

XVIII CONGRESO NACIONAL TIG

APLICACIONES DE LAS TECNOLOGÍAS DE LA INFORMACIÓN
GEOGRÁFICA (TIG) PARA EL DESARROLLO ECONÓMICO SOSTENIBLE

Málaga, 29, 30 de junio y 1 de julio de 2016
Asociación de Geógrafos Españoles (AGE)
Grupo de Tecnologías de la Información Geográfica
www.congresotig2016.com



DEPARTAMENTO DE GEOGRAFÍA



Aplicaciones geotecnológicas para el desarrollo económico sostenible

XVII Congreso Nacional de Tecnologías de la Información Geográfica

Créditos

©Edición Federico Benjamín Galacho Jiménez
 Jesús Vías Martínez
 Sergio Reyes Corredera

© Grupo de Tecnologías de la Información Geográfica. Asociación de Geógrafos Españoles (AGE) y Departamento de Geografía, European Topic Centre (ETC) y Facultad de Turismo de la Universidad de Málaga. Campus de Teatinos, 29071 Málaga (España).

ISBN- 13: 978-84-940784-9-1

Año de edición 2016

Maquetación www.everywhere.eu

Reservados todos los derechos. Ni la totalidad ni parte de esta edición puede reproducirse o transmitirse por ningún procedimiento electrónico o mecánico, incluyendo fotocopia, grabación magnética o cualquier almacenamiento de información y sistema de recuperación, sin permiso de la entidad editora.

PATRONES ESPACIALES DE CONCENTRACIÓN DE TURISTAS EN MADRID A PARTIR DE DATOS GEOLOCALIZADOS DE REDES SOCIALES: PANORAMIO Y TWITTER

Juan Carlos García Palomares¹, Javier Gutiérrez Puebla¹, Gustavo Romanillos Arroyo¹, Henar Salas-Olmedo¹

¹ Universidad Complutense, Departamento de Geografía Humana, C/Profesor Aranguren, s/n, 28040 Madrid
jcgarcia@ghis.ucm.es, javiergutierrez@ghis.ucm.es, gustavro@ucm.es, mariahenar.salas@pdi.ucm.es

RESUMEN

El comportamiento espacial de los turistas urbanos es poco conocido. Sin embargo, los turistas generan una enorme cantidad de datos cuando visitan las ciudades y estas fuentes de datos permiten seguir sus actividades. Este trabajo tiene como objetivo seguir la huella digital de los turistas urbanos de nuevas fuentes de datos, asociadas a las redes sociales. Se utilizan dos fuentes: una comunidad de fotografías compartidas (*Panoramio*) y una red social (*Twitter*). La comparación entre la densidad de turistas según las dos fuentes de datos se realiza a partir de mapas, análisis de correlación (OLS) y análisis de autocorrelación espacial (Global Moran's I statistic y LISA). Finalmente, la integración de los datos se lleva a cabo a partir de la combinación de los clusters espaciales identificados en el análisis LISA para las dos fuentes de datos. Los resultados muestran que los datos aportados por las dos fuentes son en parte redundantes y en parte complementarios espacialmente, de forma que es posible caracterizar espacios turísticos polivalentes (cubiertos por las dos fuentes) y espacios especializados en una actividad. En el caso de estudio analizado (Madrid) se observa una fuerte presencia de turistas en el centro y una creciente especialización del centro a la periferia. La principal conclusión del trabajo es que para el seguimiento de los turistas en las ciudades no basta con utilizar una fuente de datos, sino que es necesario utilizar varias de forma complementaria.

Palabras clave: turismo urbano; servicios de fotografías compartidas; redes sociales, Twitter; análisis espacial, SIG

ABSTRACT

The spatial behavior of urban tourists is not well known. However, tourists generate an enormous amount of data when they visit cities and these data sources allow us to follow their activities. This paper aims to follow the footprint of urban tourists through social networks. Two data sources are used: a community of shared photographs (*Panoramio*) and a social network (*Twitter*). The comparison between the density of tourists according to these two data sources is performed using density maps, regression analysis (OLS) and spatial autocorrelation analysis (Global Moran's I statistic and LISA). Finally, results obtained by the LISA analysis are integrated. The results show that the data provided by the two sources are partially redundant and partially complementary, so that it is possible to characterize multipurpose tourist areas (covered by the two sources) and areas specialized (covered by one source). In the analyzed case study (Madrid), it is observed a strong presence of tourists in the center and increasing specialization from the center to the periphery. The main conclusion of the study is that in order to monitor the presence of tourists in cities, it is necessary to use complementary data sources.

Keywords: Urban tourism; Photo-sharing services, Social networks; Spatial analysis, GIS

1. INTRODUCCIÓN

Uno de los campos en los que las nuevas fuentes de datos asociadas a las redes sociales y el Big Data ofrecen mayores oportunidades es el turismo. Las fuentes de datos oficiales no proveen información detallada sobre los lugares que visitan los turistas en las ciudades. En el mejor de los casos proveen información derivada de

encuestas o de registros de hoteles o de venta de entradas a museos. Sin embargo los turistas dejan una huella digital. Así, por ejemplo, toman multitud de fotografías y las suben a los servicios de fotografías compartidas, realizan pagos con tarjeta bancarias, hablan y envían mensajes por sus teléfonos móviles, participan en las redes sociales, etc. Toda esa actividad produce una enorme cantidad de datos digitales, gran parte de ellos geolocalizados, que pueden ser analizados para estudiar patrones de comportamiento. Sin embargo pocos trabajos han utilizado esta nueva información para analizar la distribución espacial de los turistas en las ciudades. Ello se debe probablemente a la novedad de estas fuentes de datos y a la dificultad para acceder a algunas de ellas.

El objetivo principal de esta comunicación es comparar dos nuevas fuentes de datos geolocalizados, procedentes de redes sociales, para el seguimiento de los turistas en las ciudades considerando diferentes la huella digital que dejan al hacer y colgar fotografías (*Panoramio*) y al conectarse a Internet para participar en las redes sociales (*Twitter*). Se toma como área de estudio la ciudad de Madrid, uno de las ciudades europeas con mayor cantidad de turistas. Para su análisis, los datos han sido tratados en un Sistema de Información Geográfica (SIG) y se han usado técnicas de estadística espacial.

Existen varias redes sociales que comparten fotografías, como Instagram, Flickr y Panoramio. Las tres proporcionan la posibilidad de geolocalizar fotografías, pero Panoramio (<http://www.panoramio.com>) pone mayor énfasis en la georreferenciación de las fotografías compartidas por los usuarios, ya que es un sitio web dedicado a mostrar fotografías de lugares o paisajes tomadas por los usuarios. Estas imágenes pueden ser vistas en la propia web de Panoramio o a través de Google Earth y Google Maps. De hecho Panoramio es un servicio de Google, que contiene más de 120 millones de fotografías geolocalizadas. Los servicios de fotografías geolocalizadas han sido utilizados en el campo del turismo con varios fines, como identificar eventos sociales, como festivales, manifestaciones, acontecimientos deportivos, etc. (Sun y Fan, 2014), estimar el número de turistas (Koerbitz et al., 2013), identificar las trayectorias más comunes que siguen los turistas (Girardin et al., 2008), proponer o evaluar rutas turísticas (Kurashima et al., 2013), sugerir viajes turísticos (Lu et al., 2010) y planificar viajes de varios días y lugares a visitar (Li, 2013). Además estas fuentes de datos permiten identificar la localización de los turistas a partir de las fotografías que toman (Kisilevich et al., 2013; Straumann et al., 2014). Estos datos pueden ser tratados con técnicas de estadística espacial para analizar los puntos calientes y medir la intensidad de uso de los mismos (García-Palomares et al., 2015).

Las redes sociales también proporcionan datos de gran interés para el estudio del turismo, si bien los trabajos que los han utilizado en este campo son poco numerosos. La mayor parte de los estudios realizados con datos masivos de redes sociales han usado Twitter (Murthy, 2013), no solo por tratarse de una plataforma de cobertura global, 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. Cada tweet geolocalizado deja una "huella" digital del lugar y el momento en que fue enviado. 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 la actividad en las redes sociales 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 y otros (2014)). Se han utilizado tweets geolocalizados también 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 (Mocanue et al., 2013). Los trabajos que utilizan tweets geolocalizados en el campo del turismo son muy escasos y han estado centrados en comparar el comportamiento espacial de los turistas a escala nacional o global (Bassolas et al., 2016; Hawelka et al., 2014; Sobolevsky et al., 2015) pero no a escala intraurbana.

2. HERRAMIENTAS INFORMÁTICAS Y DATOS

Desde la API de Panoramio fueron descargados los registros de todas las fotografías localizadas en el municipio de Madrid cubriendo el periodo temporal 2005-2014, generando ficheros "csv". La base de datos contiene los siguientes campos: latitud, longitud, identificador de usuario, un link a la url, y la fecha en que fue subida la fotografía (día, mes y año). A partir de las coordenadas geográficas se generó una capa de puntos en ArcGIS 10.3. Los registros no contienen información sobre la nacionalidad del usuario. Para identificar si la fotografía fue tomada por un residente o un turista se utilizó el periodo temporal en que ese usuario tomaba fotografías: si era inferior a un mes, el usuario era considerado turista y en caso contrario residente. Ante la falta más información que permita diferenciar entre turista y residentes en las dos fuentes, pensamos que este periodo de un mes puede ser una buena aproximación sobre todo para asegurar no incluir entre los turistas a usuarios que son claramente residentes. De hecho, esta metodología es semejante a la utilizada por Fischer en su *Geotaggers' World Atlas* y García Palomares et al. (2015) con resultados en ambos trabajos claramente satisfactorios. El resultado de este proceso fue la asignación de 52,513 fotografías a turistas.

De la API de Twitter se descargaron tweets geolocalizados enviados desde Madrid a lo largo del año 2013, generando un fichero "csv". La base de datos contiene información sobre latitud, longitud, identificador de usuario, lengua en que está configurado el dispositivo, fecha en que fue enviado el tweet (minuto, hora, día, mes y año) y texto del mensaje. Estos ficheros fueron tratados de la misma forma que los de Panoramio, primero para generar una capa de tweets con ArcGIS usando las coordenadas de los registros y después para diferenciar los tweets de turistas y residentes en función del periodo en que los tuiteros han tuiteado desde Madrid. Se trató de afinar la definición de turista con la lengua del tweet, pero los resultados no fueron satisfactorios (se deja fuera, por ejemplo, todos los turistas de otras comunidades autónomas y los hispanohablantes). Resultó así que de un total de 9.257.561 tweets, 264.462 fueron atribuidos a turistas.

Aunque las fechas de las dos fuentes son diferentes, hemos optado por trabajar con todas las fotografías descargadas para el periodo temporal 2005-2014, tratando así de tener un número muy alto de fotografías y no solo el año 2013. Para las dos fuentes, en el SIG y para la aplicación del análisis estadístico, los datos individuales (a nivel puntual) han sido agrupados a nivel de sección censal. Todos los cálculos y mapas fueron realizados usando el software ArcGIS 10.3.

3. MÉTODOS

Para analizar la distribución espacial de los turistas con ambas fuentes de datos se siguió la siguiente metodología (Figura 1):

- 1) Número de turistas por sección censal.- Para hacer comparables las dos fuentes de datos, se procedió a localizar las fotos y los tweets de turistas según secciones censales, contabilizando el número de turistas únicos en cada sección censal según cada fuente a partir del identificador de usuario.
- 2) Densidad de turistas por sección censal.- El número de turistas en cada sección censal depende de la concentración real de turistas en esa sección censal y del tamaño de la sección censal. Secciones censales más grandes tienden a registrar un mayor número de turistas. Para mitigar este problema (unidad espacial modificable) se obtuvo la densidad de turistas por sección censal para cada una de las fuentes de datos.
- 3) Reescalamiento de los datos.- Los datos de densidad de turistas fueron reescalados a una escala de 0 a 1,000 mediante una transformación lineal para hacerlos comparables.
- 4) Mapas de densidad de turistas y estadísticas descriptivas.- A partir de los datos reescalados se elaboraron mapas de densidad de turistas a nivel de sección censales y se calcularon estadísticas descriptivas.
- 5) Análisis de regresión.- El análisis de regresión bivalente se utilizó para comparar ambas fuentes de datos a partir del coeficiente de determinación (que permite conocer la parte común de variación entre ambas fuentes de datos) y los mapas de residuos estandarizados (que permiten analizar las diferencias entre ambas fuentes).

6) Análisis de autocorrelación espacial.- A diferencia de los análisis anteriores, las técnicas de autocorrelación espacial consideran cada localización no de forma aislada, sino en relación a las localizaciones de su entorno (Anselin, 1995). Se calculó el índice global de Moran y se realizó un análisis LISA (Anselin Local Moran's I) para identificar clusters espaciales, utilizando el inverso de la distancia con un radio de 500 m. Los resultados obtenidos con las dos fuentes fueron combinados para conocer la especialización turística de cada sección censal.



Figura 1. Metodología seguida

4. RESULTADOS

4.1 Mapas de densidad de turistas

La figura 2 muestra mapas de densidades de turistas según secciones censales, con datos reescalados a una escala de 0 a 1.000 y usando los mismos intervalos en los dos mapas. El análisis de los datos extraídos de Panoramio (figura 2a) muestra una clara concentración en el centro histórico y a lo largo del principal eje norte-sur de la ciudad (Paseo de La Castellana). Las áreas de mayor densidad reflejan los espacios más visitados de la ciudad, por ejemplo Plaza de Cibeles, Puerta de Alcalá, Puerta del Sol, Plaza Mayor, Palacio Real, Templo de Debod, Plaza de España, Museo Reina Sofía, Estación de Atocha, Gran Vía, los estadios del Real Madrid y Atlético de Madrid, la Plaza de Toros de Las Ventas, las Torres Kio, o Cuatro Torres, etc. Algunas secciones censales tienen una alta densidad de fotos no porque contengan un punto de atracción turística en su interior, sino porque desde ellas se pueden tomar fotografías de un elemento localizado en una sección censal adyacente. La prohibición de hacer fotos en el interior de algunos monumentos explica la relativamente poca densidad de fotos en las secciones censales que contienen algunos de ellos (por ejemplo, Museo del Prado y Palacio Real).

La densidad de turistas identificados según Twitter (figura 2b) es especialmente alta en el centro histórico, en el eje del Paseo de la Castellana (sobre todo en el Estadio del Real Madrid y la Estación de Atocha) y en menor medida en el barrio de Salamanca (hoteles y comercios), pero tiende a extenderse por un mayor número de secciones censales.

El coeficiente de variación de los datos reescalados muestra que, tal y como sugieren los mapas, la distribución de Panoramio es más concentrada que la de Twitter (Tabla 1).

Tabla 1. Estadísticas descriptivas de las densidades según secciones censales

	Densidad de turistas/Ha		Densidad de turistas/Ha: datos reescalados	
	Panamio	Twitter	Panamio	Twitter
Máximo:	28,83	28,88	1000,00	1000,00
Suma:	731,14	2080,83	25357,03	53074,24

Media:	0,30	0,86	10,53	22,03
Desviación Estándar:	1,00	1,77	34,55	62,39
Coefficiente de Variación	333,33	205,81	328,11	283,20

Numero de secciones: 2409

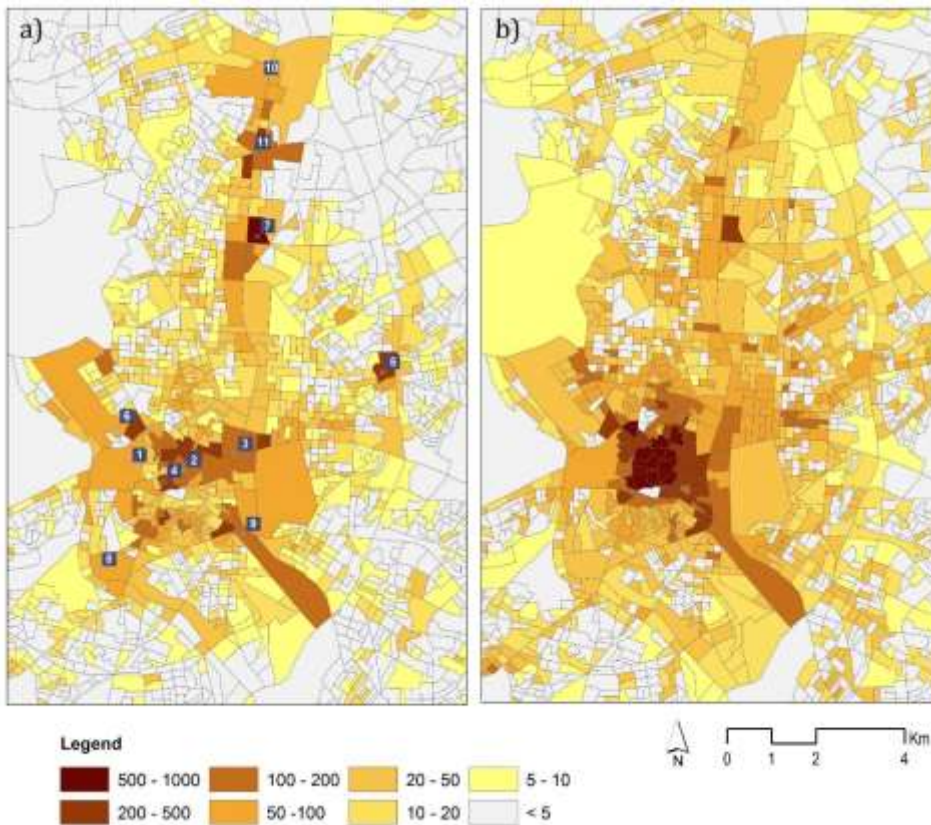


Figura 2. Densidad de turistas según: a) Panoramio, b) Twitter

Referencias: 1 Palacio Real; 2 Puerta del Sol; 3 Plaza de Cibeles; 4 Plaza Mayor; 5 Plaza de Toros; 6 Templo de Debod; 7 Estadio del Real Madrid; 8 Estadio del Atlético de Madrid; 9 Atocha – Museo Reina Sofía; 10 Cuatro Torres; 11 Torres Kio

Para conocer el grado de asociación entre las dos distribuciones se ha calculado el coeficiente de determinación, resultando un valor moderado de 0,28, pero altamente significativo (al nivel de 0,01). Los residuos estandarizados de la regresión muestran los lugares en los que las fuentes de datos presentan mayores diferencias (Figura 3). Twitter ofrece mayor densidad de turistas de lo esperado según Panoramio en el centro histórico y algunos barrios próximos (Barrio de Salamanca), espacios donde hay numerosos alojamientos y restaurantes, pero menos turistas de lo esperado en espacios muy fotografiados fuera del centro histórico: estadios de fútbol (Real Madrid y Atlético de Madrid), Plaza de Toros, Retiro Park, Torres Kio, Cuatro Torres, Templo de Debod.

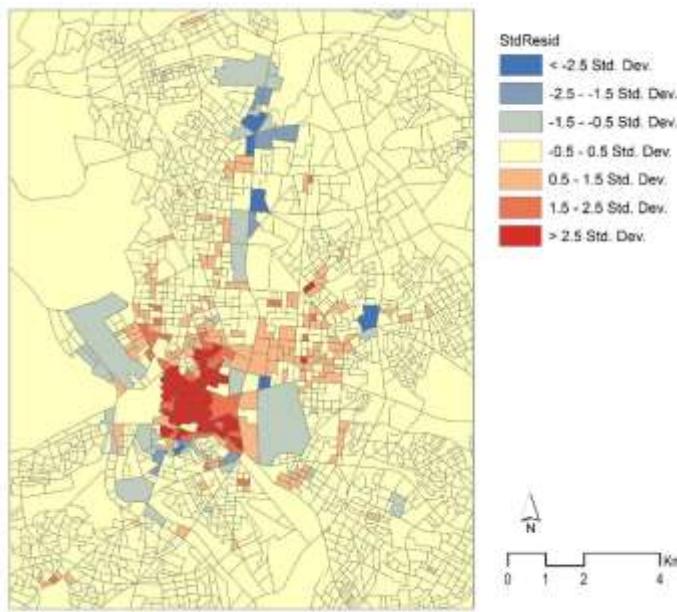


Figura 3. Residuos estandarizados de la regresión bivariada Panoramio-Twitter

4.2 Análisis de autocorrelación espacial

El análisis de autocorrelación espacial permite analizar los datos de cada sección censal en relación a los datos de las secciones censales de su entorno. Usando el inverso de la distancia (IDW) como procedimiento y un radio de distancia de 500 metros, el Índice Global de Moran muestra una autocorrelación espacial positiva significativa en los dos casos, pero mucho más alta en Twitter que en Panoramio (tabla 2). El índice Anselin Local Moran's I (LISA) muestra la distribución de los cluster espaciales (significativos al nivel de 0.01) (figura 4). Las secciones censales con valores Altos/Altos (HH, altos valores en una variable rodeados de altos valores en esa misma variable) tienden a formar un único cluster en el caso de Twitter (centro histórico) (figura 4a) y varios clusteres en las fotografías geocalizadas (centro histórico, estadio del Real Madrid, torres Kio-cuatro torres) (figura 4b).

Para integrar la información de las dos fuentes de datos, se han cruzado los resultados de los análisis univariantes anteriores. La figura 5 muestra una clasificación de secciones censales considerando los dos clusteres HH de las dos fuentes de datos conjuntamente. Así es posible identificar secciones censales que forman parte de los cluster HH en las dos fuentes de datos (centro), o solo en alguna de ellas, por ejemplo áreas especializadas en fotografías (por ejemplo, torres Kio-cuatro torres), en twitter (bordes del centro histórico) o en ambas (centro histórico). Si una sección censal forma parte de algún cluster HH en las dos fuentes de datos eso indicaría que en un radio de 500 m hay una gran densidad de oportunidades para visitar espacios de atractivo turístico o para estar conectados a internet. Por lo tanto, mediante el cruce de las dos fuentes es posible tener una visión LISA multivariante, más completa que la que obtendríamos recurriendo al análisis LISA bivariante. El mapa resultante (figura 5) muestra cómo la especialización turística de las secciones censales tiende a aumentar desde el centro histórico hacia la periferia: las secciones censales del centro con carácter mixto (polifuncional) aparecen rodeadas de otras generalmente con dos actividades y más al suelen estar especializadas en una actividad.

Tabla 2. Global Moran's I statistics (distance threshold = 500m)

	Geocalized photographs	Twitter
Global Moran's Index	0,23	0,58
z-score	22,91	55,82
p-value	0,00	0,00

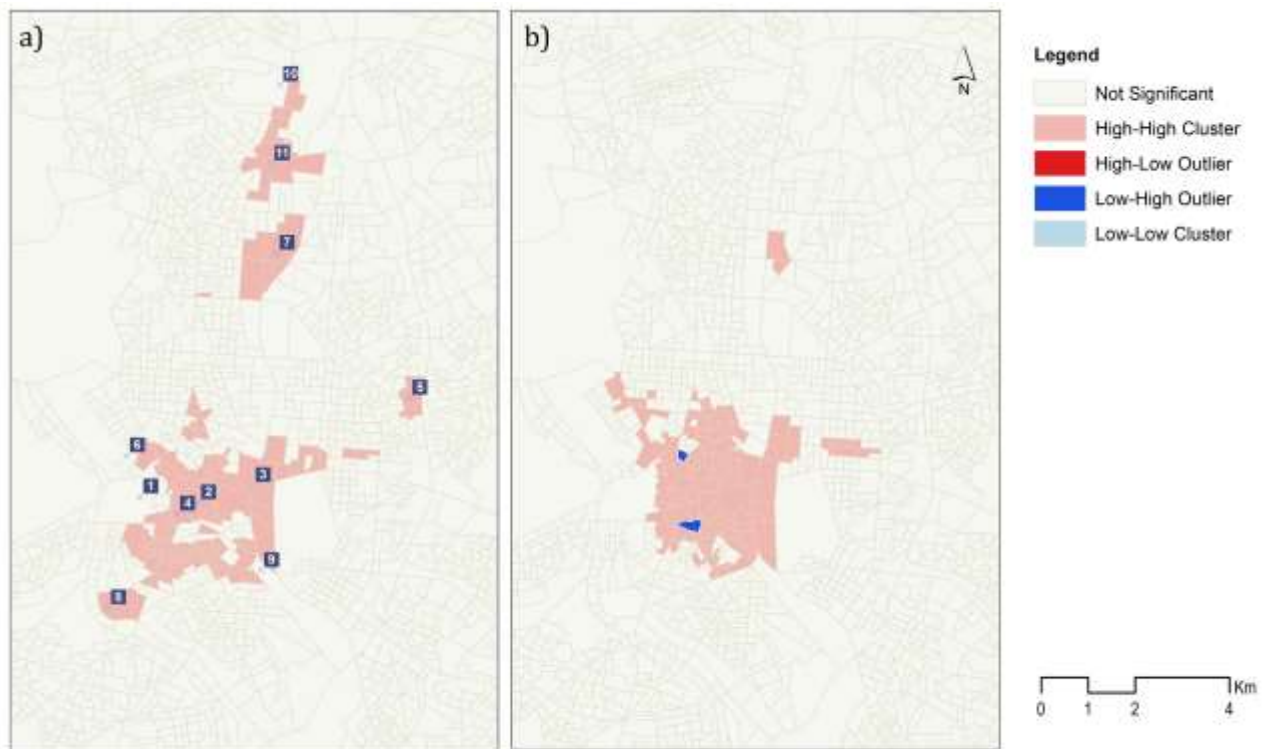


Figura 4. Resultados del análisis LISA (distance threshold = 500m). a) Panoramio, b) Twitter

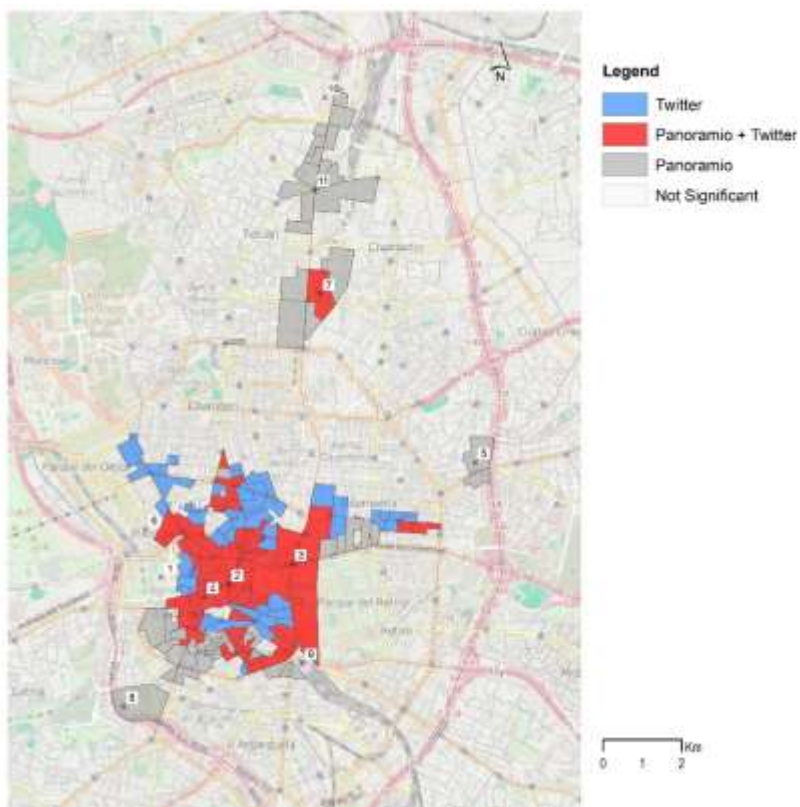


Figura 5. Tipología mediante la combinación de los cluster HH de las dos fuentes de datos

5. CONCLUSIONES

Este trabajo sigue la huella digital de los turistas urbanos, utilizando dos fuentes de datos (Panoramio y Twitter) y tomando como área de estudio la ciudad de Madrid. A diferencia de otros trabajos, aquí no se ha contabilizado la densidad de huellas digitales (por ejemplo, número de fotografías o tweets) sino la densidad de turistas. Esto mitiga problemas de posible sesgo (fotógrafos o tuiteros compulsivos) y hace comparables las dos fuentes de datos. Los resultados demuestran que no es suficiente utilizar una fuente de datos para conocer la distribución espacial de los turistas en las ciudades, ya que los turistas realizan actividades diferentes en espacios diferentes. No es de extrañar que las dos fuentes de datos muestren una alta densidad de turistas en el centro histórico, donde existe una alta concentración de monumentos, comercios, hoteles, restaurantes, etc. Pero la huella digital de los turistas se extiende también por otras áreas de la ciudad, ya se trate de espacios muy fotografiados (Panoramio) o que tienen una alta densidad de establecimientos visitados por turistas (Twitter).

Cuando se utilizan nuevas fuentes de información geográfica asociadas al Big Data y las redes sociales subyace el problema del sesgo. La mayor parte de los turistas no sube sus fotografías a las comunidades de fotografías compartidas como Panoramio e incluso parte de ellos nunca toma fotografías. Además las fotografías no siempre reflejan adecuadamente todos los monumentos de la ciudad, debido a la prohibición de hacer fotos en algunos monumentos, sobre todo en museos. Por otro lado, muchos turistas no utilizan redes sociales como Twitter y solo una pequeña proporción de los usuarios de Twitter envían tuits geolocalizados. Ciertamente el sesgo de cada fuente es difícil de conocer y de corregir. En esta comunicación el sesgo ha sido mitigado trabajando con la densidad de turistas en vez de con la densidad de sus huellas (ya se trate de fotografías o tweets), con lo que se evita contabilizar varias veces al mismo turista en la misma fuente, lo que es especialmente importante en el caso de tuiteros o fotógrafos “compulsivos”. Además la comparación entre fuentes de datos distintas permite considerar diferentes actividades de los turistas y por lo tanto compensar en parte el sesgo derivado de trabajar con solo una de ellas.

En investigaciones futuras incorporaremos nuevas fuentes de datos al estudio comparativo de la huella digital de los turistas para recoger mejor el conjunto de actividades que realizan. De hecho estamos trabajando ya con transacciones con tarjetas bancarias para reflejar las actividades de compras de los turistas, que no pueden ser captadas adecuadamente con Panoramio ni Twitter. Otra futura línea de investigación se orienta a centrar el foco de atención no en la densidad de turistas, sino en el gasto, con el objetivo de diferenciar patrones espaciales del gasto según nacionalidades y de formular modelos explicativos.

6. BILIOGRAFÍA

- Anselin, L. (1995): “Local Indicators of Spatial Association — LISA.” *Geographical Analysis*, 27 (2). Pp. 93–115.
- Bassolas, A., Lenormand, M., Gonçalves, B., Tugores, A. y Ramasco, J.J. (2016). “Touristic site attractiveness seen through Twitter”. *EPJ Data Science* 5.
- Frias-Martinez, V., Soto, V., Hohwald, H., y Frias-Martinez, E. (2012): “Characterizing urban landscapes using geolocated tweets”. In *Privacy, Security, Risk and Trust (PASSAT): International Conference on and 2012 International Confernece on Social Computing (SocialCom)*. pp. 239-248. IEEE.
- 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, pp. 408–417.
- Girardin, F., Fiore, F. D., Ratti, C., y Blat, J. (2008): “Leveraging explicitly disclosed location information to understand tourist dynamics: a case study”. *Journal of Location Based Services*, 2(1), pp. 41-56.
- Hawelka, B., Sitko, I., Beinart, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C. (2014): “Geo-located Twitter as proxy for global mobility patterns”. *Cartography and Geographic Information Science*, 41, pp. 260–271.

- Kisilevich, S., Keim, D., Andrienko, N., y Andrienko, G. (2013): "Towards acquisition of semantics of places and events by multi-perspective analysis of geotagged photo collections". In A. Moore, & Drecki (Eds.): "Geospatial Visualisation, Lecture Notes in Geoinformation and Cartography." Berlin Heidelberg: Springer-Verlag.
- Koerbitz, W., Önder, I., y Hubmann-Haidvogel, A. C. (2013): "Identifying Tourist Dispersion in Austria by Digital Footprints". Springer Berlin Heidelberg. pp. 495-506
- Kurashima, T., Iwata, T., Irie, G. y Fujimura, K. (2013): "Travel route recommendation using geotagged photos". *Knowledge and information systems*, 37(1), pp. 37–60.
- 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.
- 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.
- 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.
- 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.
- Shelton, T., Poorthuis, A., y Zook, M. (2015): "Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information". *Landscape and Urban Planning*, 142, pp. 198-211.
- Sobolevsky S, Bojic I, Belyi A, Sitko I, Hawelka B, et al. (2015): "Scaling of city attractiveness for foreign visitors through big data of human economical and social media activity". *arXiv:150406003*.
- Straumann, R. K., Çöltekin, A., y Andrienko, G. (2014): "Towards (Re)constructing narratives from georeferenced photographs through visual analytics". *The Cartographic Journal*, 51(2), 152-165.
- Sun, Y., y Fan, H. (2014): "Event Identification from Georeferenced Images. In Connecting a Digital Europe through Location and Place". Springer International Publishing, pp. 73-88.
- Wu, L., Zhi, Y., Sui, Z., y Liu, Y. (2014): Intra-urban human mobility and activity transition: evidence from social media check-in data. *PloS one*, 9(5), e97010.

AGRADECIMIENTOS

Agradecemos la financiación de la Unión Europea (proyecto INSIGHT del Séptimo Programa Marco) y de la Comunidad de Madrid (red de grupos de investigación SOCIALBIGDATA-CM S2015/HUM-3427). Agradecemos también la ayuda de Luca Piovano del CEDINT (UPM) en la descarga de los datos de Panoramio.