Understanding Urban Areas of Interest by geo-temporal photo data

Meixu Chen[*1], Dani Arribas-Bel[1], Alex Singleton[1]

[1]Department of Geography and Planning, University of Liverpool

December 5, 2016

KEYWORDS: Social Media, Urban Areas of Interests (AOIs), Flickr, Spatiotemporal Analysis

1. Introduction

The aim of this paper is to develop a method to extract urban AOIs, and understanding them in space and time dimension. The geotagged Flickr photos that fell into Inner London in 2015 will be used for this study. The clustering method DBSCAN is used to capture and cluster AOIs, and time slices on a monthly basis is produced to explore the spatiotemporal evolution of them.

Urban Areas of Interests (AOIs) should be understood as the areas in the urban built environments that are formed by people's interests and behaviours, which can be seen as perceptual or social space, as they integrate the built environment of a city with their digital representation, captured in the popularity of social networks such as Twitter, Facebook, Flickr and others (Hu et al., 2015; Hudson-Smith, 2007). At the same time, understanding urban AOIs are important in multiple application fields. For urban planners, AOIs mean higher priorities when conducting urban planning if only limited resources for allocation (Gandy, 2006). For retailing businesses, urban AOIs can help them pinpointing a large sum of potential customers, contributing to better marketing strategies.

In recent years, social media data such as Twitter and Flickr, offered a novel possibility to understand and measure urban phenomena, these social media services have a large quantity of user groups who generate their individual data every day. In this study, Flickr photo data is selected due to the locations where people take photos are highly related to their interests.

Previous works related to geotagged Flickr photo data can be summarised into two categories: 1) Just focus on spatial distribution of Flickr photos, such as the research about mobility of tourists, or identifying the events with high-density population (e.g. Chen et al., 2009; Sun et al., 2015); 2) In addition to geolocation, time period, semantics or other statistical data is considered to investigate spatiotemporal patterns, people's behaviours or characteristics (e.g. Kisilevich et al.,2010; Kennedy et al., 2007; Li, Goodchild and Xu, 2013,). Although many works have involved discovering Points of Interests (POIs), a few studies moved further into the exploration for urban AOIs. Also, research about the spatiotemporal evolution of urban patterns over time (e.g. months or years) is still rare (Davies, 2015).

2. Data Section

Flickr photo data can be retrieved and downloaded by its public API through python interface (Stüvel, 2016). To ensure the timeliness and quantities of the data, the time span is selected from 1 January 2015 to 31 December 2015. The research extent is narrowed down to Inner London, which was defined in the London Plan Consultation 2009 (London Datastore, 2010). The reasons are that approximately 83% Flickr photos of Greater London located in Inner London, and people are more

*meixu@liverpool.ac.uk

likely to visit and aggregate at this area. As this study focus on the exploration of urban AOIs in space and time rather than the photo itself, only the metadata of Flickr photos were collected, including user ID, geographic coordinates (latitude and longitude) where people took photos, and their taken date. The map below (Figure 1) shows the spatial density distribution of Flickr photos in Inner London.

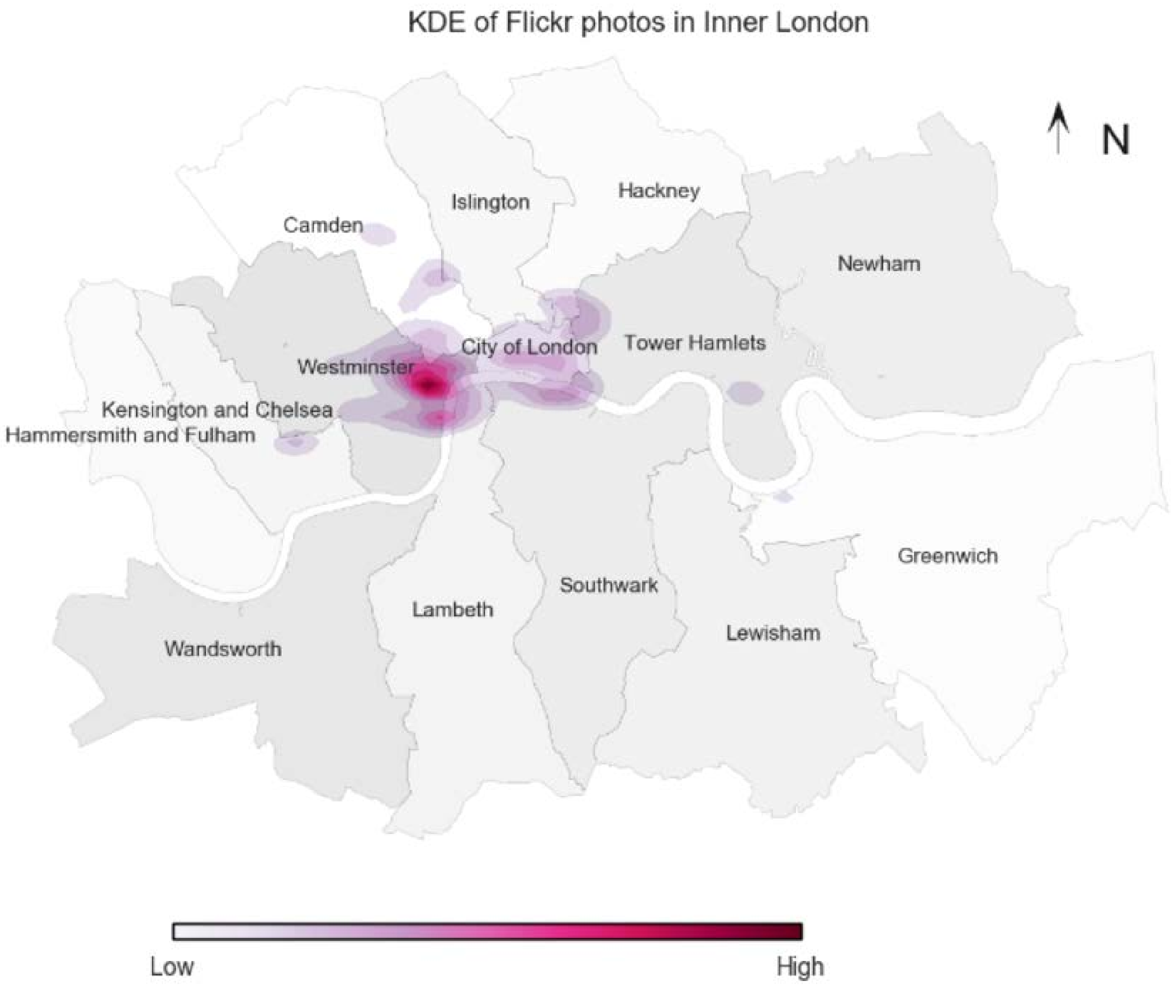

Figure 1 The spatial density distribution of Flickr photos in Inner London by the year 2015

Before extracting and understanding urban AOIs in Inner London, the data need to be pre-processed to build up time series and reduce the data noise. In order to examine the spatiotemporal changes of urban AOIs monthly, a one-year time interval was constructed and have been divided into twelve subsets by month. In addition, the raw Flickr photo data contain large numbers data noise (i.e. the user who has taken his photo at the same location and time), as a result, this set of records should be removed and only one photo remained for each of these users. It can also help avoid AOIs being dominated by those who are active users. Figure 2 shows the number of Flickr users and photos in each month of 2015 after the data pre-processing. It can be seen that the number of Flickr users are relatively stable and dropped slowly throughout the year, while the number of photos fluctuated apparently, with a higher number from March to May and lower number at the last two months. From the statistics perspective, each Flickr user approximately took four photos on average, which is useful for the setting of the parameter in DBSCAN.

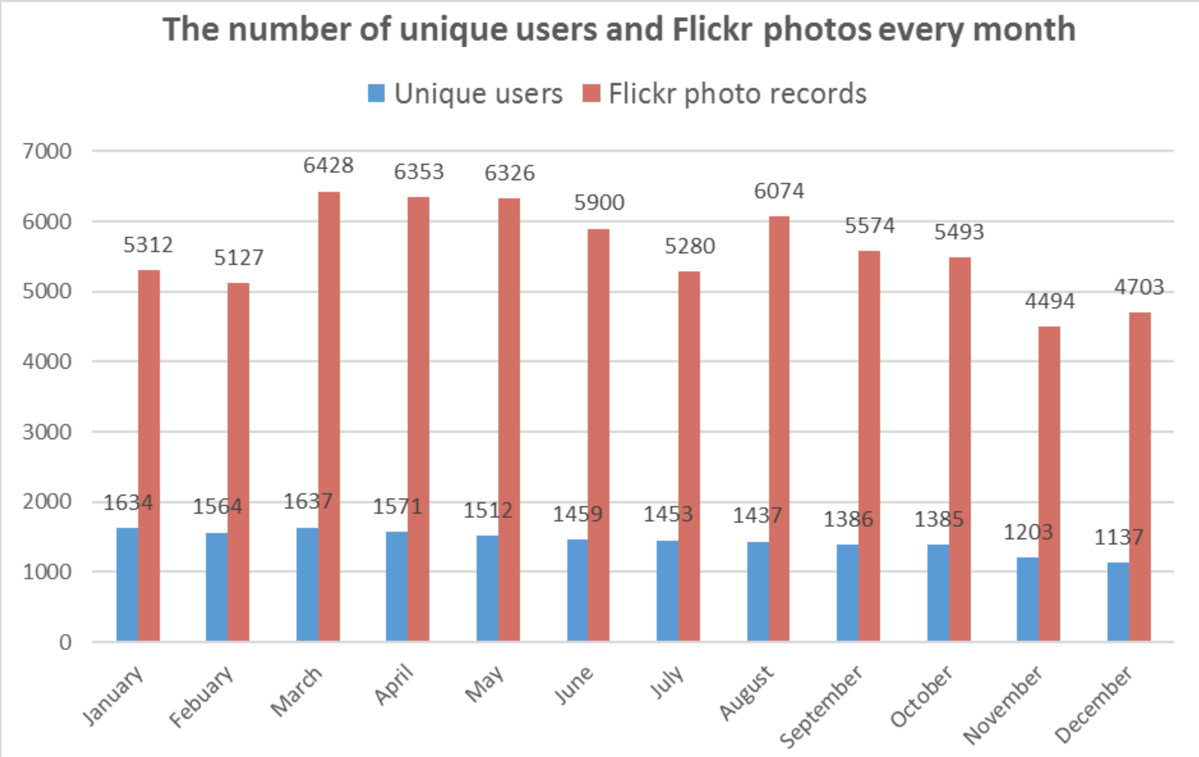

Figure 2 The number of unique users and Flickr photos every month

3. Methods

In this study, a widely-used clustering method DBSCAN is applied to extract urban AOIs. DBSCAN (Density-based spatial clustering for applications with noise) is a density-based clustering method, the core concept is to ensure a minimum number of points contained in a given neighbourhood radius for every cluster (Ester et al., 1996). Compared to other clustering methods, such as k-means clustering, DBSCAN presents several advantages. First, it does not need much-related knowledge to decide the parameters. For this research, it is difficult to pre-determine the number of clusters (i.e. AOI) in various years. Moreover, urban AOIs can be in any shape, DBSCAN has a benefit in identifying clusters with arbitrary shapes (e.g. spherical, linear, drawn-out). Besides, it can process big data efficiently and can filter the outliers or data noises in clusters (Hu et al., 2015).

For the selection of the MinPts parameter, due to the number of Flickr users and photos vary in different months, it is inappropriate to set it as an absolute value. Furthermore, based on the definition of urban AOIs, AOIs are areas where photos have been taken by multiple Flickr users. Therefore, MinPts can be set based on the number of Flickr users. For example, if at least 1% of Flickr users should be ensured in each urban AOI (i.e. each cluster) in each month, the value of MinPts can be set as 4 times of 1% Flickr users explained before. For another parameter Eps, according to Han et al. (2004), k-dist graph based on MinPts can be used to estimate an appropriate value for Eps. After multiple empirical evidence, 4% Flickr photos in each month (i.e. 1% Flickr users) and 250 meters are selected for MinPts and Eps to extract urban AOIs at the neighbourhood level. After setting two important parameters, clusters and noise points are identified.

4. Preliminary Results

Figure 3 The spatiotemporal changes of urban AOIs in the half year of 2015

The preliminary result of extracted AOIs are partly shown in Figure 3. From a spatial perspective, it can be seen that urban AOIs mainly aggregated at four boroughs over the year, including City of London, Westminster, Camden and Tower Hamlets. The most significant AOI is identified in the east of Westminster and the south of Camden by more than 400 Flickr users monthly, which contains Covent Garden, Soho area, St James's Park, The National Gallery, Big Ben, to name only a few. The result corresponds to the empirical knowledge that famous tourist attractions, public entertainment, food and shopping areas are more attractive to people than other areas. City of London is occupied by two extracted AOIs, one extend from St Paul's Cathedral, and another contains numerous important landmarks such as The Monument, Bank of England, Mansion House.

As time is considered, hence, the spatiotemporal changes of urban AOIs can also be investigated. The Figure 3 shows that some urban AOIs showed up and disappeared quickly in just one month, such as the AOI formed around Greenwich Park in March and disappeared in the other months. It is likely because 'The Tough Runs' event held on 28th March 2015 in the park, so that many people took photos in that days. This change implied that the formation of urban AOI is related to some big event. In addition, the evolution in space and time or urban AOIs are also helpful for business marketing. The extracted AOI located in Canary Wharf (a major business zone in Tower Hamlets) is a good example. It only existed for a few months from April to August in 2015, which can be used to deduce that spring and summer are the consumption peak period for Canary Wharf.

5. Conclusion and future work

This paper has proposed an initial framework to extract and understand urban AOIs from Flickr data. DBSCAN clustering method was used to identify AOIs where high-density photos taken from multiple users in a specific neighbourhood distributed. A one-month time interval was constructed to explore the spatiotemporal changes of AOIs.

This work seeks a better understanding of how urban AOIs form and the spatiotemporal evolution over time. It is planned to be extended in the future by investigating semantics in urban AOIs, which aims to further understand what are people's interests in different AOIs. Meanwhile, the clusters are expected to be identified based on both spatial density and photo's semantics.

6. Biography

Meixu Chen is a first-year PhD student under the supervision of Dani Arribas-Bel and Alex Singleton, in Geographic Data Science Lab at University of Liverpool. Her research interest is understanding cities by a series of volunteered geographic image information, includes spatiotemporal analysis, image analysis and machine learning.

Dani Arribas-Bel is a Lecturer in Geographic Data Science at the Department of Geography and Planning , and member of the Geographic Data Science Lab, at the University of Liverpool (UK), where I direct the MSc in Geographic Data Science. He is also part of the development team of the open source library PySAL for spatial analysis in Python. His main research interests are Urban Economics and Regional Science, Spatial Analysis and Spatial Econometrics, and Open source scientific computing.

Alex Singleton is a Professor of Geographic Information Science at the University of Liverpool and Deputy Director of the ESRC Consumer Data Research Centre (CDRC). In a general sense his research is concerned with how the complexities of individual behaviours manifest spatially and can be represented and understood though a framework of geographic data science. In particular, this research has extended a tradition of area classification and he has developed a broad critique of the ways in which geodemographic methods can be refined through modern scientific approaches to data mining, geographic information science and quantitative human geography.

References:

Chen, W. C., Battestini, A., Gelfand, N., and Setlur, V., Visual summaries of popular landmarks from community photo collections. ed. Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on, 2009, 1248-1255.

Davies, J. (2016). The virtual and the physical | Nesta. [online] Nesta.org.uk. Available at: http://www.nesta.org.uk/blog/virtual-and-physical [Accessed 29 Nov. 2016].

Ester, M., Kriegel, H. P., Sander, J., & Xu, X., (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. ed. Kdd, 226-231.

Gandy, M. (2006). Urban nature and the ecological imaginary. In the nature of cities: Urban political ecology and the politics of urban metabolism, 63.

Hudson-Smith, A; (2007) Digital Urban - The Visual City. (CASA Working Paper Series 124). Centre for Advanced Spatial Analysis: London.

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W. and Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. Computers, Environment and Urban Systems, 54, pp.240-254.

Inner and Outer London Boundaries (London Plan Consultation 2009). (2010). London Datastore. [online] Available at: https://data.london.gov.uk/dataset/inner-and-outer-london-boundaries-london-plan-consultation-2009 [Accessed 5 Jan. 2017].

Kennedy, L., Naaman, M., Ahern, S., Nair, R., & Rattenbury, T., 2007. How Flickr helps us make sense of the world: context and content in community-contributed media collections. ed. Proceedings of the 15th international conference on Multimedia, 631-640.

Kisilevich, S., Mansmann, F., Bak, P., Keim, D., & Tchaikin, A., 2010. Where would you go on your next vacation? A framework for visual exploration of attractive places. ed. Advanced Geographic Information Systems, Applications, and Services (GEOPROCESSING), 2010 Second International Conference on, 21-26.

Li, L., Goodchild, M. and Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. Cartography and Geographic Information Science, 40(2), pp.61-77.

Stüvel, S. (2016). Python Flickr API kit - Stüvel Stuff. [online] Stuvel. Available at: https://stuvel.eu/flickrapi [Accessed 5 Aug. 2016].

Sun, Y., Fan, H., Bakillah, M. and Zipf, A. (2015). Road-based travel recommendation using geo-tagged images. Computers, Environment and Urban Systems, 53, pp.110-122.