURA 4
USUARIOS DETECTADOS EN TWITTER DURANTE LA *WORLD PRIDE* SEGÚN PROVINCIA DE PROCEDENCIA



Fuente: Elaboración propia a partir de datos de Twitter.

FIGURA 5
USUARIOS DETECTADOS EN TWITTER DURANTE LA *WORLD PRIDE* SEGÚN PAÍS DE PROCEDENCIA



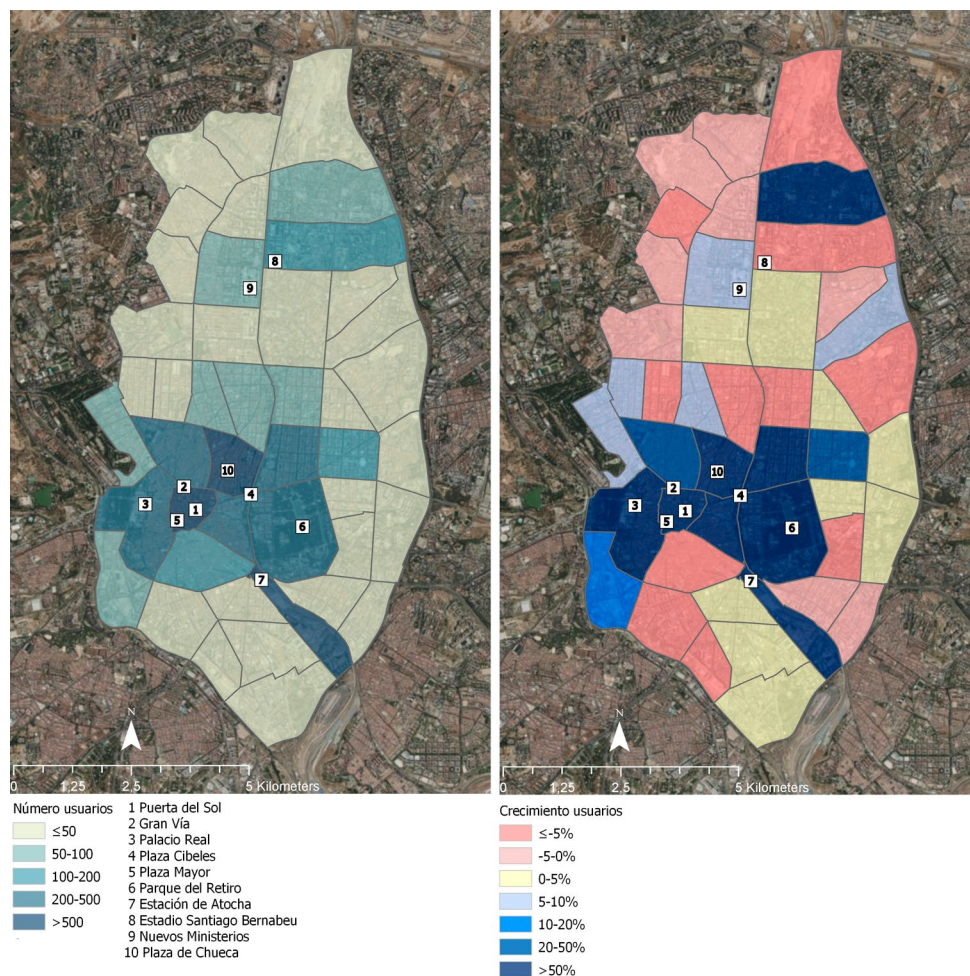
Fuente: Elaboración propia a partir de datos de Twitter.

teció en la semana habitual, se puede apreciar que estos tres barrios obtuvieron un aumento destacable de la actividad respecto a una semana habitual. Destaca también el aumento de actividad en espacios de especial interés turístico, como el barrio central de Cortes (barrio entre Sol, Justicia, y el Paseo del Arte), y los barrios de Jerónimos (ubicación del Parque del Retiro), Recoletos (por donde transcurre el paseo de Recoletos, vía donde transcurrió el desfile del Orgullo del sábado 1 de julio), y Nueva España (barrio cercano a la estación de trenes de Chamartín, a la zona de las Cuatro Torres, y al estadio de fútbol Santiago Bernabéu). También se detecta un crecimiento notable en el barrio central de Universidad (o Malasaña), Goya (donde se encuentra el Palacio de Deportes de la Comunidad

de Madrid), e Imperial (barrio que transcurre paralelamente al camino verde Madrid Río). Por otro lado, se puede apreciar un descenso de actividad respecto a una semana habitual en el barrio central de Embajadores (o barrio de Lavapiés), y los barrios de los distritos periféricos del norte, el este y el sur de la almendra central de Madrid (Figura 6b).

Si reducimos la escala de análisis y trabajamos con secciones censales, se puede apreciar con mayor detalle como los usuarios se concentran en el eje de la Gran Vía y en las plazas de Chueca. Así mismo, también se observa una importante presencia de usuarios en los principales lugares turísticos del área de estudio (triángulo Sol-Callao-Opera, Palacio Real, estadios Santiago Bernabéu y Vicente Calderón, Palacio de los

FIGURA 6
NÚMERO USUARIOS (IZQUIERDA) Y CAMBIO PORCENTUAL (DERECHA) RESPECTO A SEMANA HABITUAL EN LA ALMENDRA CENTRAL POR BARRIOS DURANTE LA *WORLD PRIDE*



Fuente: Elaboración propia a partir de datos de Twitter.

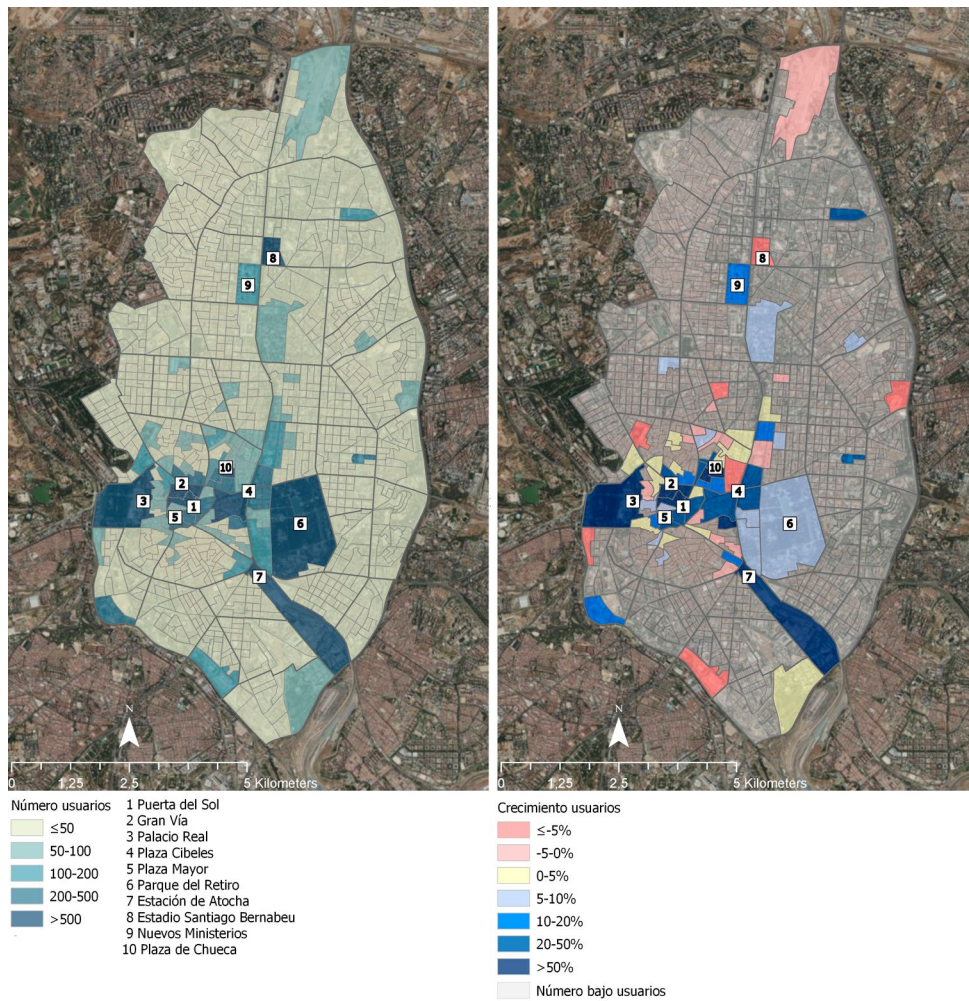
Deportes de la Comunidad de Madrid, Paseo de Recoletos y Parque del Retiro) (Figura 7a).

En cuanto al porcentaje de crecimiento respecto a la semana habitual, para una mejor visualización de los principales cambios, se ha optado por ignorar las secciones con menos de 10 usuarios. De esta forma, se puede apreciar de nuevo el fuerte crecimiento en el barrio de Chueca y otros puntos centrales como la Puerta de Sol, la Gran Vía, o el Palacio Real (los principales puntos turísticos del casco histórico), además del Palacio de los Deportes de la Comunidad de Madrid, el área de Nuevos Ministerios, o la estación de Atocha. A su vez, destaca un porcentaje negativo de crecimiento en el estadio Santiago Bernabéu pese a

ser uno de los lugares con mayor número de usuarios, y se puede apreciar un fuerte decrecimiento en otras zonas como la plaza de toros de Las Ventas, puntos de Madrid Río, el Palacio de Conde-Duque, o la estación de trenes de Chamartín. Por otra parte, el parque del Retiro o el Museo Arqueológico no presentan crecimiento ni decrecimiento respecto a una semana de actividad habitual. Esto se debe a que mientras en una semana habitual son puntos de atracción de visitantes, durante el evento pierden interés (Figura 7b).

El Índice de Moran revela durante la *World Pride* (valor z 3,37 y valor p 0,0007) indica una fuerte clus-terización debido a la presencia causada durante el evento. Esta misma situación ocurre durante una se-

FIGURA 7
NÚMERO USUARIOS (IZQUIERDA) Y CAMBIO PORCENTUAL (DERECHA) RESPECTO A SEMANA HABITUAL EN LA ALMENDRA CENTRAL POR SECCIONES CENSALES DURANTE LA *WORLD PRIDE*



Fuente: Elaboración propia.

mana habitual debido a la fuerte presencia turista que suele tener la ciudad de Madrid, pero a una menor intensidad respecto a la *World Pride*. En este caso (valor z 2,2 y valor p 0,027), se aprecia una menor clusterización respecto a la semana del evento.

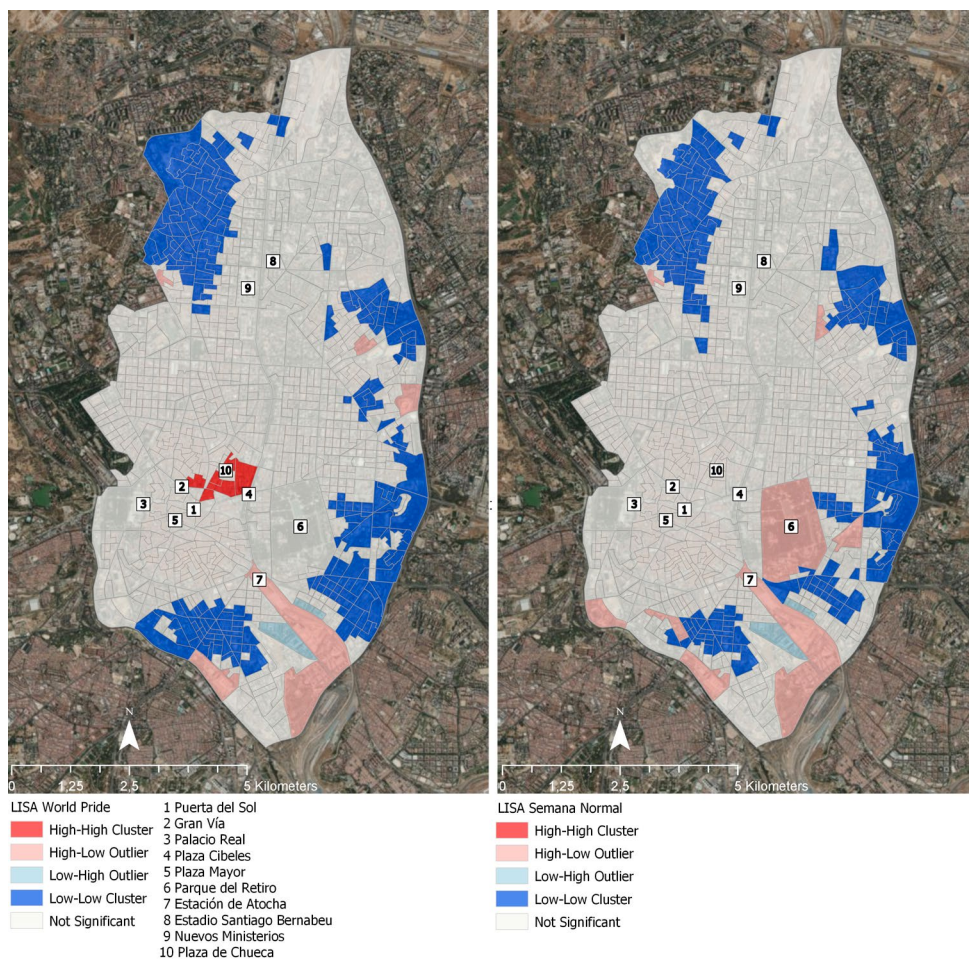
El Índice LISA permite cartografiar la distribución de los clústeres. Tanto en la semana del festival como en la semana habitual hay un valor alto rodeado de áreas de valor bajo en la estación de Atocha, y clústeres de poca actividad el distrito de Tetuán y en barrios del este y sur de la almendra central. Sin embargo, durante la *World Pride* se ha visto que se genera un clúster de valor alto rodeado de áreas de mucha actividad en el barrio de Chueca, mientras que el valor alto de Retiro en la semana normal desaparece (Figura 8).

La evolución espaciotemporal del evento

La actividad detectada a través de Twitter varía dependiendo de la hora, el día y el lugar. Estudiando la actividad detectada durante una semana habitual, se puede observar como el mayor foco de usuarios corresponde con la estación de trenes de Atocha, y luego una actividad centrada de manera uniforme en el casco histórico, el parque del Retiro, y en secciones de los distritos de Salamanca y Chamberí. En cambio, se puede apreciar una mayor actividad a rasgos generales (un número mayor de círculos, y también círculos de mayor tamaño en un mayor número de franjas horarias) en estas zonas durante la *World Pride*.

Además, se observa durante el evento un mayor número de usuarios en las secciones correspondien-

FIGURA 8
CLÚSTERES DE ASOCIACIÓN ESPACIAL LOCAL EN LA ALMENDRA CENTRAL POR SECCIONES CENSALES DURANTE LA *WORLD PRIDE* (IZQUIERDA) Y DURANTE UNA SEMANA HABITUAL (DERECHA)



Fuente: Elaboración propia a partir de datos de Twitter.

tes con el casco histórico, y se puede apreciar en determinados momentos (como la tarde-noche del sábado o el mediodía del domingo) un mayor número de usuarios en el barrio de Chueca que en la estación de Atocha. Al contrario, el estadio Santiago Bernabéu es el único lugar con círculos de mayor tamaño durante una semana habitual. La Figura 9 incluye un mapa donde se pueden comparar los momentos más importantes del evento (las franjas horarias en fin de semana) con la actividad en esas mismas franjas en una semana normal. El Archivo Multimedia 1 compara de forma animada el número de usuarios detectados en cada franja horaria a lo largo de la semana de la *World Pride* con la actividad detectada en una semana habitual.

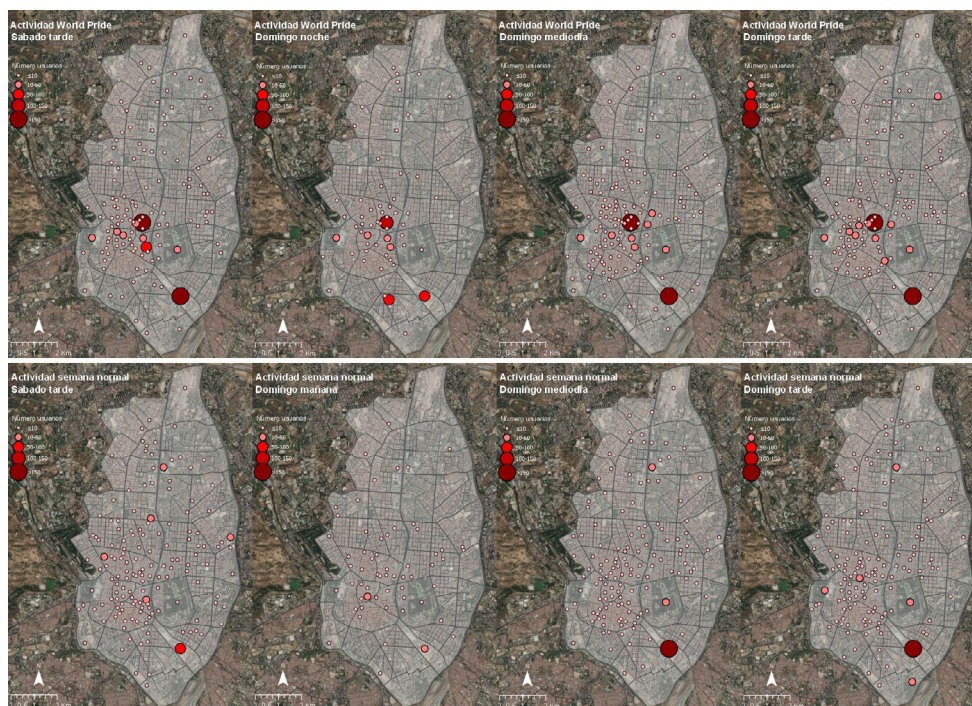
CONCLUSIONES

Los mega eventos destacan por generar una gran actividad en internet, por lo que redes sociales como Twitter pueden convertirse en fuentes válidas para analizar su impacto en una ciudad y ayudar a su organización. Este artículo ha trabajado con datos de Twitter para obtener información sobre las caracte-

rísticas y el impacto que ha tenido la *World Pride* sobre el centro de la ciudad de Madrid respecto a la situación en una semana de actividad corriente. Al tener acceso solo a una pequeña parte de los datos, se ha trabajado con una muestra pequeña. Sin embargo, se ha podido comprobar que los datos geolocalizados de Twitter son una fuente eficaz para estudiar y analizar de forma libre y gratuita los patrones espaciales y temporales del evento con un alto detalle. Es interesante destacar que los resultados de los datos obtenidos son muy similares a datos de otras fuentes oficiales o de empresas privadas (como es el caso de la procedencia de turistas).

Los resultados obtenidos han mostrado un destacable aumento de la actividad turista en la *World Pride* respecto a semanas de actividad habitual en la zona central de la ciudad. Se han podido hallar picos de actividad que coincidían con los días de algunas de las actividades más importantes del evento (como es el caso de la manifestación o la clausura del festival). La impronta espacial del evento se ha notado en la formación de determinados puntos calientes con alto número de usuarios y elevado porcentaje de actividad, principalmente en el barrio de Chueca, pero también en menor medida en otros puntos de la ciu-

FIGURA 9
NÚMERO DE USUARIOS POR SECCIONES CENSALES EN LA ALMENDRA CENTRAL EN FRANJAS DEL FIN DE SEMANA



Fuente: Elaboración propia.

dad, destacando las principales zonas turísticas del casco histórico de la ciudad. Igualmente se puede apreciar disminución de la actividad en puntos importantes de la ciudad, pero alejados del centro (cómo el estadio Santiago Bernabéu). Analizando las variaciones espaciotemporales, mientras en una semana habitual el principal foco de usuarios se concentra en la estación de Atocha, en la semana del evento diversas secciones del casco histórico, y más concretamente del barrio de Chueca, llegan a superar en número de usuarios a la estación de Atocha, sobre todo en momentos de la tarde-noche.

Una de las innovaciones realizadas en este trabajo es el diseño de una metodología para identificar usuarios que han viajado al festival observando si han publicado tweets en Madrid fuera del evento, y sus lugares de procedencia a partir de la descarga de los 3.200 últimos mensajes de casa usuario e identificación de la provincia o país con mayor número de mensajes. Se ha observado que el principal

lugar de procedencia de los visitantes en España es la propia Comunidad de Madrid seguida de las provincias de mayor densidad de población (Barcelona, Sevilla, Málaga, Cádiz, Valencia, Alicante), y fuera de España, las regiones con un mayor número de visitantes fueron Estados Unidos, América Latina, y Europa occidental.

Finalmente, pese a la demostración de Twitter como fuente veraz para aplicar en este tipo de estudios, hay que resaltar algunas limitaciones de cara a futuras líneas de investigación, como la cantidad limitada de mensajes con la que se ha trabajado, o el sesgo de la red social. Disponer de muestras mucho mayores de datos puede permitir investigaciones en otros campos como la distribución de visitantes por sexo o edad, el análisis de sentimientos del evento a partir de las palabras encontradas en los mensajes, o la evaluación de los *hashtags* más comunes durante el evento. El uso de otras fuentes de datos masivas cómo teléfonos, tarjetas bancarias o tarjetas de trans-

ARCHIVO MULTIMEDIA 1

NÚMERO DE USUARIOS POR SECCIONES CENSALES EN LA ALMENDRA CENTRAL A LO LARGO DE LA SEMANA



Fuente: Elaboración propia.

porte podrían mitigar los sesgos de Twitter, pero hay que tener en cuenta que son fuentes menos accesibles. La privacidad y seguridad de los datos son retos a tener en cuenta que se han tratado en este artículo a través del uso de identificadores numéricos anónimos de usuarios, la sumarización de datos, y el uso de la escala a nivel de provincia y país para tratar con el lugar de procedencia de los visitantes.

Aprovechando el auge de las fuentes basadas en *Big Data* para la administración y gestión de las ciu-

dades por parte de organismos públicos y privados, este trabajo puede ser de gran interés ya que muestra una metodología para hallar los principales puntos de impacto en una ciudad, las principales horas del día de actividad, o realizar un conteo de visitantes y estudiar su procedencia. Al contar con datos masivos, gratuitos, y de fácil tratamiento, estos organismos pueden convertirlos en información con la que tomar medidas para la organización, aprovechamiento y mejora de futuros eventos de masas en la ciudad.

NOTAS

- 1 <https://www.sevilla.org/ayuntamiento/alcaldia/comunicacion/noticias/el-estudio-smart-city-sevilla-contabiliza-medio-millon-de-desplazamientos-a-pie-al-dia-en-el-entorno-de-la-carrera-oficial-durante-el-arranque-de-la-semana-santa-2018>
- 2 http://www.saladeprensa.vodafone.es/c/notas-prensa/np_soluciones_mtvmusicweek/
- 3 https://issuu.com/cibbva/docs/big_data_english
- 4 <http://www.worldpridemadrid2017.com/home/acercade>
- 5 <https://diario.madrid.es/blog/notas-de-prensa/exito-del-operativo-de-seguridad-movilidad-y-limpieza-del-world-pride-madrid-2017/>
- 6 <https://diario.madrid.es/blog/notas-de-prensa/el-gasto-en-el-centro-de-la-ciudad-aumento-un-15-durante-el-world-pride/>
- 7 https://www.hosteltur.com/comunidad/nota/019268_argentinos-y-colombianos-los-mas-orgullosos-de-viajar-al-worldpride-2017.html

REFERENCIAS BIBLIOGRÁFICAS

- Abdelhaq, H., Sengstock, C., y Gertz, M. (2013). EventTweet: Online localized event detection from Twitter. *Proceedings of the VLDB Endowment*, 6(12), pp. 1326–1329. <https://doi.org/10.14778/2536274.2536307>
- Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2), pp. 93–115.
- Bar-Gera, H. (2007). Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel. *Transportation Research Part C: Emerging Technologies*, 15(6), pp. 380–391. <http://dx.doi.org/10.1016/j.trc.2007.06.003>
- Batista e Silva, F., Marin Herrera, M., Rosina, K., Ribeiro Barranco, R., Freire, S., y Schiavina, M. (2018). Analyzing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources. *Tourism Management*, 68, pp. 101–115. <http://dx.doi.org/10.1016/j.tourman.2018.02.020>
- Cáceres, N. (2012). Traffic Flow Estimation Models Using Cellular Phone Data. *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–12. <http://dx.doi.org/10.1109/TITS.2012.2189006>
- Cáceres, N., Wideberg, J.P, y Benítez, F.G. (2007). Deriving origin–destination data from a mobile phone network. *IET Intelligent Transport Systems*, 1, pp. 15–26. <http://dx.doi.org/10.1049/iet-its:20060020>
- Frías-Martínez, V., Soto, V., Hohwald, H., y Frías-Martínez, E. (2012). Characterizing urban landscapes using geolocated tweets. *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom)*, pp. 239–248. IEEE. <http://dx.doi.org/10.1109/SocialCom-PASSAT.2012.19>
- Gutiérrez-Puebla, J., y García-Palomares, J.C. (2016). Big (Geo) Data en Ciencias Sociales: Retos y Oportunidades. *Revista de Estudios Andaluces*, 33(331), pp. 1–23. <http://dx.doi.org/10.12795/rea.2016.i33.01>

- Gutiérrez-Puebla, J. (2018). Big Data y nuevas geografías: la huella digital de las actividades humanas. *Documents d'Anàlisi Geogràfica*, 64, pp. 195–217. <https://doi.org/10.5565/rev/dag.526>
- Huang, Q., y Wong, D. (2016). Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? *International Journal of Geographical Information Science*, 30(9), pp. 1873–1898. <https://doi.org/10.1080/13658816.2016.1145225>
- Huang, Y., Li, Y., y Shan, J. (2018). Spatial-Temporal Event Detection from Geo-Tagged Tweets. *ISPRS International Journal of Geo-Information*, 7(4). <https://doi.org/10.3390/ijgi7040150>
- Kim, K-S., Kojima, I., y Ogawa, H. (2016). Discovery of local topics by using latent spatio-temporal relationships in geo-social media. *International Journal of Geographical Information Science*, 30(9), pp. 1899–1922. <https://doi.org/10.1080/13658816.2016.1146956>
- Kirilenko, A. P., y Stepchenkova, S. O. (2017). Sochi 2014 Olympics on Twitter: Perspectives of hosts and guests. *Tourism Management*, 63, pp. 54–65. <https://doi.org/10.1016/j.tourman.2017.06.007>
- Kitchin, R. (2013). Big Data and human geography: opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3), pp. 262–267. <https://doi.org/10.1177/2043820613513388>
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), pp. 1–12. <https://doi.org/10.1177/2053951714528481>
- Knott, B., Swart, K., y Visser, S. (2015). The impact of sport mega-events on the quality of life for host city residents: reflections on the 2010 FIFA World Cup. *African Journal of Hospitality, Tourism and Leisure*, 4, pp. 1–16.
- Lansley, G., Smith, M. De, Goodchild, M., y Longley, P. (2018). Big Data and Geospatial Analysis. *Geospatial Analysis 6th Edition*, pp. 547–570. Edimburgo, Reino Unido: The Winchelsea Press.
- Lee, R., y Sumiya, K. (2010). Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks - LBSN '10*, pp. 1–10. New York, ACM Press. <https://doi.org/10.1145/1867699.1867701>
- Leszczynski, A. y Crampton, J. (2016). Introduction: spatial big data and everyday life. *Big Data & Society*, 3(2), pp. 1–6. <https://doi.org/10.1177/2053951716661366>
- Leung, D., Law, R., van Hoof, H., y Buhalis, D. (2013). Social Media in Tourism and Hospitality: A Literature Review. *Journal of Travel and Tourism Marketing*, 30(1–2), pp. 3–22. <https://doi.org/10.1080/10548408.2013.750919>
- Li, H., Ji, H., y Zhao, L. (2015). Social Event Extraction: Task, Challenges and Techniques. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 - ASONAM '15*, pp. 526–532, New York, New York, ACM Press.
- Liu, H., Ge, Y., Zheng, Q., Lin, R., y Li, H. (2018). Detecting global and local topics via mining twitter data. *Neurocomputing*, 273, pp. 120–132. <https://doi.org/10.1145/2808797.2809413>
- Marine-Roig, E., y Anton Clavé, S. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing and Management*, 4(3), pp. 162–172. <https://doi.org/10.1016/j.jdmm.2015.06.004>
- Masala, E., y Pallares-Barbera, M. (2016). When Internet became Geography. Spatial patterns on urban open spaces through the analysis of user-generated data in Barcelona. *ICiTy – Enhancing Places through Technology. MALTA: April, 18-19. The mid-term conference for COST Action TU1306 CyberParks: Fostering Knowledge about the Relationship between Information and Communication Technologies and Public Spaces*. <https://www.um.edu.mt/ev>, Valletta, Malta.
- Miller, H. J. (2010). The data avalanche is here. Shouldn't we be digging? *Journal of Regional Science*, 50(1), pp. 181–201. <https://doi.org/10.1111/j.1467-9787.2009.00641.x>
- Miller, H.J., y Goodchild, M.F. (2014). Data-driven geography. *GeoJournal*, 80(4), pp. 449–461. <https://doi.org/10.1007/s10708-014-9602-6>
- Moya-Gómez, B.; Salas-Olmedo, M. H.; García-Palmares, J. C. y Gutiérrez-Puebla, J. (2017). Dynamic accessibility using Big Data: The role of the changing conditions of network congestion and destination attractiveness. *Networks and Spatial Economics*, 18(2), pp. 273–290. <https://doi.org/10.1007/s11067-017-9348-z>

- Osorio, J., y García-Palomares, J.C. (2017). Nuevas fuentes y retos para el estudio de la movilidad urbana. *Cuadernos Geográficos*, 56(3), pp. 247–267.
- Osorio, J., y García-Palomares, J.C. (2019). Social media and urban mobility: using Twitter to calculate home-work travel matrices. *Cities*, 89, pp. 268–280. <https://doi.org/10.1016/j.cities.2019.03.006>
- Popescu, A., y Pennacchiotti, M. (2011). Dancing with the Stars, NBA Games, Politics: An Exploration of Twitter Users' Response to Events. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pp. 594–597. <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewPDFInterstitial/2866/3233>
- Raun, J., Ahas, R., & Tiru, M. (2016). Measuring tourism destinations using mobile tracking data. *Tourism Management*, 57, pp. 202–212. <https://doi.org/10.1016/j.tourman.2016.06.006>
- Steiger, E., Ellersiek, T., Resch, B., y Zipf, A. (2015). Uncovering latent mobility patterns from Twitter during mass events. *Journal for Geographic Information Science*, 1, pp. 525–534. <https://doi.org/10.1553/giscience2015s525>
- Tascón, M., y Coullaut, A. (2016). *Big Data y el internet de las cosas: qué hay detrás y cómo nos va a cambiar*. Madrid, Catarata, 128 pp.
- Versichele, M., Neutens, T., Delafontaine, M., y van de Weghe, N. (2012). The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. *Applied Geography*, 32(2), pp. 208–220. <https://doi.org/10.1016/j.apgeog.2011.05.011>
- Weng, J., Yao, Y., Leonardi, E., y Lee, F. (2011). Event Detection in Twitter. *Proceedings of the Fifth International Conference on Weblogs and Social Media*.
- Xu, Y., y González, M. C. (2017). Collective benefits in traffic during mega events via the use of information technologies. *Journal of The Royal Society Interface*, 14(129), pp. 1–10. <https://doi.org/10.1098/rsif.2016.1041>
- Zeng, B., y Gerritsen, R. (2014). What do we know about social media in tourism? A review. *Tourism Management Perspectives*, 10, 27–36. <https://doi.org/10.1016/j.tmp.2014.01.001>
- Zhou, X., y Xu, C. (2017). Tracing the Spatial-Temporal Evolution of Events Based on Social Media Data. *ISPRS International Journal of Geo-Information*, 6(3), 88. <https://doi.org/10.3390/ijgi6030088>