Wu, Wenjie; Wang, Jianghao

Conference Paper

Exploring city social interaction ties in the big data era: Evidence based on location-based social media data from China

55th Congress of the European Regional Science Association: "World Renaissance: Changing roles for people and places", 25-28 August 2015, Lisbon, Portugal

Provided in Cooperation with:
European Regional Science Association (ERSA)

Exploring city social interaction ties in the big data era: Evidence based on location-based social media data from China

Abstract:

Mountains of public expenditures have been invested on cyberspace infrastructure investments for mapping digital business, social networking and cities with big data. Despite intense public and policy interests, location-based social media data is largely unexplored in conventional urban computational modeling studies due to fundamental challenges in dealing with dynamics of unstructured texts, videos, pictures and web-based big data. This paper explores the geographical ties of urban network patterns by using social media users' mobility flows at a detailed spatial degree (the city-pair level) based on location-based social media (Weibo, a Chinese version of Twitter) data from China. At its heart a geo-tagged computational framework is designed to extract and integrate billions of social media users' space–time footprints for spatiotemporal topology network mining analysis at different geographical scales. A real-time intercity human mobility flow mapping matrix is developed to track spatial patterns of migration flows across regions before and after traditional Chinese Festivals' periods. Our visual exploration results also suggest the significantly heterogeneity in core-periphery urban systems in China, an issue that is highly sensitive in the regional cohesion policy decision making.

Keywords: Big data, Social media, Urban network, China

JEL code: C55, P25

1 Introduction

Location-based social media data is, increasingly, an important facilitator of exploring the movement of goods and people in and between countries across the globe. Typical examples include Twitter, Facebook, Foursquare. Twitter, in particular, has dramatically expanded its user-network worldwide over the past few years. International active Twitter users created approximately 500 million tweets (individual user posts by Twitter) per day in 2013 by using more than 35 languages. As with all social media data outputs, the fundamental value of location-based social media (e.g. twitter and facebook) data is for sensing users' space–time trajectories, and thus, makes social media data a new platform for understanding business and social interactions in the spatial context.

Individual intercity movements (that are reflected from 'geo-tagged' twitter data) are seen as an important way to boost business opportunities and productivity spillover, and inadequate social interactions are seen as an obstacle to the persistence of urban networks. In large developing and emerging economies with massive social media users via computers and mobile phones, real-time 'geo-tagged' human mobility information from social media data sources are clearly potentially large. In these settings, cyberspaces are often built and expanded with the explicit aim of stimulating digital socioeconomic activities and balancing regional disparities. However, despite intense policy and public enthusiasms, there is virtually no direct evidence on exploring the configuration of urban network patterns by using social media users' mobility flows within a large developing country context.

The scarcity of empirical evidence is not surprising, given that mining location-based social media data faces serious identification challenges. First, location-based social media data, as a type of big data resource, are often featured by the dynamic, massive information generated by billions of users across space (Manyika et al., 2011). In truth, despite of the recent development of intensive-computational geographic information system (GIS) modeling programs, social media data with precise individual-level location information is still extremely large to proceed by using the GIS techniques at multiple geographical scales (Wang, 2010; Wright &Wang, 2011). Furthermore, conventional GIS-based computational methods cannot directly read the unstructured social media datasets (e.g. words, pictures, videos). Additional big data mining methods are often needed to transform social media data information from unstructured data formats to structured, and ready-to-use spatial datasets. In this paper, we tackle these problems by analysing the configuration of intercity connection patterns in China to provide new evidence to the applications of location-based social media data in urban and regional studies.

China provides an ideal laboratory for our investigation. First, expansion of the internet infrastructure has been rapid. Since 2006, China embarked on an ambitious programme of civil internet infrastructure investment investing over 1500 billion RMB (1 USD=approximately 6.5 RMB). Today, China is world's largest social media market, carrying approximately 1 billion of tweets (*weibo*) per day in 2013. This rapid expansion means that there are massive new location-based social media users and intercity mobility flows on which we can base our data mining of the geographical ties of urban networks. Second, China is a vast country that consists of political subunits: provinces, prefecture cities, and counties, thus providing ample space for the emergence of urban networks. Our units of analysis are small scale regions (prefecture cities). Third, China has undergone a process of marketization during the past two decades, with the substantial relaxation of labor mobility policies since the 1990s. Millions of people in China moved across cities for leisure, business and social interactions. By applying big data mining techniques, 'geo-tagged' social media users' mobility flows between city-pairs can be measured with the explicit aim of exploiting urban network patterns.

This research contributes to several strands of literature. First, it adds to the work on applications of social media data resources. Despite the booming trend of social media users in developing countries, the empirical literature has mostly focused on the U.S. and European countries. These studies have allowed for the simulation and modeling of distribution and dynamics of traffic flows (Steiger et al., 2014), mobile users (Malleson and Andresen, 2014), urban population (Aubrecht et al., 2011), LBSM users' social network (Ahern et al., 2007; Backstrom et al., 2010; Sun et al., 2013) and food health (Widener and Li, 2014), as well as spatio-temporal predictions of natural disaster progresses (e.g. earthquake, forest fire). These studies are of interest in their own right and are important for the development of optimal public policy. We are–for the first time in the literature–to comprehensively measure urban network patterns at a detailed spatial degree (the city-pair level) based on location-based social media data from a large developing country context.

There is a substantial literature that investigates various aspects of human mobility behaviors. Much of it is concerned with variation in local amenities (e.g. crimes) and population distribution within cities, an issue not directly related to our work. Only a small a number of papers look at the social and spatial interactions of individuals and cities (Crandall et al., 2009; Cranshaw et al., 2012; Gao et al., 2012; Stephens, 2013; Rosler and Liebig, 2013; Stefanidis et al., 2013; Liu et al., 2014; Lovelace et al., 2014; Hollenstein and Purves, 2014). In what is probably the most closely related paper to our own, Jiang and Miao (2014) uses location-based social media data to examine the evolution of the rank-size distribution of cities in the mainland US. Put differently, we look at the dynamics of inter-city connection patterns in China, an important complementary inquiry.

Finally, our work is related to the spatial economic literature dealing with the spillovers of agglomeration effects (Andersson et al., 2004; Rosenthal and Strange, 2008; Arzaghi and Henderson, 2008). By showing that migration flows from periphery cities tend to cluster into large metropolitan cities, recent economic studies suggest that external economies of agglomeration are substantial, but sharply attenuated by geographical distance across cities. But migration flow information from conventional census data cannot capture the real-time dynamics of human mobility flows between city pairs. In our analytical framework, we use space–time trajectories to track the spatial and temporal dimensions of social media users' activities. Our examination of changes in human mobility patterns by months by city-pairs throughout China by months involves many potential stages of big data mining analysis. We stratify cities by core-periphery urban systems, by regions and by calendar months, finding that human mobility flows are not distributed evenly over time and across space. Interestingly, we find larger human mobility flows around the Chinese New Year month and the summer months. Our evidence suggests the significantly heterogeneity patterns of core-periphery urban systems as reflected from real-time human mobility flows.

The remainder of this paper is organized as follows. Section 2 introduces the spatially-integrated social media data modeling framework for extracting and aggregating location-based social media data information. In Section 3 we describe the data for our empirical implementation. Section 4 implements the methodology into mapping spatiotemporal patterns of migration flows of social media users in Chinese cities. Section 5 discusses the economic geography implications of social interactions of cities and potential channels at work. Section 6 presents the concluding remarks.

2 Spatially-integrated social media data modeling framework

A spatially-integrated social media data modeling framework is coded in R after a substantial modification and secondary programming based on the existing literature. The specific model setup is detailed below.

2.1 Defining activity trajectories of social media users

The starting point for our analysis is to use the space-time trajectories of social media users for identifying individuals' footprints in a geographic space. Assume that there is a country space which contains *M* cities available for individuals' mobility. A set of *N* individuals would post their daily social activities (e.g., traveling) through a location-based social media platform[1]. We seek to use a credible measure of "space–time trajectory "for capturing human mobility pattern. Our guiding principle in defining the space–time trajectory has been to follow Hägerstraand (1970)'s implicit function used in the geographical analysis (Zheng & Zhou, 2011; Cao et al., 2015).

We define that a social media user, $u_i\ (i \in 1, N)$, has a true space–time trajectory W_i within a country. This real-life trajectory W_i is approximately identified by WT_i; Where WT_i represents a set of geographically-tagged footprints of location (l_i), timestamp (t_i) and message content (c_i) posted in social media. So, for each user u_i, $WT_i=\{(l_i^j, t_i^j, c_i^j),(l_i^{j+1}, t_i^{j+1}, c_i^{j+1}),(l_i^{j+2}, t_i^{j+2}, c_i^{j+2}),,,(l_i^{j+k}, t_i^{j+k}, c_i^{j+k}),,,\}$.where $j \ge 0; k \ge 0; t_i^{j+k} \ge t_i^{j+k-1} \ldots \ge t_i^j$. One thing to note is that, unlike traditional spatial trajectories of geographic objects, social media users' space-time trajectories (WT_i) are not likely to be sampled at regular time intervals (Zheng & Zhou, 2011; Gao & Liu, 2013). As suggested by Andrienko et al. (2012), space-time trajectories of social media users have the 'episodic" nature. Indeed, it is possible that users may not frequently share their mobility behaviors, and users can even choose to disable

[1] To avoid personal confidential data concerns, we assume that the location-based social media data company can protect the information about individual socioeconomic characteristics.

location positions when posting their social activities. Conventional methods (e.g., interpolation and map-matching) in used spatial trajectory analysis would at best capture the estimated footprint patterns of social media users. For simplicity of implementation, we solve Hägerstraand (1970)'s implicit function numerically and write social media users' W_i as a function of time t, defined by space-time trajectories WT_i, i.e., a social media user u_i stays at the origin location l_i during $[t_i, t_{i+1}]$ until a new activity is posted at t_{i+1} in a different location l_{i+1}.

In light of precision issues, we restrict our focus on social media users' inter-city mobility activities, rather than intra-city mobility behaviors. Building on the sociology literature, people are likely to share their activities if they are visiting a city that is different from their origins or current residences. But one fundamental threat to identification is how to identify where is a social media user's *origin* city. Existing studies have often used to the most frequently visited city as the origin city of the social media user, and used a spatial radius to track individuals' footprints (Gonzalez et al., 2008; Cao et al., 2015). Cities that have higher visiting frequency rates are potentially very different from those that do not. These differences may arise through disparities in individuals' initial motivations. For example, it is likely that the most frequent visited city for a businessman is his project's location, rather than his home location. For a more rigorous assessment, we apply the text mining methods (Rao et al., 2010; Burger et al., 2011; Wang et al., 2013) to derive social media users' current residence information. To be specific, we define a social media user's current residence as his or her origin city, while other cities in space-time trajectories as visited destination cities.

2.2 Dimensional mobility algorithm

Adopted from Leonardi et al. (2014), a graphical framework for data warehousing and data cuboid, is employed in this study to represent the dimensional mobility algorithm of social media users across cities. In the social media data cuboid, we stratify three dimensions: First, user dimension. We apply the text mining methods to read the individuals' socioeconomic information (such as gender, number of friends). To avoid data privacy concerns, we restrict our focus onto extracting users' origin city information and geo-tagged footprints information from text-based contents. Second, spatial dimension. It creates $1km^2$ grid cells, identifies the number of users in each cell unit as the basic spatial cubioid scenario. We define the cubioid (***C***)-based geometric measures as follows:

$C(u_i)$: the number of social media users who are residents in the origin location O;

$OutC(u_i)$: the number of mobility visits made by the user i from the origin location O to other destinations.

$InC(u_i)$: the number of mobility visits made by the user i from other locations into the origin location O.

The third dimension is related to the temporal information. We break down the temporal measurement intervals into days as our baseline temporal cuboid. Evidently, by interacting the temporal dimension with spatial-user dimensions, we can quantify human mobility flow patterns between cubiod pairs over time and space.

2.3 Aggregation function

Incorporating individuals' space-time trajectories into the dimensional mobility geometrics measures requires appropriate aggregation functions for efficient data query operations (Gray et al., 1997). Assume that U is an aggregated spatiotemporal hierarchy corresponding to a set of human mobility patterns of social media users u_i between a pair-wise cities: p_1 and p_2. Specifically, p_1 and p_2 represent the higher levels of the hierarchy cubiods (e.g. month&city-based cubiod), aggregated by a series of basic cubiod measures (e.g. day&grid unit-based cubiod): $p_1 = \sum_{i=1}^{k} p_{1,i}$,where $\{p_{1,i}, i = 1,2,\ldots k\}$; and $p_2 = \sum_{j=1}^{k} p_{2,j}$, where $\{p_{2,j}, j = 1,2,\ldots k\}$. Thus we can write the aggregation function for measuring mobility flows between p_1 and p_2 as follows: $F(c_1, c_2) = \sum_{i,j=1}^{i,j=k} F(c_{1,i}, c_{2,j})$.

In addition to the mobility flows between city pairs, it would also be interesting to know the total out-flow volume and in-flow volume of social media users for each city. Recall that we assume that there are M cities (p_m) available for individuals' mobility in China. By deducting the space–time trajectories that occurred within the city boundaries of p_m, we can finalize the aggregation function for measuring total out-flow volume ($OutC(p_m)$) and in-flow volume ($InC(p_m)$) of social media users of p_m as:

$$OutC\left(p_m\right) = \sum_{i=1}^{k} OutC\left(p_{m,i}\right) = \sum_{i=1}^{k} F\left(p_{m,i}\right)$$

$$InC\left(p_m\right) = \sum_{i=1}^{k} InC\left(p_{m,i}\right) = \sum_{i=1}^{k} F\left(p_{m,i}\right)$$

While the above aggregation function is robust, this specification was unworkable in practice as location-based social media data are extremely massive and dynamic over time and space. To facilitate the big data processing, we adopt Leonardi et al (2014)'s framework for trajectory data warehousing and visual OLAP and measure the aggregated cubiod space-time trajectories of social media users between city pairs in our implementation.

2.4 Mapping network topology pattern

In this study, we focus on assessing the spatiotemporal patterns of intercity social interaction connections by using social media users' mobility data. Hypothetical intercity connections are defined as individuals' intercity mobility behaviors---reflected from aggregated 'geo-tagged' location-based social media data. The rationale behind this is that, an intercity connection "geo-tagged" linkage will be created if an individual's twitter is registered in city A but sends a tweet from city B. This assumption is critical because it has allowed for the transformation of social media users' geo-tagged records into the topology of intercity connection networks as detailed below.

We adopt the directed star network topology approach to construct the matrix between a social media user's origin city (corresponding to individuals' current residence city) and destination cities (corresponding to individuals' geo-tagged cities). In line with the common practice, we calculate the origin-destination matrix using a two-step procedure.

In the first step, we use the data mining methods to retrieve all users' geo-tagged records throughout China. This involves of massive computing exercises due to the large amount of LBSM users' information (around 300 billion records). A typical example for illustrating a user's geo-tagged record is, {< origin city=City O>, <City-D1, January.10>, <City-D2, Feburary.12>, <City-D3, Feburary.15>, <City-D4, March.15>, <City-D2, April.3>, <City-D3, June.30>…}. This simple example illustrates the trajectory of a user from the origin city (City-O) to four geo-tagged destination cities (City-D1, City-D2, City-D3, City-D4) in six times from January to June.

The second step reads and calculates the intensity of linkages between city pairs within the directed star topology network. Our guiding principles include: First, the origin city are defined as the core node in the network, whereas other geo-tagged destination cities are defined as leaf nodes that include connections to the origin city. Second, we characterize the direction of city-pair linkages from the core code to leaf codes. This means that directed intercity connection flows are drawn from assigning an outward direction from the core code to each leaf code in the topology network. Third, we use the frequencies of geo-tagged destination cities to weight the linkages of intercity connections in the topology network. By doing so, we can adjust the origin-destination human mobility flows along the network to reflect the intensity of intercity linkages. To identify the overall volume of migration flows, we also calculate the accumulative migration flows out of this city and towards this city during our study period. We then use the aggregation function outlined above to aggregate space-time trajectories of social media users into the prefecture city level by months. One thing to note is that, many computing runs with this computing step specification failed due to insufficient hardware resources and successful runs were implemented by using the super-machines installed in the Institute of Geographical Sciences and Natural Resource Research, Chinese Academy of Science.

3 Empirical implementation

We focus on the most commonly used location-based social media data available in China---*Weibo*. Weibo (literally means 'microblog' in English), which is often seen as the Chinese version of Twitter, is essentially a web-based social media platform. Similar like Twitter, Weibo users can post a short text message (with a 140-character limit) for showing subjective impressions and daily activities (Java et al. 2007). With more than 100 million daily active users and more than 1 billion monthly active users, Weibo (www.weibo.com) has provided us a new way to explore real-time migrant flows between city pairs in China. We improve on the applications of Weibo data in two ways: First, we design a sophisticated big data mining programme to search, gather and extract billions of Weibo records from the Weibo's public application programming interface (API) system. Second, Weibo provides a location-tracked application tool (commonly known as a 'geo-tagged' service) for identifying users' geographical locations. A geo-tagged record, therefore, includes spatial information for studying users' mobility behaviors across cities. By mining and aggregating millions of users' geo-tagged records, we can construct the origin-to-destination city matrix based on the directed star topology network method.

In our work, the Weibo API was used to gather the geo-tagged records submitted within the mainland China over the period from September 2013 to September 2014. Our overall sample contains 30.53 billion geo-tagged records, where each record provides a Weibo user's basic information (such as user-ID, origin city, gender, the number of connected Weibo friends, and so on), messages sent by days and by locations (geographic coordinates). Figure 2 shows the 'heat map' of these geo-tagged records.

We adopt the spatially-integrated social media mining framework and structure billions of data records by (1) only keeping information about user ID, origin city and geo-tagged destination cities; (2) joining individual trajectory records based on the same user ID; (3) dividing each individual's geo-tagged records into two samples: origin city and geo-tagged destination cities; (4) matching the precise geographic coordinate information of tagged locations with corresponding prefecture city names; and (5) applying the directed star topology network method to extract geo-tagged information from our database and aggregating millions of individual's intercity mobility behaviors by months by prefecture cities.

There are several important limitations underlying our implementation. First, we focus on the network of intercity connections within the mainland China (excluding

HongKong, Macau, and Taiwan). Our measurement captures the aggregated city-level human mobility flows between city pairs in one-year period. Lacking long-term LBSM data and systemic Chinese users' international tweet data, we are pushed to focus on the short-run, intra-country-variation of urban networks. But our methodology can be easily applied into the international level when better data are available. Second, our network data does not allow us to observe the drivers and motivations of human mobility behaviors at the individual level. It is highly possible that some people may visit other cities for business purposes, whereas others may travel to different cities for family union, leisure and study purposes. This is a so-called second-order effect that may be partly related to the cultural identity of cities. In the empirical analyses, we are pushed to focus on the first order effect that only considers an intercity connection as a human mobility linkage along the origin-to-destination routes between city pairs. But one advantage for using the human mobility linkage is that any measured changes in visiting frequencies between city pairs over time are determined by variation in the location of cities. Despite of potential limitations, our spatially-integrated data cuboid model allows to transform the massive, dynamic and unstructured location-based social media data into a structured framework for exploring spatiotemporal patterns of urban networks in a large developing country context. This is novel.

4 Exploration of social interaction patterns in China

Seeing city and regional disparities in a large developing country through social media user migration flows is new to the existing literature. The visualized flow mapping methods have been recently applied in the literature to represent dynamics of goods and people across space (Guo et al., 2006; Wang, 2010; Verbeek et al, 2011). By using real-time space-time trajectories of location-based social media data, we are able to investigate geographical implications of migration flows of social media users across spatiotemporal scales. The model results are presented below in two ways. First, we monitor the transition trajectories of social media users' migration flows and characterize the aggregated spatiotemporal outcomes of city and regional disparities. The second way of presentation assesses spatiotemporal patterns of pair-wise social connection networks by different geographical scales (city pairs, province pairs and region pairs).

Other sections will be added in March

References

Ahern, S., Naaman, M., Nair, R. & Yang, J. H.-I. (2007). World explorer: visualizing aggregate data from unstructured text in geo-referenced collections. In: *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries. ACM,* 1-10.

Andersson, R., Quigley, J. M., Wilhelmson, M. (2004). University decentralization as regional policy: the Swedish experiment, *Journal of Economic Geography*, 4(4), 371-388.

Andrienko, N., Andrienko, G., Stange, H., Liebig, T., & Hecker, D. (2012). Visual analytics for understanding spatial situations from episodic movement data. *KI – Künstliche Intelligenz*, 26, 241–251.

Arzaghi, M., Henderson, J. V. (2008). Networking off Madison Avenue, *Review of Economic Studies*, 75(4), 1011-1038.

Aubercht, C., Ungar, J. & Freire, S. (2011). Exploring the potential of volunteered geo-graphic information for modeling spatio-temporal characteristics of urban population. *Proceedings of 7VCT,* 11, 13.

Backstrom, L., Sun, E. & Marlow, C. (2010). Find me if you can: improving geographical prediction with social and spatial proximity. In: *Proceedings of the 19th international conference on World wide web. ACM,* 61-70.

Baum-Snow, N., Brandt, L., Henderson, J., Turner, M., & Zhang, Q. (2013). Roads, railroads and decentralization of Chinese cities. *mimeo, Brown University.*

Burger, J. D., Henderson, J., Kim, G., & Zarrella, G. (2011). Discriminating gender on Twitter. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 1301–1309). Association for Computational Linguistics.

Cao, G., Wang, S., Hwang, M., Padmanabhan, A., Zhang, Z., & Soltani, K. (2015). A scalable framework for spatiotemporal analysis of location-based social media data. *Computers, Environment and Urban Systems*, 51, 70–82.

Crandall, D. J., Backstrom, L., Huttenlocher, D. & Kleinberg, J. (2009). Mapping the world's photos. In: *Proceedings of the 18th international conference on World wide web. ACM,* 761-770.

Cranshaw, J., Schwartz, R., Hong, J. I. & Sadeh, N. M. (2012). The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In: *ICWSM.*

Gao, H. & Liu, H. (2013). Data analysis on location-based social networks. In A. Chin & D. Zhang (Eds.), *Mobile social networking: An innovative approach* (pp. 165– 194).

Gao, H., Tang, J. & Liu, H. (2012). Exploring social-historical ties on location-based social networks. In: *ICWSM.*

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453, 779–782.

Gray, J., Chaudhuri, S., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M., et al. (1997). Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-totals. *Data Mining and Knowledge Discovery*, 1, 29–53.

Guo, D., Chen, J., MacEachren, A. M., & Liao, K. (2006). A visualization system for space-time and multivariate patterns (vis-stamp). *IEEE Transactions on Visualization and Computer Graphics*, 12, 1461–1474.

Hägerstraand, T. (1970). What about people in regional science? *Papers in Regional Science*, 24, 7–24.

Hollenstein, L. & Purves, R. (2014). Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, 21-48.

Java, A., Song, X., Finin, T. & Tseng, B. 2007. Why we twitter: understanding microblogging usage and communities. In: *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*. ACM, 56-65.

Jiang, B. & Miao, Y. (2014). The evolution of natural cities from the perspective of location-based social media. *arXiv preprint arXiv:1401.6756.*

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. The *American Economic Review*, 70(5), 950–959.

Leonardi, L., Orlando, S., Raffaetà, A., Roncato, A., Silvestri, C., Andrienko, G., et al. (2014). A general framework for trajectory data warehousing and visual OLAP. *GeoInformatica*, 18, 273–312.

Liu, Y., Sui, Z., Kang, C. & Gao, Y. (2014). Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *Plos One*, 9, e86026.

Lovelace, R., Malleson, N., Harland, K. & Birkin, M. (2014). Geotagged tweets to inform a spatial interaction model: a case study of museums. *arXiv preprint arXiv:1403.5118.*

Malleson, N. & Andersen, M. A. (2014). The impact of using social media data in crime rate calculations: shifting hot spots and changing spatial patterns. *Cartography and Geographic Information Science*, 1-10.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, (pp. 1–137).

Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010). Classifying latent user attributes in Twitter. In *Proceedings of the 2nd international workshop on search and mining user-generated contents* (pp. 37–44). ACM.

Rosenthal, S. S., Strange, W. C. (2008). The attenuation of human capital spillovers, *Journal of Urban Economics*, 64(2), 373-389.

Rosler, R. & Liebig, T. (2013). Using Data from Location Based Social Networks for Urban Activity Clustering. *Geographic Information Science at the Heart of Europe*. Switzerland:

Springer.

Stefanidis, A., Cotnoir, A., Croitoru, A., Crooks, A., Rice, M. & Radzikowski, J. (2013). Demarcating new boundaries: mapping virtual polycentric communities through social media content. Cartography and Geographic Information Science, 40, 116-129.

Steiger, E., Ellersiek, T. & Zipf, A. (2014). Explorative public transport flow analysis from uncertain social media data. In: *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information. ACM,* 1-7.

Stephens, M. (2013). Featured graphic. Mapping the geoweb: a geography of Twitter. *Environment and Planning A*, 44, 100-102.

Sui, D. & Goodchild, M. (2011). The convergence of GIS and social media: challenges for GIScience. *International journal of geographical information science*, 25, 1737-1748.

Sun, Y., Fan, H., Helbich, M. & Zipf, A. (2013). Analyzing human activities through volunteered geographic information: Using Flickr to analyze spatial and temporal pattern of tourist accommodation. In: *KRISP, J. M. (ed.) Progress in Location-Based Services.* Berlin Heidelberg: Springer.

Verbeek, K., Buchin, K., & Speckmann, B. (2011). Flow map layout via spiral trees. *IEEE Transactions on Visualization and Computer Graphics*, 17, 2536–2544.

Wang, S. (2010). A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. Annals of the Association of American Geographers, 100, 535–557.

Wang, S., Cao, G., Zhang, Z., & Zhao, Y. (2013). A CyberGIS environment for analysis of location-based social media data. Advanced Location-based Technologies and Services, p. 187

Widener, M. J. & Li, W. (2014). Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Applied Geography*, 54, 189-197.

Wright, D., & Wang, S. (2011). The emergence of spatial cyberinfrastructure. Proceedings of the National Academy of Sciences, 108, 5488–5491.

Zheng, Y., & Zhou, X. (2011). Computing with spatial trajectories. Springer.