

Revista de Estudios Andaluces (REA)

e-ISSN: 2340-2776.

REA Vol. 33 (2016). <http://dx.doi.org/10.12795/rea.2016.i33>

Big (Geo)Data en Ciencias Sociales: Retos y Oportunidades

Big (Geo)Data in Social Sciences: Challenges and Opportunities

Javier Gutiérrez-Puebla

Universidad Complutense de Madrid

javiergutierrez@ghis.ucm.es

Juan Carlos García-Palomares

Universidad Complutense de Madrid

jcgarcia@ghis.ucm.es

María Henar Salas-Olmedo

Universidad Complutense de Madrid

mariahenar.salas@pdi.ucm.es

Formato de cita / Citation: Gutiérrez-Puebla, J.; García-Palomares, J.C.; Salas-Olmedo, M.H. (2016): Big (Geo)Data en Ciencias Sociales: Retos y oportunidades. *Revista de Estudios Andaluces*, vol. 33 (1), 1-23. <http://dx.doi.org/10.12795/rea.2016.i33.01>

Enlace artículo / to link to this article: <http://dx.doi.org/10.12795/rea.2016.i33.01>



Esta obra se distribuye con la licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional

<http://editorial.us.es/es/revista-de-estudios-andaluces>

<https://ojs.publius.us.es/ojs/index.php/REA>

Big (Geo)Data en Ciencias Sociales: Retos y Oportunidades

Big (Geo)Data in Social Sciences: Challenges and Opportunities

Javier Gutiérrez-Puebla

Universidad Complutense de Madrid

javiergutierrez@ghis.ucm.es

Juan Carlos García-Palomares

Universidad Complutense de Madrid

jcgarcia@ghis.ucm.es

María Henar Salas-Olmedo

Universidad Complutense de Madrid

mariahenar.salas@pdi.ucm.es

Recibido: 12 de abril, 2016

Revisado: 16 de abril, 2016

Aceptado: 22 de abril, 2016

Resumen

Actualmente asistimos a una verdadera revolución en la producción y el tratamiento de datos masivos (Big Data). Aunque los principales usuarios de este tipo de datos son las empresas, el mundo de la investigación ha encontrado también interesantes posibilidades en el análisis de Big Data, con abordajes nuevos a viejos problemas o incluso con el planteamiento de cuestiones que no podían ser abordadas con datos tradicionales. El presente artículo constituye una revisión de trabajos de investigación que utilizan datos masivos geolocalizados, Big (Geo)Data, y muestra ejemplos de aplicación en la investigación, ordenando los trabajos revisados según fuentes de datos: registros de llamadas de teléfonos móviles, redes sociales, comunidades de fotografías geolocalizadas, registros de transacciones con tarjetas de crédito, tarjetas inteligentes de transporte, navegadores, etc. El trabajo concluye con unas reflexiones sobre las ventajas que ofrece el Big (Geo)Data para el investigador, como la alta resolución espacial y temporal de los datos y, en muchos casos, su cobertura global y su carácter gratuito, pero también resalta algunos de los principales inconvenientes que plantea su uso, como el sesgo y la dificultad de su proceso y, en muchos casos, de acceso a los mismos.

Palabras clave: Big Data, datos geolocalizados, redes sociales, teléfonos móviles, tarjetas inteligentes.



Abstract

Currently we are witnessing a revolution in the production and processing of massive data (Big Data). Although the main users of such data are companies, social researchers have also found interesting possibilities in the analysis of Big Data, with new approaches to old questions or even with the approach to issues that could not be addressed with traditional data. This article is a review of research papers using geolocated massive data, Big (Geo)Data, and shows examples of their application in research, grouping the papers according to data sources: mobile phone calls records, social networks, communities of geolocated photos, credit card transactions records, transport smart cards, car navigators, etc. The paper concludes with some reflections on the advantages of Big (Geo)Data in social sciences research (high temporal and spatial resolution, and, in many cases, global coverage and free of charge), but it also highlights some of the main problems arising from their use, such as bias, processing capacity and access barriers.

Keywords: Big Data, geolocalized data, social networks, mobile telephones, smart cards



1. INTRODUCCIÓN

La producción de información geográfica ha cobrado un ritmo antes insospechado. Raras son las instituciones que no ponen a disposición del público su información geográfica, ya sea a través de zonas de descarga en sus webs, plataformas OpenData o mediante la creación de Infraestructuras de Datos Espaciales (IDEs). También las empresas se están sumando con fuerza a la producción de datos masivos, tratando de mejorar sus procesos productivos y su competitividad o abriendo nuevos nichos de mercado mediante la oferta de nueva información y servicios. Pero es, posiblemente, desde las acciones voluntarias, apoyadas en el desarrollo de la Web 2.0, desde donde más se ha acelerado la producción de datos masivos geolocalizados. Un ejemplo muy conocido al respecto es el de Open Street Maps, un proyecto colaborativo que consiste en la creación de un callejero digital de cobertura global elaborado y constantemente actualizado por voluntarios, que se puede descargar gratuitamente en la red. En tan solo unas décadas se ha pasado de una situación de escasez de datos digitales a otra de sobreabundancia, donde la dificultad surge no tanto de la disponibilidad de los datos necesarios para llevar a cabo nuestra investigación como de ser capaces de seleccionarlos y procesarlos adecuadamente. James Cheshire y Oliver Uberti (2014) en un atractivo libro, apoyado en mapas y gráficos, que pretende dar una nueva visión de la ciudad de Londres usando datos hasta hace poco tiempo no disponibles, comienzan diciendo: Este libro no podría haberse escrito hace diez años. Los ordenadores no eran lo suficientemente potentes. Los datos tenían tanto nivel de detalle o no eran de libre acceso. No existían tecnologías como Twitter y los teléfonos inteligentes (Cheshire and Uberti, 2014).

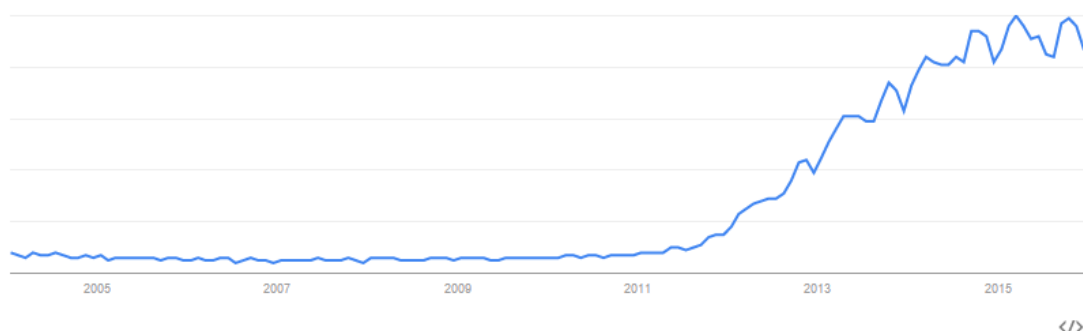
Big Data es un concepto emergente, que se ha popularizado enormemente en los últimos dos o tres años (Figura 1) y hace referencia a la producción de cantidades ingentes de datos por medio de múltiples redes de sensores y dispositivos. No solo hay que enfatizar el carácter masivo de los datos, sino que también que se trata de datos (estructurados y no estructurados) de distinta naturaleza a los convencionales, por lo que resultan complementarios a los suministrados por las estadísticas oficiales¹. En la era tecnológica actual, las actividades humanas, de forma voluntaria o involuntaria, dejan un rastro digital y con frecuencia esa huella está geolocalizada. Así, por ejemplo, generamos cantidades ingentes de datos geolocalizados cuando nos desplazamos (registrados a través del GPS de nuestros smartphones), cuando hacemos una llamada con nuestro teléfono móvil, cuando pagamos con tarjetas de crédito o presentamos

¹ Recientemente se está planteando el uso de big data para la producción de estadísticas oficiales. Es el caso del Instituto Nacional de Estadística de España, que en sus trabajos preparatorios para el Censo de 2021 está trabajando con datos de consumo de energía eléctrica para la clasificación de las viviendas (principales, secundarias y vacías) y registros de llamadas de teléfonos móviles para confeccionar estadísticas de movilidad diaria.



una tarjeta de fidelización, cuando interactuamos en las redes sociales o cuando somos captados por cámaras de video en un centro comercial. Como señala Batty (2013), la mayor parte de la información que ahora llamamos Big Data se produce de forma automática, de forma rutinaria, y por diversas formas de sensores. Casi todos estos datos en principio son capturados y almacenados para llevar a cabo los procesos de control y gestión en las empresas (por ejemplo, gestión de cargos en las tarjetas de crédito), pero después han sido utilizados para usos distintos a los que fueron almacenados, como el análisis del comportamiento de los consumidores para diseñar estrategias de marketing, la predicción de tendencias de mercado, el control del fraude o la generación de nuevas estadísticas más actualizadas y detalladas (Heershap et al., 2014). Un caso muy conocido en el campo de las estrategias de marketing es el de Facebook, que desarrolla algoritmos a partir del comportamiento de sus usuarios en la red (likes, contenidos visitados) para crear un perfil de cada uno de ellos y enviar publicidad de forma personalizada.

Figura 1. Evolución del número de consultas en el buscador Google sobre Big Data (Datos relativos. Google Trends reescala los datos hasta un valor máximo de 100)



Fuente: Google Trends, 2015.

En el mundo de la web 2.0, los usuarios de internet ya no adoptan una actitud pasiva, de meros receptores de información, sino que se convierten en productores de ingentes cantidades de datos, particularmente a través de las redes sociales. La geolocalización de tweets o el análisis de las relaciones a partir de Facebook, son dos de los ejemplos más conocidos y brindan grandes posibilidades para el análisis de las redes sociales y su impronta espacial (ver, por ejemplo, los famosos mapas de redes de usuarios de Facebook o algunos análisis con datos de Twitter como el de las lenguas de Londres que se puede encontrar en <http://mappinglondon.co.uk/2012/londons-twitter-tongues/>).



Varios trabajos recientes aparecidos en la revista *Dialogues in Human Geography* ponen el foco en los retos y las oportunidades del uso de Big (Geo)Data en la Geografía Humana en general, y en el estudio de la ciudad y la planificación urbana en particular (ver Batty, 2013; Kitchin, 2013; Graham and Shelton, 2013). Como sugiere el título del trabajo de Graham y Shelton (2013) "Geography and the future of Big Data, Big Data and the future of geography", el futuro de la Geografía y los datos masivos va de la mano. Las fuentes de datos disponibles permiten a los investigadores poner fin a su dependencia de las estadísticas oficiales, en campos tan diversos como la demografía, la actividad económica, la movilidad y los flujos, y cualquier otro de los aspectos urbanos (Shelton et al, 2015). No es de extrañar, pues, el cada vez mayor interés que despiertan las nuevas fuentes de datos entre los geógrafos, cada vez más atraídos por el carácter innovador de estos datos y la relevancia que empiezan a adquirir en el estudio de la ciudad actual.

Entre las principales ventajas del Big Data se pueden citar la alta resolución espacial y temporal, lo que permite la monitorización de procesos espacio-temporales, así como el ofrecer información complementaria a la de las fuentes oficiales. Dado que una buena parte del Big Data contiene datos geolocalizados, éstos pueden ser tratados y analizados con Sistemas de Información Geográfica y técnicas de estadística espacial, además de con herramientas de la estadística clásica. La visualización animada cobra una especial importancia en este tipo de datos geolocalizados, precisamente por su alta resolución espacial y temporal. En resumen, los datos masivos pueden ser procesados para convertirlos en información y a partir de ésta generar conocimiento.

Aunque no hay un consenso claro al respecto, se citan como principales características del Big Data las siguientes (Kitchin, 2013):

- Su enorme volumen (terabytes o petabytes de datos).
- La enorme velocidad a la que son generados (en tiempo real o casi tiempo real).
- Su diversidad y variedad, incluyendo datos estructurados (en formato base de datos) y no estructurados (por ejemplo, líneas de texto).
- Su exhaustividad, cubriendo poblaciones o sistemas enteros.
- Su resolución de grano fino, con identificadores que permiten seguir a individuos u objetos.
- Su naturaleza relacional, al tener campos comunes que permiten combinar bases de datos distintas.
- Su flexibilidad, tanto desde el punto de vista de su extensión (se pueden añadir nuevos campos fácilmente) como de la escalabilidad (se puede aumentar el volumen de datos).



Big (Geo)Data cubre ámbitos muy distintos: actividad de los teléfonos móviles, registros de operaciones con tarjetas de crédito, datos recogidos en tiempo real con GPS, redes sociales, registros de consumo de agua y electricidad, imágenes grabadas con cámaras de video, tarjetas de transporte público o de uso de sistemas públicos de bicicletas, etc. Las aplicaciones en la investigación son muy variadas, cubriendo campos como el marketing, la movilidad, el turismo, la diferenciación social, etc. En este trabajo se realiza una revisión de diferentes trabajos de investigación que han utilizado Big Data, ordenándolos según las fuentes utilizadas. La revisión no es exhaustiva, pero permite una aproximación inicial al investigador que por primera vez se adentra en el mundo del Big (Geo)Data.

2. REGISTROS DE LA ACTIVIDAD DE LOS TELÉFONOS MÓVILES

Una de la fuentes de Big Data más populares en la investigación del comportamiento humano es sin duda la constituida por los registros de la actividad de los teléfonos móviles (llamadas, mensajes), los denominados CDR (Call Detail Records). Esta actividad es captada por las antenas más próximas al teléfono móvil y puede ser agregada y representada a partir de las áreas de influencia de las antenas, generalmente delimitadas mediante polígonos de Thiessen. Además, a partir de cálculos geométricos es posible conocer con bastante exactitud la localización del usuario (con un error de unos 50 metros). Esta fuente de datos, facilitada preservando el anonimato de los clientes de las operadoras, abre expectativas prometedoras en la investigación. Las fuentes tradicionales (censo y padrón) ofrecen información sobre la distribución espacial de la población en la noche (dónde vive la población), pero no sobre la forma en que utiliza el espacio a lo largo del día. Así, los datos recolectados a partir de los teléfonos móviles se pueden usar para cartografiar la densidad de llamadas a distintas horas del día como un indicador de la evolución espacio-temporal de la intensidad de actividades en la ciudad (Ratti et al., 2006, Reades et al., 2009). Dado que esta densidad de llamadas varía según franjas horarias, reflejando las densidades cambiantes de población, es posible analizar los patrones de la estructura urbana (Reades et al., 2009), deduciendo los usos del suelo a partir de los patrones temporales de las llamadas, muy frecuentes en áreas de actividad en las horas centrales del día y en áreas de residenciales en la tarde y primeras horas de la noche (Louail et al., 2014; Grawing et al., 2015).

Los registros de la actividad de los teléfonos móviles han sido también utilizados en estudios de movilidad y transporte. El seguimiento de un mismo individuo a partir de la actividad de su teléfono móvil permite conocer sus pautas de movilidad y elaborar modelos predictivos (De Domenico et al., 2013), generar matrices origen-destino que cuantifican el volumen de movimientos según franjas horarias (Cáceres et al., 2007), estimar velocidades de desplazamiento y tiempos de viaje (Bar-Gera, 2007) o aforar el tráfico de la red viaria en carreteras que no disponen de aforadores convencionales a partir de correlaciones obtenidas en los tramos que sí disponen de estos dispositivos (Cáceres, 2012). Así mismo esta fuente de datos ha sido utilizada para estudiar las



redes de relaciones de amistad (Eagle et al., 2009) o para, en el campo de la geografía de los riesgos naturales, hacer un seguimiento del comportamiento de la población ante una catástrofe, como por ejemplo las inundaciones (Pastor-Escuredo, 2014).

3. REDES SOCIALES: TWITTER Y FOURSQUARE

En los últimos años se ha producido una verdadera proliferación de redes sociales, comunidades de individuos con intereses o actividades comunes que a través de Internet se comunican e intercambian información. Las redes sociales ofrecen datos de gran interés para el investigador en ciencias sociales. Una de las más difundidas es sin duda Twitter, una plataforma que permite enviar mensajes con un máximo de 140 caracteres, lo denominados tweets, que son por defecto son públicos. Este servicio cuenta con 500 millones de usuarios en todo el mundo y genera unos 65 millones de tweets al día. Una pequeña parte de ellos (alrededor de un 3% hasta abril de 2015 y de un 1% en la actualidad) son tweets geolocalizados, es decir, mensajes para los que se conoce la localización del emisor por sus coordenadas geográficas.

La mayor parte de los estudios realizados con datos masivos de redes sociales han utilizado Twitter (Murthy, 2013), no solo por tratarse de una plataforma de alcance mundial, sino también por el hecho de que sus datos (los tweets) están disponibles en la red de forma gratuita a medida que se producen, es decir, en tiempo real. Se pueden descargar tweets geolocalizados (o no), registrando el hashtag, indicador de usuario, momento en el que se envía, idioma, tipo de dispositivo, texto del mensaje, etc. Cada tweet deja una “huella” digital del lugar y el momento en que fue emitido. Si se procesan los datos según identificador de usuario, se puede tener una aproximación de los lugares que visita cada usuario en los distintos momentos del día y días de la semana, es decir, su perfil espacio-temporal. Así es posible utilizar esta fuente como una proxy para analizar las densidades de población cambiantes a lo largo del día en la ciudad (Ciuccarelli et al., 2014) y las pautas de movilidad de la población (Wu et al., 2014). La fiabilidad de este tipo de datos ha sido validada en el trabajo de Lenormand et al. (2014), quienes tras contrastar datos de Twitter con información de teléfonos móviles y datos oficiales (censos) concluyeron que las tres fuentes ofrecen resultados comparables en términos de distribución espacial de la población, evolución temporal de las densidades de población y pautas de movilidad de los individuos.

A partir del perfil horario de los tweets es posible deducir las características de la estructura urbana. Así, Frías-Martínez et al. (2012) utilizaron la variación espacio-temporal de los tweets para diferenciar el uso del suelo urbano e identificar los principales puntos de interés (monumentos) de Nueva York. Por otro lado se ha constatado que el perfil horario de los tweets es diferente según se trata de ocupados (mayor cantidad de tweets enviados en las primeras horas de la mañana) y desempleados (más tweets al final de la tarde y primeras horas de la noche), con lo



que del comportamiento en las redes sociales se pueden estimar los índices de paro de los municipios (Llorente et al., 2015). Los residuos obtenidos en este modelo, una vez cartografiados, demostraron ser un buen indicador de la economía sumergida de los municipios. De la actividad en Twitter también se puede extraer información relevante relativa al impacto económico de las catástrofes naturales y por lo tanto de los pagos que deben afrontar las compañías de seguros

Así mismo se han utilizado tweets geolocalizados para analizar el grado de mezcla social en el uso del espacio, rastreando el movimiento de los grupos sociales en ciudades fuertemente segregadas como Río de Janeiro (Netto et al., 2005) y Louisville (Shelton et al., 2015). A diferencia de la información suministrada por las fuentes oficiales, que ofrecen datos relativos al lugar de residencia, en estos estudios los indicadores de multiculturalidad y mezcla elaborados a partir de Big Data se refieren al uso del espacio a lo largo del día. Así, por ejemplo, hay estudios sobre la diversidad lingüística en ciudades y regiones, a partir de las lenguas utilizadas en los tweets, como indicador de diversidad cultural (Mocanu et al., 2013). Por su parte, Takhteyev et al. (2012) estudiaron el papel de la distancia, las lenguas y las fronteras en la configuración de las redes de tweets, concluyendo que las relaciones que se establecen sobre largas distancias son similares a las que se observan en los flujos de transporte aéreo.

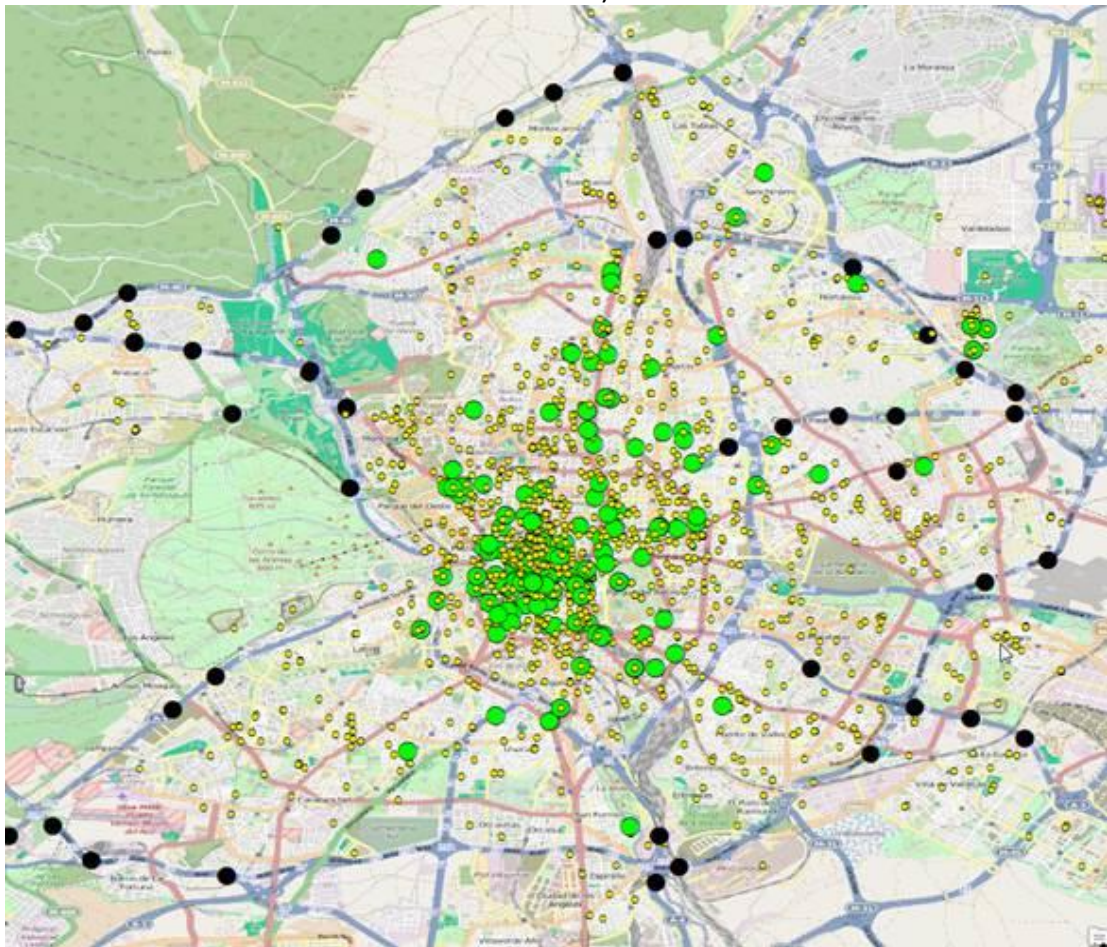
Una línea prometedora de investigación es el análisis de movimientos sociales analizando los contenidos de los tweets, como es el caso de las protestas contra el cambio climático (Segeberg and Bennett, 2011) o las concentraciones del movimiento 15M (Ferrerias, 2011). El uso de Big Data también permite medir la concentración de personas en protestas sociales por medio de tweets geolocalizados, como por ejemplo el movimiento Occupy Wall Street (Conover et al., 2013). Estas concentraciones pueden ser monitorizadas en tiempo real, como ocurrió en el caso con las manifestaciones de Vía Catalana y la V catalana <http://ifisc.uibcsic.es/humanmobility/>. En efecto, los tweets pueden ser utilizados para el seguimiento de fenómenos en tiempo real, filtrando por el hashtag del mensaje (la intensidad de tweets con el hashtag de la manifestación es un indicador de la concentración de población en un determinado momento y un determinado lugar). También los hashtags pueden ser utilizados para la descarga de tweets geolocalizados sobre un problema determinado en una ciudad, como por ejemplo los atascos de tráfico (Figura 2), o para seguimiento de catástrofes, como los terremotos (Sakaki et al., 2010).

Otra red social de interés es Foursquare (<https://es.foursquare.com>). Su idea principal es que los usuarios puedan indicar (check-in) los lugares específicos en que se encuentran y hacer comentarios sobre los mismos. A partir de la información facilitada por los usuarios, el servicio ha ido evolucionando hacia un motor de recomendaciones que sugiere lugares interesantes de manera inteligente. Frente a Google Places, centrado en suministrar datos de oferta, Foursquare añade la perspectiva del uso real de los lugares y permite conocer cuáles son los lugares preferidos o socialmente



relevantes de una ciudad (Serrano-Estrada et al. 2014). Por su parte, Google Places, servicio basado en la oferta, proporciona diferentes datos para cada establecimiento: nombre, coordenadas, dirección, categoría, teléfono, dirección web y horarios. Es posible descargar las coordenadas de cada negocio y cartografiar la densidad y diversidad de actividades (comercios, restaurantes, oficinas bancarias, etc.) (Nolasco-Cirugeda, A. y García Mayor, C. (2014).

Figura 2. Tweets geolocalizados en un día laborable a primera hora en Madrid (enero de 2016).



Los puntos negros son tweets que mencionan los atascos, los puntos verdes mencionan Madrid (excluidos los atascos) y los puntos amarillos son el resto de tweets enviados desde Madrid.

Fuente: Imagen cedida por Henar Salas.



4. REDES SOCIALES DE FOTOGRAFÍAS GEOLOCALIZADAS

Un tipo especial de redes sociales es el de las comunidades que comparten fotografías. Estos servicios tienen el potencial de poder medir el atractivo del espacio (Kachkaev and Wood, 2013). Destacan particularmente tres comunidades: Instagram, Flickr y Panoramio, aunque existen otras como Picasa, etc. Las tres proporcionan la opción de geolocalizar las fotografías.

- Instagram (<https://instagram.com/>) es una red social que permite a los usuarios aplicar efectos a sus fotografías o videos y después compartirlos en diferentes redes sociales como Facebook, Twitter, Tumblr o Flickr. A finales de 2014 Instagram contaba con más de 300 millones de usuarios.

- Flickr (<http://www.flickr.com/map>), de Yahoo, cuenta con una importante comunidad de usuarios que comparten fotografías y videos creados por ellos mismos. Las dos tienen la posibilidad de geolocalizar las fotografías. Sin embargo, tanto en Instagram como en Flickr el acento se pone en el tratamiento y retoque de las fotografías.

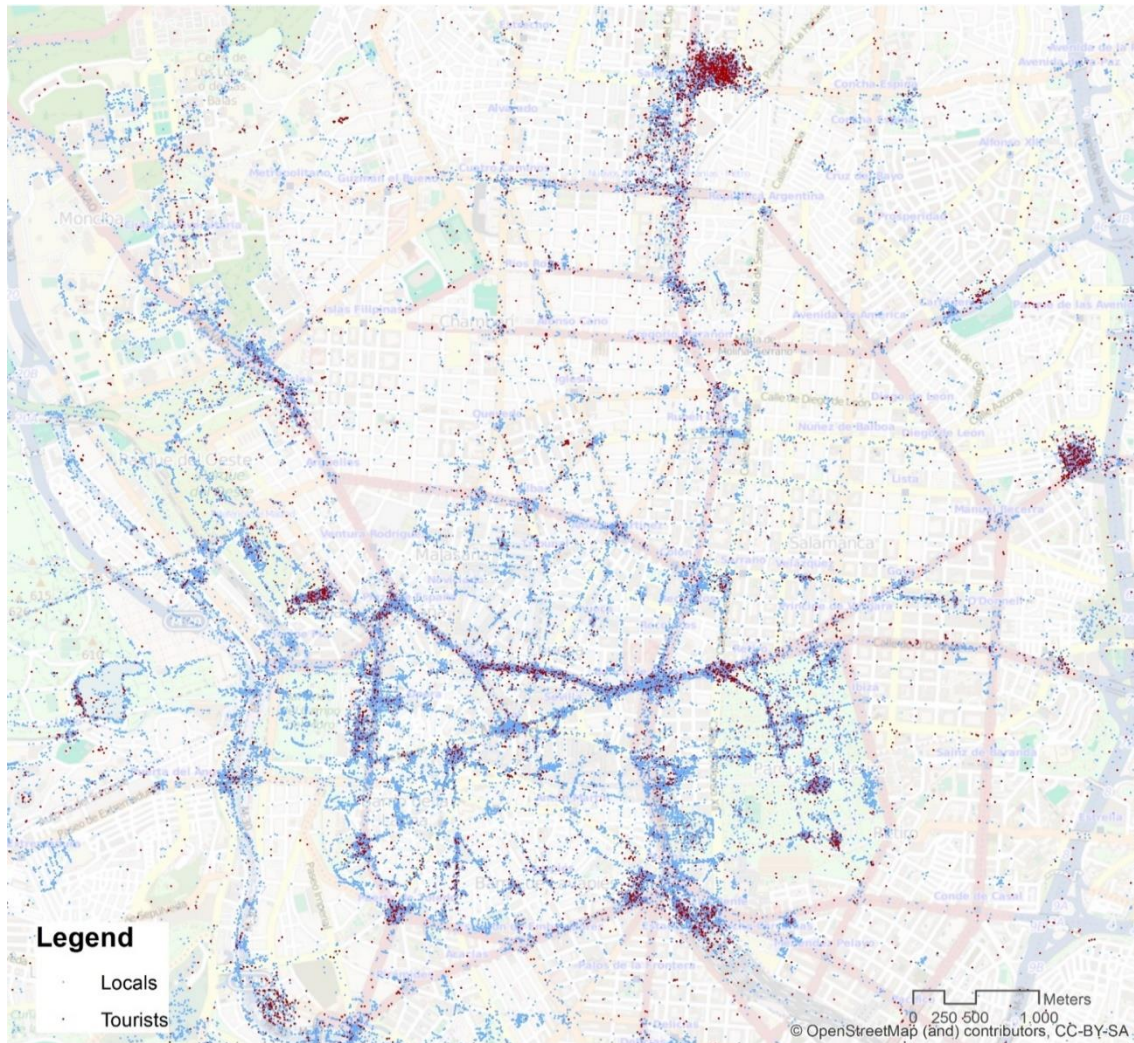
- Panoramio (<http://www.panoramio.com>) pone mayor atención en la georreferenciación de las fotografías compartidas por los usuarios. Panoramio es un sitio web dedicado a exhibir fotografías de lugares o paisajes que los usuarios toman y cuelgan una vez han sido georreferenciadas. Su objetivo es permitir a los interesados conocer más sobre una zona específica del globo, observando las fotografías que otros han tomado allí. Estas imágenes pueden ser vistas en la propia web de Panoramio o a través de Google Earth. Desde su creación en octubre de 2005, su éxito ha sido enorme, y en apenas dos años recibió más de 5 millones de fotografías. En julio de 2007, Panoramio fue adquirido por Google. Desde entonces, las subidas han crecido exponencialmente hasta superar las 100 millones de imágenes en diciembre de 2013.

Los servicios de fotografías compartidas han sido utilizados para identificar los principales puntos de atracción turística y la intensidad de uso de los mismos (Popescu et al., 2009; Kisilevich et al., 2010; Gavric et al., 2011; Straumann et al., 2014). Sun y Fan (2014) utilizan fotografías geolocalizadas y regresión binaria logística para identificar eventos sociales, como festivales, manifestaciones, acontecimientos deportivos, etc. Koerbitz et al. (2013) a partir de datos de Flickr y regresión polinómica llegan a estimar el número de turistas en Austria. Varios trabajos han usado la información de photo-sharing services para analizar la movilidad de los turistas, y proponer o evaluar rutas turísticas, como Kurashima et al (2013) o De Choudhury et al. (2010). Lu et al (2010) aprovechan estas fuentes para sugerir viajes turísticos y Li (2013) para planificar viajes de varios días y lugares a visitar. Finalmente, García-Palomares et al. (2015) aplican técnicas de estadística espacial para identificar las concentraciones de turistas (clústers espaciales) a partir de fotografías descargadas de Panoramio (Figura 3). La diferenciación entre turistas y locales se realiza a partir del identificador de usuario y el tiempo que utiliza el servicio en una ciudad: cuando ese



uso es continuo (más de un mes) ese usuario se identifica como local; cuando por el contrario un usuario sube fotografías de una ciudad durante un periodo muy corto de tiempo (menos de un mes) entonces pasa a ser considerado turista. Los resultados obtenidos mostraron distribuciones espaciales claramente diferentes para turistas y locales.

Figura 3. Fotografías geolocalizadas de turistas (en rojo) y residentes (en azul) en el centro de la ciudad de Madrid.



Fuente: García Palomares et al., 2015.



5. PLATAFORMAS WEB PARA LA RESERVA DE ALOJAMIENTOS

En los últimos años han tenido un gran éxito los portales web para la reserva de hoteles, como TripAdvisor o Booking, que albergan información sobre millones de establecimientos en todo el mundo. Los clientes no adoptan ya una actitud pasiva de meros receptores de la información que se les facilita, sino que adquieren un papel protagonista, publicando comentarios e incluso calificando al establecimiento en una escala de 0 a 10. Esta información (cualitativa y cuantitativa) es extraordinariamente rica. Estos datos, una vez procesados, permiten abordar viejos problemas con enfoques nuevos o incluso analizar cuestiones que no tenían respuesta con datos convencionales. Los trabajos pioneros de O'Connor (2008 y 2010) han puesto de manifiesto que el proceso de los datos contenidos en TripAdvisor no solo suministran información de gran interés para el turista a la hora de planificar los viajes, sino que también contribuyen de forma decisiva a conformar la imagen de las empresas del sector turístico. Esto es reconocido por las empresas del sector cuando en la recepción de los hoteles sitúan carteles con la puntuación media concedida por los clientes en Tripadvisor, Expedia o Booking.

Frente a los modelos de negocio de la economía formal, que conectan la empresa con el consumidor (B2C), como es el caso de Expedia o Booking para la reserva de hoteles, el modelo de negocio de la economía colaborativa utiliza plataformas alternativas para poner directamente en contacto a unos individuos con otros (P2P), ya se trate de anfitriones y viajeros (Airbnb) o de gente que quiere intercambiar sus alojamientos sin ninguna contrapartida económica (CouchSurfing). Estos portales contienen datos de gran interés para conocer el desarrollo de la economía colaborativa en el campo del sector turístico. Airbnb, cuyo modelo de negocio ha sido considerado como una disrupción en el sector de los alojamientos turísticos (Guttentag, 2013), provee una muy amplia oferta de habitaciones y viviendas en todo el mundo (más de dos millones de reseñas, sumando habitaciones y viviendas enteras), mayor a la de las grandes cadenas internacionales de hoteles. Los datos de su portal referentes a diferentes ciudades pueden ser descargados utilizando su API, con lo que es posible disponer de datos geolocalizados de cada uno de sus alojamientos. A partir de estos datos se han podido estudiar las pautas de concentración espacial de la oferta de Airbnb, comprobándose que, en el caso de Barcelona, los alojamientos ofertados se localizan preferentemente en el centro, en un cinturón que rodea al eje de mayor concentración de hoteles (Ramblas – Paseo de Gracia), aumentando así la presión turística sobre el centro y favoreciendo los procesos de gentrificación (Figura 4).

Las páginas web con comentarios y valoraciones de consumidores también tienen gran interés para el análisis del sector de la restauración. Zhai et al. (2015) descargaron datos de la ciudad de Hangzhou (China) desde un portal web que contiene puntuaciones de los clientes para restaurantes según distintas dimensiones (calidad, servicio, recomendaciones...). Estas dimensiones fueron tratadas como variables en un análisis de componentes principales que permitió obtener un índice de popularidad,

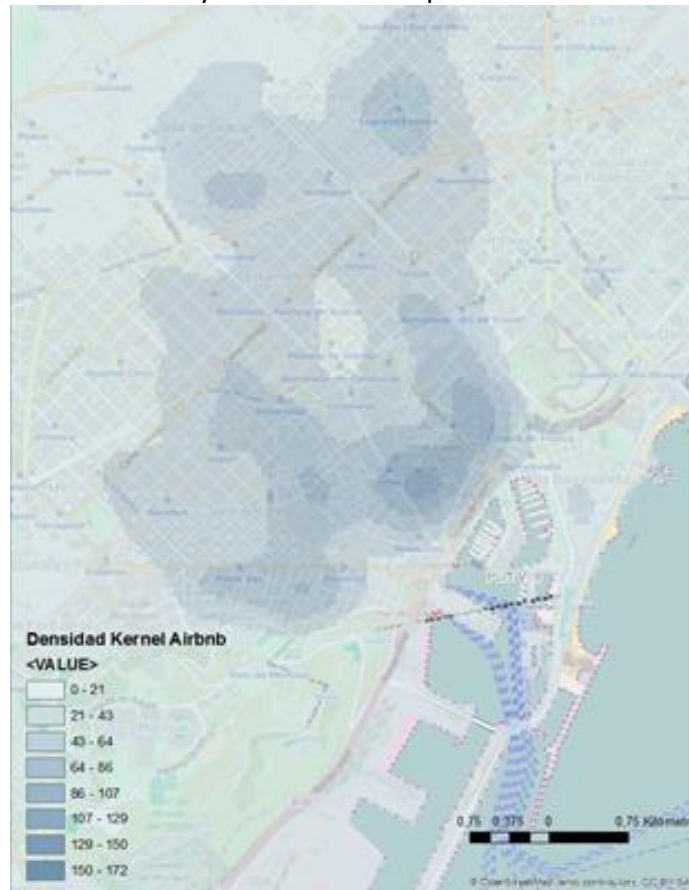


realizar un análisis de clústers espaciales y relacionar los resultados obtenidos con variables del entorno (hoteles, bancos, centros comerciales, etc.).

6. REGISTROS DE TRANSACCIONES DE TARJETAS BANCARIAS

Otra interesante fuente de datos masivos, explotada solo muy recientemente, son las tarjetas bancarias. Las entidades financieras disponen de información sobre los poseedores de estas tarjetas (domicilio, edad, sexo...) y sobre cada una de sus compras con tarjeta (hora, día, lugar, tipo de establecimiento, volumen de gasto, etc.). Cruzando estos datos y agregándolos geográficamente se puede obtener información de gran interés desde la perspectiva del consumo y de los viajes a compras y actividades de ocio. Seleccionando las transacciones realizadas por personas no residentes en una determinada ciudad se pueden analizar las pautas espacio-temporales del gasto realizado por el colectivo de turistas (Figura 5).

Figura 4. Densidad de oferta de camas de Airbnb en Barcelona. Kernel decreciente con radio de 500 metros y datos de oferta previamente normalizados.

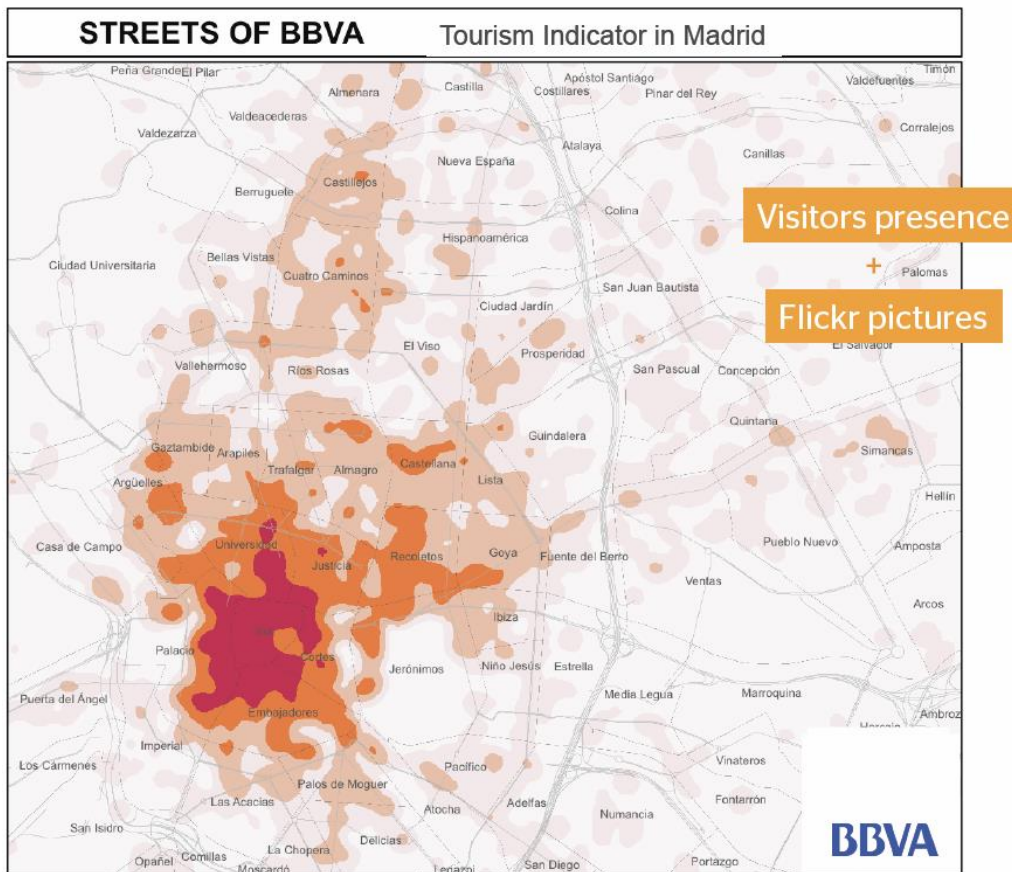


Fuente: Elaboración propia



La mayor parte de los trabajos realizados con registros de transacciones con tarjetas bancarias están orientados a optimizar los procedimientos bancarios y al control del fraude. Desgraciadamente los trabajos de investigación sobre el comportamiento humano basados en este tipo de registros son todavía escasos, debido a la dificultad para acceder a estos datos. Krumme et al. (2013) utilizaron datos de tarjetas de crédito para predecir las pautas de comportamiento de los consumidores. Sobolevsky et al., (2014a y 2015a) partieron de datos de los registros de transacciones con tarjetas de crédito de un gran banco para determinar el atractivo de cada una de las ciudades españolas. Así mismo, Sobolevsky et al. (20015b) utilizaron esta misma fuente de datos para caracterizar el perfil del gasto de los residentes en cada una de las mayores ciudades españolas y establecer una clasificación de las mismas.

Figura 5. Indicador de presencia de turistas a partir de registros de transacciones de tarjetas bancarias y fotografías geolocalizadas en Flickr.



Fuente: <https://www.centrodeinnovacionbbva.com/noticias/6564-el-uso-eficiente-de-la-informacion-y-el-modelo-de-smart-cities-que-queremos>

Dado que los registros de datos de transacciones con tarjetas de crédito contienen las coordenadas espacio-temporales de cada transacción y un identificador de usuario, es posible seguir la movilidad espacial de cada consumidor. En este sentido puede destacarse el artículo de Lenormand et al. (2015) en el que se analizan las pautas de

Revista de Estudios Andaluces, vol. 33, núm. 1 (2016) pp. 1-23. e-ISSN: 2340-2776
<http://dx.doi.org/10.12795/rea.2016.i33.01>



Esta obra se distribuye con la licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional

movilidad en los desplazamientos a compras y ocio, así como el de Sobolesky et al. (2014b) sobre las redes de movilidad conformadas al considerar la procedencia de los turistas y su modelización mediante un modelo gravitatorio.

7. TARJETAS INTELIGENTES DE TRANSPORTE

Una fuente de gran interés para el estudio de la movilidad en transporte público es la enorme cantidad de datos generada por las tarjetas inteligentes de transporte (transport smart cards). Se trata de las conocidas tarjetas de abono de transporte público, que ahora permiten almacenar datos sobre los viajes de los usuarios. Existen básicamente dos tipos de tarjetas. Las tarjetas sin contacto registran automáticamente la entrada y salida del viajero de una estación o de un vehículo de transporte público; en cambio, las tarjetas de contacto exigen que el usuario pase la tarjeta por un lector, cosa que hace solamente en la entrada. Por lo tanto mediante estas tarjetas se dispone no solamente de las características del usuario (registradas al adquirir o renovar la tarjeta), sino también de todos sus movimientos en transporte público: el lugar de acceso (y, en su caso, de salida) a los medios de transporte público (paradas o estaciones), la hora del día y el día de la semana. La dificultad a la hora de procesar estos datos es lógicamente mayor en el caso de las tarjetas de contacto, ya que -al registrar datos solamente sobre la entrada del viajero- para conocer la salida se deben desarrollar algoritmos que tengan en cuenta la parada o estación de entrada en el viaje de vuelta, considerando la posibilidad de realización de viajes multietapa y de viajes no basados en casa.

A partir de los datos de las tarjetas inteligentes de transporte es posible calcular la longitud de las rutas y generar matrices origen-destino (Munizaga et al., 2010). Además la información espacio-temporal se puede cruzar con los datos socioeconómicos de los viajeros, lo que proporciona una muy valiosa información para la planificación (Bagchi y White, 2003, Pelletier et al, 2011). De especial interés es el trabajo de Tao et al. (2014) explorando las dinámicas espacio-temporales y los flujos en transporte público a partir de las smart cards, con predominio de los viajes de componente centrípeta en la mañana y centrífuga en la tarde.

8. NAVEGADORES Y SERVICIOS WEB PARA EL CÁLCULO DE RUTAS ÓPTIMAS

La información sobre redes de transporte privado ha mejorado decisivamente en los últimos años gracias a las compañías de navegadores para automóviles, particularmente Navteq y Teleatlas (ésta última recientemente adquirida por TomTom). Estas detalladas redes viarias contienen información sobre las características del viario (tiempos de recorrido de cada tramo, direcciones privadas, prohibiciones de giros, etc.) para permitir el cálculo de rutas óptimas en los navegadores. Algunos investigadores han utilizado estas redes en sus estudios sobre



movilidad sin contar con datos de velocidades (por ejemplo, Neutens et al., 2010 o Zielstra y Hochmair, 2011). Pero lo que merece ser destacado como realmente novedoso es el nuevo producto de TomTom “Speed Profiles”, una red de transporte privado que se puede utilizar en ArcGIS y que incluye las velocidades de los vehículos en cada uno de los tramos de la red viaria cada cinco minutos y por lo tanto permite realizar análisis dinámicos de accesibilidad considerando el efecto de la congestión. Se trata de datos históricos obtenidos por distintos dispositivos, incluyendo los propios navegadores y los GPS de teléfonos móviles. Esta nueva fuente ha sido muy poco explotada hasta la fecha. Un análisis dinámico de cómo los cambios en la congestión afectan a la accesibilidad en Madrid y Barcelona, respectivamente, se puede consultar en (Moya-Gómez y García-Palomares, 2015).

Por su parte, Google Maps ha seguido mejorando su aplicación para el cálculo de caminos mínimos en la web. Utilizando su API es posible calcular matrices origen-destino de mínimo tiempo en los distintos modos de transporte (vehículo privado, transporte público, a pie y en bicicleta). En función del número de orígenes y destinos la consulta es gratuita o no. En el caso del transporte privado se puede elegir la hora y el día de la semana (datos históricos) o directamente pedir el cálculo en tiempo real (estimado). La red de transporte privado se alimenta fundamentalmente de datos de smartphones que llevan el GPS activado. Además es posible calcular rutas en transporte público con información suministrada por las autoridades de transporte. Estos ficheros de transporte público pueden ser descargados para realizar análisis de accesibilidad en un SIG mediante rutinas de análisis de redes o para analizar el nivel de cobertura de la red de transporte público, las velocidades medias o las superposiciones de las líneas (Hadas, 2013).

9. CONSIDERACIONES FINALES

La generación de datos geolocalizados aumenta de forma exponencial a partir de multitud de sensores que registran la actividad humana y de dispositivos con los que los usuarios interactúan. El proceso de Big Data implica la realización de una serie de operaciones de forma secuencial, como son la captura de datos, almacenamiento, depuración, agregación, análisis, explicación y predicción. Debido a la enorme cantidad de datos a procesar, normalmente se recurre a la computación en la nube con plataformas especialmente diseñadas para el tratamiento de datos masivos.

El uso de Big (Geo)Data en la investigación en ciencias sociales está solo en sus comienzos, pero no cabe duda de que está abriendo nuevas y prometedoras posibilidades. No solo se pueden contestar a algunas de las preguntas tradicionales desde ópticas distintas (por ejemplo, por la mayor resolución espacial y temporal de los datos), sino que se pueden formular preguntas de investigación que no podían obtener respuesta recurriendo a las fuentes tradicionales.



El uso de Big Data aporta importantes ventajas al investigador. Ofrecen una información complementaria a la de las fuentes de datos tradicionales, lo que permite responder a las preguntas de investigación desde otra perspectiva. Muchas de estas fuentes tienen una cobertura global, con lo que se pueden abordar estudios comparativos entre ciudades o entre países. Además los datos masivos geolocalizados tienen una muy alta resolución espacial y temporal: espacial, porque cada dato está localizado por sus coordenadas geográficas y no agregado espacialmente; temporal, porque para cada dato se almacena el momento en que se genera (año, mes, día, hora, minuto y segundo), de forma que se dispone de información siempre actual y se pueden realizar estudios evolutivos y monitorizar procesos. Algunas de estas fuentes de datos son gratuitas y se pueden descargar directamente desde Internet utilizando la correspondiente API.

Sin embargo, no todos son ventajas. Algunas de estas fuentes de datos masivos son de difícil acceso, bien porque las compañías que los generan no están dispuestas a compartirlos con los investigadores o porque son puestos a la venta a precios elevados. Un segundo problema deriva de la dificultad de procesar los datos masivos, ya que su volumen excede las capacidades de los gestores de bases de datos convencionales. Finalmente este tipo de datos presenta el inconveniente de que generalmente son sesgados, lo que obliga a utilizar otras fuentes de forma complementaria para intentar compensar el sesgo. Piénsese, por ejemplo, en la utilización de las redes sociales, que no son utilizadas por toda la población ni en la misma intensidad en cada segmento de usuarios. El corolario de lo anterior es que los análisis llevados a cabo con Big (Geo)Data tienen fundamentalmente un carácter exploratorio, sin que generalmente sea posible determinar estadísticamente los márgenes de error y los niveles de significación.

REFERENCIAS

Bagchi, M., & White, P. R. (2005): "The potential of public transport smart card data", *Transport Policy*, 12(5), 464-474. <http://dx.doi.org/10.1016/j.tranpol.2005.06.008>

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): 380-391. <http://dx.doi.org/10.1016/j.trc.2007.06.003>

Batty, M. (2013): "Big Data, smart cities and city planning", *Dialogues in Human Geography*, 3(3), 274-279. <http://dx.doi.org/10.1177/2043820613513390>

Revista de Estudios Andaluces, vol. 33, núm. 1 (2016) pp. 1-23. e-ISSN: 2340-2776
<http://dx.doi.org/10.12795/rea.2016.i33.01>



Esta obra se distribuye con la licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional

Cáceres, N. (2012): "Traffic Flow Estimation Models Using Cellular Phone Data", *IEEE Transactions on Intelligent Transportation Systems*, pages 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, 15–26.
<http://dx.doi.org/10.1049/iet-its:20060020>

Chershire, J. and Uberti, O. (2014): London: The Information Capital. 100 Maps and Graphics that Will Change How You View the City. Penguin Group. London.

Ciuccarelli, P., Lupi, G., & Simeone, L. (2014): Visualizing the Data City. (pp. 17-22). Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-02195-9_3

Conover, M. D., Davis, C., Ferrara, E., McKelvey, K., Menczer, F. y Flammini, A. (2013): "The geospatial characteristics of a social movement communication network", *PloS One*, 8(3), e55957. <http://dx.doi.org/10.1371/journal.pone.0055957>

De Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R. y Yu, C. (2010): Automatic construction of travel itineraries using social breadcrumbs [Electronic Version], http://research.microsoft.com/en-us/um/people/munmund/pubs/ht_10_long.pdf

De Domenico, M., Lima, A. y Musolesi, M. (2013): "Interdependence and predictability of human mobility and social interactions", *Pervasive and Mobile Computing*, 9(6), 798-807. <http://dx.doi.org/10.1016/j.pmcj.2013.07.008>

Eagle, N., Pentland, A. y Lazer, D. (2009): Inferring friendship network structure by using mobile phone data", *PNAS*, 9 (36): 15274-15278.
<http://dx.doi.org/10.1073/pnas.0900282106>

Ferreras Rodríguez, E. M. (2011): "El movimiento 15 m y su evolución en Twitter", *Telos, Cuadernos de Comunicación e Innovación*, nº 89.

Frías-Martínez, V., Soto, V., Hohwald, H., y Frías-Martínez, E. (2012): "Characterizing urban landscapes using geolocated tweets", in *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>

García-Palomares, J.C., Gutiérrez, J. y Mínguez, C. (2015): "Identification of tourist hot spots based on social networks: a comparative analysis of European metropolises using photo-sharing services and GIS", *Applied Geography*, 63, 408–417.
<http://dx.doi.org/10.1016/j.apgeog.2015.08.002>

Revista de Estudios Andaluces, vol. 33, núm. 1 (2016) pp. 1-23. e-ISSN: 2340-2776
<http://dx.doi.org/10.12795/rea.2016.i33.01>



Esta obra se distribuye con la licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional

Gavric, K. D., Culibrk, D. R., Lugonja, P. I., Mirkovic, M. R. y Crnojevic, V. S. (2011): "Detecting attractive locations and tourists' dynamics using geo-referenced images", in 2011 10th International Conference on Telecommunication in Modern Satellite Cable and Broadcasting Services (TELSIKS) (pp. 208–211), Belgrade, Oct 5–8.

<http://dx.doi.org/10.1109/TELSIKS.2011.6112035>

Graham, M., & Shelton, T. (2013): "Geography and the future of big data, big data and the future of geography", *Dialogues in Human Geography*, 3(3), 255-261.

<http://dx.doi.org/10.1177/2043820613513121>

Guttentag, D. (2013): "Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector", *Current Issues in Tourism*, 18, 1192-1217.

<http://dx.doi.org/10.1080/13683500.2013.827159>

Hadas, Y. (2013): "Assessing public transport systems connectivity based on Google Transit data", *Journal of Transport Geography*, 33, 105–116.

<http://dx.doi.org/10.1016/j.jtrangeo.2013.09.015>

Heerschap, N., Ortega, S., Priem, A. y Offermans, M. (2014, May): "Innovation of tourism statistics through the use of new big data sources", in 12th Global Forum on Tourism Statistics, Prague, CZ.

http://tsf2014prague.cz/assets/downloads/Paper%201.2_Nicolaes%20Heerschap_NL.pdf

Kachkaev, A. & Wood, J. (2013): "Investigating Spatial Patterns in User-Generated Photographic Datasets by Means of Interactive Visual Analytics", Paper presented at the GeoViz Hamburg: Interactive Maps that Help People Think, 6 - 8 Mar 2013, HafenCity University, Hamburg, Germany

http://openaccess.city.ac.uk/2829/1/kachkaev_geovizhamburg_2013-final.pdf

Kisilevich, S., Krstajic, M., Keim, D., Andrienko, N., & Andrienko, G. (2010): "Event-based analysis of people's activities and behavior using Flickr and Panoramio geotagged photo collections", in 2010 14th International Conference Information Visualisation (pp. 289-296). IEEE. <http://dx.doi.org/10.1109/iv.2010.94>

Kitchin, R. (2013): "Big Data and human geography Opportunities, challenges and risks", *Dialogues in Human Geography*, 3(3), 262-267.

<http://dx.doi.org/10.1177/2043820613513388>

Koerbitz, W., Önder, I., y Hubmann-Haidvogel, A. C. (2013): *Identifying Tourist Dispersion in Austria by Digital Footprints* (pp. 495-506). Springer Berlin Heidelberg.

Revista de Estudios Andaluces, vol. 33, núm. 1 (2016) pp. 1-23. e-ISSN: 2340-2776

<http://dx.doi.org/10.12795/rea.2016.i33.01>



Esta obra se distribuye con la licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional

http://dx.doi.org/10.1007/978-3-642-36309-2_42

Krumme, C., Llorente, A., Cebrian, M., y Moro, E. (2013): "The predictability of consumer visitation patterns", *Scientific Reports*, 3.
<http://dx.doi.org/10.1038/srep01645>

Kurashima, T., Iwata, T., Irie, G. y Fujimura, K. (2013): "Travel route recommendation using geotagged photos", *Knowledge and Information Systems*, 37(1), 37–60.
<http://link.springer.com/article/10.1007/s10115-012-0580-z>
<http://dx.doi.org/10.1007/s10115-012-0580-z>

Lenormand, M.; Picornell, M.; Cantú-Ros, O.; Tugores, A.; Louail, T.; Herranz, R.; Barthelemy, M.; Frías-Martínez, E. y Ramasco, J.J. (2014): "Cross-checking different sources of mobility information", *PLoS ONE*, 9, e105184.
<http://dx.doi.org/10.1371/journal.pone.0105184>

Lenormand, M., Louail, T., Cantú-Ros, O. G., Picornell, M., Herranz, R., Arias, J. M. y Ramasco, J. J. (2015): "Influence of sociodemographics on human mobility", *Scientific Reports*, 5. <http://dx.doi.org/10.1038/srep10075>

Llorente, A., García-Herranz, M., Cebrián, M., & Moro, E. (2015): "Social media fingerprints of unemployment", *PLoS One*, 10(5), e0128692.
<http://dx.doi.org/10.1371/journal.pone.0128692>

Louail, T.; Lenormand, M.; García Cantú, O.; Picornell, M.; Herranz, R.; Frías-Martínez, E.; Ramasco, J.J. y Barthélemy, M. (2014): "From mobile phone data to the spatial structure of cities", *Scientific Reports* 4, 5276. <http://dx.doi.org/10.1038/srep05276>

Li, X. (2013): "Multi-day and multi-stay travel planning using geo-tagged photos", in Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information (pp. 1-8). ACM.
<http://dx.doi.org/10.1145/2534732.2534733>

Lu, X., Wang, C., Yang, J. M., Pang, Y. y Zhang, L. (2010): "Photo2trip: generating travel routes from geo-tagged photos for trip planning", in Proceedings of the international conference on Multimedia (pp. 143–152). ACM.
<http://dx.doi.org/10.1145/1873951.1873972>

Mocanu, D., Baronchelli, A., Perra, N., Gonçalves, B., Zhang, Q. y Vespignani, A. (2013): "The twitter of babel: Mapping world languages through microblogging platforms", *PLoS One*, 8(4), e61981. <http://dx.doi.org/10.1371/journal.pone.0061981>

Revista de Estudios Andaluces, vol. 33, núm. 1 (2016) pp. 1-23. e-ISSN: 2340-2776
<http://dx.doi.org/10.12795/rea.2016.i33.01>



Esta obra se distribuye con la licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional

Moya-Gómez, B. and García-Palomares, J.C. (2015): "Working with the daily variation in infrastructure performance. The cases of Madrid and Barcelona", *European Transport Research Review* 7(2), 20, 1-13.

Munizaga, M., Palma, C. y Mora, P. 2010: "Public transport OD matrix estimation from smart card payment system data", in Proceedings from 12th World Conference on Transport Research, Lisbon, Paper (No. 2988).

Murthy, D. (2013): *Twitter: Social communication in the Twitter age*. John Wiley & Sons.

Netto, V. M., Pinheiro, M., Meirelles, J. V. y Leite, H. (2015): Digital footprints in the cityscape. International Conference on Social Networks, Athens, USA.

Neutens, T., Schwanen, T., Witlox, F. y De Maeyer, P. (2010): "Evaluating the temporal organization of public service provision using space-time accessibility analysis", *Urban Geography*, 31(8), 1039-1064. <http://dx.doi.org/10.2747/0272-3638.31.8.1039>

Nolasco-Cirugeda, A. y García Mayor, C. (2014): Aplicación de los indicadores de complejidad urbana a través de las redes sociales y TIG: El caso de los paseos marítimos de Levante y Poniente en Benidorm. Alicante, XVI Congreso Nacional de Tecnologías de la Información Geográfica, 25, 26 y 27 de Junio de 2014.

O'Connor, P. (2008). "User-generated content and travel: A case study on Tripadvisor. Com", *Information and Communication Technologies in Tourism*, 47-58.

O'Connor, P. (2010): "Managing a hotel's image on TripAdvisor", *Journal of Hospitality Marketing & Management*, 19(7), 754-772. <http://dx.doi.org/10.1080/19368623.2010.508007>

Pastor-Escuredo, D., Morales-Guzmán, A., Torres-Fernández, Y., Bauer, J. M., Wadhwa, A., Castro-Correa, C., y Luengo-Oroz, M. (2014): "Flooding through the lens of mobile phone activity", in Global Humanitarian Technology Conference (GHTC), 2014 IEEE (pp. 279-286). IEEE. <http://dx.doi.org/10.1109/GHTC.2014.6970293>

Pelletier, M. P., Trépanier, M. y Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4), 557-568. <http://dx.doi.org/10.1016/j.trc.2010.12.003>

Popescu, A., Grefenstette, G. y Moellic, P.-A. (2009): Mining Tourist Information from User-Supplied Collections. Paper presented at the Conference on Information and Knowledge Management. <http://comupedia.org/adrian/articles/sp0668-popescu.pdf>

Revista de Estudios Andaluces, vol. 33, núm. 1 (2016) pp. 1-23. e-ISSN: 2340-2776
<http://dx.doi.org/10.12795/rea.2016.i33.01>



Esta obra se distribuye con la licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional

<http://dx.doi.org/10.1145/1645953.1646211>

Ratti, C., D. Frenchman, R.M. Pulselli y S. Williams (2006): "Mobile Landscapes: using location data from cell phones for urban analysis", *Environment and Planning B: Planning and Design*, 33: 727-748. <http://dx.doi.org/10.1068/b32047>

Reades, J., F. Calabrese y C. Ratti, Eigenplaces (2009): "Analyzing Cities Using the Space-time Structure of the Mobile Phone Network", *Environment & Planning B*, 36, 824-836. <http://dx.doi.org/10.1068/b34133t>

Sakaki, T., Okazaki, M. y Matsuo, Y. (2010): "Earthquake shakes Twitter users: real-time event detection by social sensors", in Proceedings of the 19th international conference on World Wide Web (pp. 851-860). ACM. <http://dx.doi.org/10.1145/1772690.1772777>

Seegerberg, A. y Bennett, W. L. (2011): "Social media and the organization of collective action: Using Twitter to explore the ecologies of two climate change protests", *The Communication Review*, 14(3), 197-215. <http://dx.doi.org/10.1080/10714421.2011.597250>

Shelton, T., Poorthuis, A., & Zook, M. (2015): "Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information", *Landscape and Urban Planning*, 142, 198-211. <http://dx.doi.org/10.1016/j.landurbplan.2015.02.020>

Serrano-Estrada, L., Serrano Salazara, S. y Álvarez Álvarez. F.J. (2014): Las redes sociales y los SIG como herramientas para conocer las preferencias sociales en las ciudades turísticas: el caso de Benidorm. Alicante, XVI Congreso Nacional de Tecnologías de la Información Geográfica, 25, 26 y 27 de Junio de 2014.

Sobolevsky S, Sitko I, Grauwin S, Des Combes RT, Hawelka B, et al. (2014a): "Mining Urban Performance: Scale-Independent Classification of Cities based on Individual Economic Transactions", in: Big Data Science and Computing, 2014 ASE International Conference on, May 27–31, Stanford University. p. 10.

Sobolevsky S, Sitko I, Tachet des Combes R, Hawelka B, Murillo Arias J, et al. (2014b): "Money on the Move: Big Data of Bank Card Transactions as the New Proxy for Human Mobility Patterns and Regional Delineation. The Case of Residents and Foreign Visitors in Spain", in: Big Data (BigData Congress), 2014 IEEE International Congress on, Jun 27-Jul 2, Anchorage, AK. pp.136–143. doi:10.1109/BigData.Congress.2014.28. <http://dx.doi.org/10.1109/BigData.Congress.2014.28>

Sobolevsky S, Bojic I, Belyi A, Sitko I, Hawelka B, et al. (2015a): "Scaling of city attractiveness for foreign visitors through big data of human economical and social media activity", *arXiv:150406003*.

Revista de Estudios Andaluces, vol. 33, núm. 1 (2016) pp. 1-23. e-ISSN: 2340-2776 <http://dx.doi.org/10.12795/rea.2016.i33.01>



Esta obra se distribuye con la licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional

Sobolevsky, S., Sitko, I., Combes, R. T. D., Hawelka, B., Arias, J. M., & Ratti, C. (2015b): "Cities through the Prism of People's Spending Behavior", *arXiv preprint arXiv:1505.03854*.

Straumann, R. K., Çöltekin, A., & Andrienko, G. (2014): "Towards (Re) Constructing Narratives from Georeferenced Photographs through Visual Analytics", *The Cartographic Journal*, 51(2), 152–165.
<http://dx.doi.org/10.1179/1743277414Y.0000000079>

Sun, Y., & Fan, H. (2014): "Event Identification from Georeferenced Images", in *Connecting a Digital Europe through Location and Place*. (pp. 73-88). Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-03611-3_5

Takhteyev, Y., Gruzd, A., & Wellman, B. (2012): "Geography of Twitter networks", *Social Networks*, 34(1), 73-81. <http://dx.doi.org/10.1016/j.socnet.2011.05.006>

Tao, S., Rohde, D., & Corcoran, J. (2014): "Examining the spatial-temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap", *Journal of Transport Geography*, 41, 21-36. <http://dx.doi.org/10.1016/j.jtrangeo.2014.08.006>

Wu, L., Zhi, Y., Sui, Z., & Liu, Y. (2014): "Intra-urban human mobility and activity transition: evidence from social media check-in data", *PloS One*, 9(5), e97010.
<http://dx.doi.org/10.1371/journal.pone.0097010>

Zielstra, D., & Hochmair, H. H. (2011): "Comparative Study of Pedestrian Accessibility to Transit Stations Using Free and Proprietary Network Data", *Transportation Research Record: Journal of the Transportation Research Board*, 2217(1), 145-152.
<http://dx.doi.org/10.3141/2217-18>

Zhai, S., Xu, X., Yang, L., Zhou, M., Zhang, L. y Qiu, B. (2015): "Mapping the popularity of urban restaurants using social media data", *Applied Geography*, 63, 113-120.
<http://dx.doi.org/10.1016/j.apgeog.2015.06.006>

