CHAPTER 18

Research on social media feeds – A GIScience perspective

Enrico Steiger, Rene Westerholt and Alexander Zipf*
*zipf@uni-heidelberg.de

Introduction

During the last two decades, the role of internet users changed dramatically. While they were mostly passive content consumers before, they are now considered proactive data producers. This phenomenon is summarized by the term "Prosumer" (Ritzer & Jurgenson (2010) and gets facilitated through major technological advancements such as ubiquitous access to the mobile Internet and a widespread use of smartphones equipped with positioning and sensing capabilities. These outlined developments do not just happen recently, but trace back to the much older development around the so called "Web 2.0" (ITU 2014). In geospatial terms, these developments are well reflected by Mike Goodchild's popular definition of 'Citizens as Sensors' (Goodchild 2007), where ordinary people capture and disseminate "Volunteered Geographic Information (VGI)." Haklay further puts this development into broader context and rather coined the term "GeoWeb" (Haklay *et al.* 2008). OpenStreetMap (OSM) is probably the most prominent example of VGI.

Projects like OSM provide a well-defined data capturing protocol as well as a clear mission regarding their contributed contents. In contrast, data originating from online social networks (another source of VGI) is way more heterogeneous and diverse. At the same time, however, it may also provide high levels of semantic detail and is generated by a larger number of users. Consequently, it gained the interest of various research disciplines. These range from sociology

How to cite this book chapter:
Steiger, E, Westerholt, R and Zipf, A. 2016. Research on social media feeds – A GIScience perspective. In: Capineri, C, Haklay, M, Huang, H, Antoniou, V, Kettunen, J, Ostermann, F and Purves, R. (eds.) *European Handbook of Crowdsourced Geographic Information*, Pp. 237–254. London: Ubiquity Press. DOI: http://dx.doi.org/10.5334/bax.r.

toward linguistics, and of course geography and GIScience. The latter one is facilitated by the fact that a great deal of information contributed to social media is geotagged. Thus, the remainder of this chapter focuses on the spatial aspects of social media, and potential applications that can be derived from this kind of data.

Section A highlights the general potential of social media analysis for investigating social phenomena. We do that by outlining selected case studies from the exemplary field of human mobility analysis. These have demonstrated the usefulness of social media for investigating mobility patterns as well as human spatial behavior. Section B then provides an overview of several different application domains of social media analyses, with a particular focus on Twitter. Finally, Section C discusses some technical issues of established spatial analysis methods and their application to social media data. We conclude the chapter by summarizing its different parts. We further provide recommendations regarding future areas for GIScience research on social media data.

Utilization of social media data for investigating urban environments

The spatial and social structures of a city as well as the dynamic nature of human activities result in certain collective and individual human behavior patterns. Social media data can help to "sense" this type of information from urban environments in an in-situ manner. GIScience research thereby is focused on the overall question how corresponding spatiotemporal patterns from ubiquitous sensor networks and heterogeneous data streams can be explored, extracted, validated and aggregated. In turn, such information might enable us to sense everyday spatial processes and to gain knowledge about urban environments, especially with respect to collective human dynamics. The study of these issues has become one of the primary objectives of GIScience (Giannotti & Pedreschi 2008).

The information originating from social media messages (e.g., Tweets in case of Twitter) may contain spatial, temporal and semantic attributes. Considering these dimensions, social media can be considered as a (partial) proxy of real world happenings. However, space, time as well as semantics are influenced by each user's individual perception of the surrounding space. It is thus important to figure out ways to circumvent these issues for gaining trustworthy and objective information from these data sources.

The following short paragraphs outline case studies in which a range of GIScience researchers has drawn human mobility and urban study related knowledge from Twitter. We group these studies in accordance to their underlying research goals. The listed paragraphs thus provide the reader a quick overview of both the types of studies that have been conducted as well as methods and outcomes.

Mobility and social behavior. Studying the social dynamics of a city remains a challenging endeavor, which has recently been carried in a qualitative manner. Thus, social media might be a promising source of information in order to provide a better understanding of social dynamics within urban environments and resulted in various research efforts. Regarding the analysis of collective human mobility and activity patterns from social media, Cho et al. (2011) investigate social ties and their influence on human mobility patterns by comparing social media check-in data and cellphone location data. They found a stronger association of social network ties influencing long-distance travel than short range spatially and temporally periodic movements. Within the observed Twitter user pattern, Lee & Sumiya (2010) study user behavior by measuring geographic regularities and detecting geo-social events through identifying Regions of Interests (RoI). Another approach conducted by Noulas et al. (2011), Cranshaw et al. (2012) and Kafsi and Cramer (2015) is the identification of characteristic neighborhoods, collective movement patterns and social ties within certain user communities from Foursquare and other Social media data. In a similar approach for Twitter, Li *et al.* (2014) measure the spatial dispersion of users in a community and their trajectories. Hawelka et al. (2012) aim to further empirically validated the observed human behavior patterns and found a correlation between the conducted Twitter census and economic key figures. Furthermore, Li et al. (2013) explore spatiotemporal patterns of Twitter and Flickr data and investigated a relationship between socioeconomic characteristics of people who are generating social media posts in the US.

Mobility and underlying urban structures. The exploration of the relationships and the impact of urban structures on human mobility is an interesting study area for social media researcher. Wakamiya et al. (2011) investigate temporal patterns of crowd behavior over Japan by spatial partitioning Tweets in order to extract urban characteristics. On a smaller scale several studies investigate the connection with extracted urban activities from social media and their connection with the underlying urban structure. Kling et al. (2012) were able to detect spatiotemporal clusters of frequently occurring urban topics in New York. Furthermore, Ferrari et al. (2012) also work with georeferenced Tweets and a semantic probabilistic topic modeling approach to automatically extract urban patterns from location-based social networks. The study concluded that extracted urban motion patterns and identified hotspots in the city allow the inference of crowd behaviors that recur over time and space. A similar approach by using Foursquare data by Cheng et al. (2011) and Hasan et al. (2013) also resulted in the characterization of urban human mobility and activity patterns. Andrienko & Andrienko (2013) correlated the spatiotemporal clusters of keyword based filtered georeferenced Tweets of places where people Tweet with US population densities. The results have shown strong correlations between the observed Twitter distribution and census data, suggesting that social media is a reliable proxy for the inference of mobility patterns. One further application is to derive intra-urban events showing distinct mobility patterns over time. This

spatiotemporal movement has proven to reflect typical mobility behavior in the underlying urban structures (Steiger, Westerholt, et al. 2015).

Mobility and human activities. Several studies infer individual and collective human daily activity patterns by analyzing crowdsourced information, such as taxi trip records (Liang et al. 2012), GPS traces (Azevedo & Bezerra 2009) (Jiang, Yin, & Zhao 2009) or mobile phone records (Candia & González 2008) (Gao 2014). Consequently, a large literature body also focus on studying human mobility and activity pattern from social media data. Krumm et al. (2011) estimate individual home locations of heavy Twitter user and apply machine learning algorithms to classify and predict individual travel behavior. Jin *et al.* (2014) developed a method to infer users' mobility patterns from check-ins in Foursquare. Coffey & Pozdnoukhov (2013) go one step further and semantically annotate mobility flow datasets with activity information and trip purposes from Tweets. Similarly, Wu *et al.* (2015) utilize social media to annotate the location history of mobile phone users for the characterization of certain social activities. Focusing on the content of Tweets, Grinberg *et al.* (2013) proposed a method to detect semantic patterns to infer clusters of users' real world activity. Gao (2014) developed a probabilistic approach to make place recommendations based on the users' geo-social circles, as extracted from Foursquare. In another study, the authors estimate spatiotemporal mobility flows from Twitter for the area of greater Los Angeles to infer origin- and destination trips (Gao et al. 2014). Results have shown similar pattern when comparing with community survey data. In a previous study we introduced a semantic and spatial analysis method (Steiger, Lauer, *et al.* 2014), through which we were able to extract geographic features from uncertain Twitter data and have shown that observed clusters correspond to landmarks, such as highly frequented squares and major transportation hubs. A further investigation revealed similar semantic layers that represent collective human mobility flows in co-occurrence with underlying social activity (Steiger, Ellersiek, *et al.* 2014) and could thus lead to new insights in characterizing urban mobility.

Future research recommendations

Further research needs to be conducted to assess the reliability of social media datasets. It also must be noted that the data collected from wireless devices are influenced by GPS/WIFI inaccuracy issues (Zandbergen and Barbeau 2011). Moreover, users can individually choose to share their precise location to a Tweet or just a general location information (such as a city or neighborhood). This resulting location uncertainty leads to imprecise location information of geotagged Tweets(Li *et al.* 2011).

Within the semantic attribute one must consider that the containing information may relate to events in the past, present or even future (Sengstock and

Gertz 2012). Principally the text corpora as such in social media posts are relatively sparse and vague. It may also be fairly ambiguous and hence observed phenomena may only be a weak indicator of a real world event. This uncertain semantic knowledge is a result of the fact that people using Twitter have individual motivations to post information and their main intention is to primarily serve their own communication needs. One further typical characteristic of social media is that users do not post equally distributed in geographic space and time leading to a heterogeneous dispersion of posts. Jatowt *et al.* (2015) further assess these varying temporal patterns and dynamics within social media. Furthermore, georeferenced social media posts only represent a small fraction of the overall available data. Not all user groups use all types of social media platforms similarly, which produces a potentially strong socio-demographic bias (Longley and Adnan 2015). Last, the application of spatial and semantic methods themselves creates uncertainties, since the distribution of specific geographic phenomena and their semantic complexities within Tweets are not known beforehand (Westerholt *et al.* 2015). Hence, it is important to compare and validate results with other acquired sensor data.

Conducting further research in this area however will be worthwhile, since study results may provide new additional insights into the complex human-sensor-city relationship at a much more fine-grained spatial and temporal level than before. New knowledge gained from this research will provide a better understanding of individual and collective human behavior within urban environments and may assist stakeholders and decision makers in their planning processes.

Application Domains of Social Media Analyses

Location-based social networks (LBSN) (Roick and Heuser 2013) offer a vast amount of voluntary content. The investigation of human activities in location-based social networks is one promising example of exploring spatial structures in order to infer underlying spatiotemporal patterns. Twitter for example is more and more recognized by numerous research domains. In particular it provides an opportunity for GIScience to understand geographic processes and spatial relationships comprised in social networks. Summarizing the current state of research concerning the application for spatiotemporal analyses, one outcome of a previously conducted systematic literature (Steiger, Albuquerque, *et al.* 2015) revealed that Twitter analyses are mainly focused on the spatiotemporal classification and detection of events. Principal investigated application domains are:

Event Detection. To detect events, researchers are currently looking for spatial, temporal and semantic patterns within Twitter. In this respect people act as a social sensors for events (Yardi and Boyd 2010, Chae *et al.*

2012). *Disaster- and emergency management* as one event detection subfield has been the primarily identified application in nearly a third of all reviewed studies (Sakaki *et al.* 2010, Murthy and Longwell 2013, Crooks *et al.* 2013). Further research has been conducted on utilizing Twitter in *traffic management.* This can be found in 14% of reviewed studies (Kosala and Adi 2012, Wakamiya and Lee 2012, Lenormand *et al.* 2014). Another area which seems to be quite popular is research on Twitter data for *disease/ health management* adding up to another 5% of the reviewed studies (Lampos and Cristianini 2010, Veloso and Ferraz 2011, Sofean and Smith 2012). A famous example is the derivation and prediction of information on infection sources and the spreading of an illness from Twitter messages (Culotta 2010, Collier *et al.* 2011). One prominent example is earthquake detection from Twitter data (Longueville *et al.* 2010, Zook *et al.* 2010). This has been successfully accomplished in a number of studies correlating results with official earthquake sensor data (Tapia *et al.* 2011, Thomson *et al.* 2012). Sakaki *et al.* (2010) have developed an algorithm that uses Twitter to calculate earthquakes' epicenters and the typhoons' trajectories. Moreover, situational information can be derived from location-related short messages to coordinate emergency responses (Vieweg *et al.* 2010). Also in the context of disease and health management similar outcomes have been derived. Tweets showing disease incidents have shown similar spatiotemporal distributions as those in with official reports. With these studies research has proven the trustworthiness and a high level of representativeness of Tweets throughout different application domains (Albuquerque *et al.* 2015).

Location Inference. Locations of users within social networks can be inferred or even predicted with the help of direct or indirect geolocation information derived from the provided metadata or from the semantic content (Kinsella *et al.* 2011, Hong *et al.* 2012, Hiruta *et al.* 2012). The geographic accuracy could be increased by extracting the textual information from the Tweet or from the metadata itself. For example, Lamprianidis and Pfoser (2011) have extracted locations and their names from Flickr pictures by clustering user-generated data points associated with geo-referenced pictures. Kelm *et al.* (2013) discusses various methods to extract place names from textual data from articles, posts or tags in geo-social networks, including place name gazetteer and statistical language modeling. Some methods follow an opposite approach and infer the location of a feature from implicit location information. Serdyukov *et al.* (2009) model the probability that a group of tags be assigned to a location. Similarly, (Gallagher *et al.* 2009) used location probability maps generated from tags for the same purpose. Van Laere *et al.* (2010) have pursued the same goal using k-medoids and Naive Bayes clustering methods. Some approaches focus on inferring a user's or a group of users' location. Cheng *et al.* (2010) have proposed a probabilistic method to determine users´

location from the content of their Twitter messages. Other authors have proposed to use the location of users´ friends to achieve the same goal (Backstrom *et al.* 2010). Stefanidis *et al.* (2013) have proposed a framework to harvest ambient geospatial information from social media feeds to locate social hotspots or to map social networks in a given geographical area. Ajao *et al.* (2015) summarize the broad range of available techniques applied to infer direct and indirect location from Twitter messages and social media users.

Geo-Social Network Analysis. Another important domain of research is social analysis which investigates relationships of individual users within a social network (Wu *et al.* 2011, Cranshaw *et al.* 2012). Geo-social network analysis seeks to identify the structure of social networks and their distribution in geographic space (Scellato and Mascolo 2010, Lee and Sumiya 2010). Social ties may feature distinct spatial distributions enabling spatiotemporal analyses. These distributions can help finding collective social activities and ultimately understanding geographical processes. A subfield of geo-social network analysis are *sentiment and emotion analysis* (Wang *et al.* 2012, Quercia *et al.* 2012). This field of research also offers a great potential for GIScience in the context of extracting contextual emotional information within urban and rural environments. One promising further field of research within social analysis which should be mentioned is urban planning and management which also could benefit from the rich data found in location based social networks such as Twitter. In the context of disaster management, several studies aim to infer the social dimensions within certain geo-located communities in twitter during disaster events (Conover *et al.* 2013, Bakillah *et al.* 2014).

Future research recommendations

Social Media data for research has proven to be a valuable source, as it not only comes for free, but also features a high spatiotemporal resolution. This kind of data especially enables possibilities to find spatial patterns and events which can help validating existing information sources. One identified main research gap is the exploration of human spatial behavior (Miller & Goodchild 2014) in order to gain knowledge about the underlying geographic processes and dynamics. Furthermore, the current research foci allow to transfer established methods from various disciplines (e.g. Computer- and Information Science, Social Science etc.) into other disciplines and enhancing new applications. As one example, more use of computer linguistic approaches to leverage knowledge from textual information, combined with methods for spatiotemporal analysis from computational sciences could lead to new insights within specific geographic application domains, such as disaster management or human mobility analysis.

Spatial Analysis of Social Media Feeds – Challenges and Approaches

The primary goal of spatial analysis is to explore structures within spatial data. This typically involves tasks like finding clusters on a map or figuring out distributional characteristics of data. One theoretical field underlying spatial analysis is spatial statistics. This field provides the basic principles that are underlying many spatial analysis problems. Key to this field is identifying spatial correlations, and thus hints on systematic patterns in geographic data (Fischer & Getis 2010). Respective methods and techniques are thus useful tools for gaining geographic insight into social media data.

The spatial analysis of social media data is typically conducted in an exploratory manner. This is due to lacking knowledge about potential underlying spatial processes, and thus about social media messages and their dispersal in geographic space in general. Useful tools on that regard are the K-Function (Ripley 1976) (purely geometric) and the mark correlation function (Stoyan & Stoyan 1994) (attribute values), both originating from spatial point pattern analysis. These methods allow identifying significant geometric clustering and regularity within stochastic point patterns. When the geometry is fixed (or rather treated as such) spatial autocorrelation statistics like Moran's I (Moran 1950, Cliff & Ord 1973) and hot spot statistics like Getis-Ord's G statistics (Getis & Ord 1992, Ord & Getis 1995) are suitable alternatives. These assess the degree of randomness within georeferenced attributes associated to units on a fixed geographic layout. In fact, many of the latter are essentially identical to different variants of the mark correlation function (see, e.g., Shimatani 2002). Thereby, Moran's I tests for correlations between neighbored observations across space, while G separates between extremal values (i.e., high and low).

As mentioned earlier, thorough spatial knowledge about social media datasets is typically lacking. Consequently, analysts oftentimes proceed with a trial-and-error approach when parameterizing the methods mentioned above. It is common practice to apply these techniques to different scales. The goal then is to sort out that scale at which patterning seems to be most pronounced. However, the techniques mentioned so far were designed long before the appearance of social media and similar kinds of user-generated data. The idea of the following two sections is thus to briefly reflect differences between social media and more traditional data, and to give some recommendations with respect to the spatial analysis of these.

Potential issues and pitfalls

The issues presented in the following are likely to occur when analyzing social media feeds with established methods from spatial analysis. It is important to

note that social media feeds provide a mixture of indications from different real-world (and also some solely virtual) phenomena. This is due to the autonomous manner in which the data is being collected. Users can contribute any type of content from any place at any time. Such a mixture might be beneficial in terms of the wealth of contained information about the users' everyday lives. However, it also imputes some critical problems when it comes to spatial analysis. Probably the most trivial yet critical among these is the mere mixture of information as such. Any attribute which is derived from social media is highly likely to include information from several different real-world phenomena. Analyzing social media therefore comes at the risk of drawing conclusions about a mixture population that might not exist in reality. In most circumstances this is not desirable, since it does not lead to reasonable insight about any real-world process. One way to overcome this problem would be an accurate a priori semantic separation. However, that is a non-trivial task on its own right given the colloquial language used in corresponding messages.

Another issue with social media data is the implicit subjectivity that is per se introduced by the notion of "humans as sensors" (Goodchild 2007). One implication from that concept is the diversity at which people perceive environments (see also Section A). Similar phenomena might lead to varying responses among different users. This inevitably leads to an increased difficulty in analyzing the semantics (i.e. the attribute value) of the observations; and thus to a potential misclassification of phenomena. The implication of that for spatial analysis is crucial: techniques such as measures of spatial autocorrelation or spatial regression techniques are based on both, spatial characteristics as well as the attribute values. Consequently, spatial analysis techniques might end up in spurious results when the analyst fails controlling such effects.

The analysis of social media can also lead to an artificial increase in the number of type I / type II errors. This problem is likely to occur whenever testing hypotheses about spatial patterns with social media datasets. One might be interested in assessing spatial heterogeneity by means of local statistics like local Moran's I (Anselin 1995) or G_i^*(Ord and Getis 1995). It is common sense that these methods lead to an increase in type I errors due to alpha error inflation (Nelson 2012). Thus, it is important to control the alpha level accordingly (e.g., through techniques such as False-Discovery-Rate (Benjamini & Hochberg 1995)). With social media datasets, however, phenomena operating at smaller scales than the adjusted analysis scale might be considered by accident; and inadvertently influence the analysis. This is due to the mixture described above which is leading to spatially overlapping representations of different phenomena. The result is an increased amount of spurious indications of significant spatial effects.

Another critical implication of the scale-mixture outlined above is a potential creation of wrong and misleading relationships across scale levels. Recall that observations from smaller scale levels are prone to inherently being included in analyses at larger scales due to potential geometric mixture. Effects from smaller scales are therefore likely to be propagated towards analyses at larger

scales. Due to this effect, some results become impossible, e.g., in scenarios where one wants to assess spatial autocorrelation at some large scale that is influenced by highly autocorrelated observations from smaller scales. If there is spatial autocorrelation present at some small scale (e.g. one "heavy" Twitter user recurrently posting from a particular location), it will be carried through to all larger scales being observed in the same geographic neighborhood.

Further discussion of these and related problems (including some empirical results) can be found in Westerholt *et al.* (2015) (including a discussion of a multi-scale modification of the local G statistic) and Lovelace *et al.* (2016). The presented list of effects is of course not exhaustive. There might be many more effects, some of which are still about to be discovered. The subsequent section provides some hints and recommendations about how to precede with the spatial analysis of social media data.

Some recommendations

Spatial autocorrelation is the core principle underlying a great deal of spatial analysis methodology. Therefore, it is crucial to accurately assess this characteristic in order to design applicable methods, and for drawing reasonable geographic conclusions. This is not just important for exploratory tests on spatial clustering and heterogeneity, but also crucial for model-driven spatial regression scenarios such as Geographically Weighted Regression (GWR) (Fotheringham *et al.* 2003) and for assessing model misspecification (Cliff and Ord 1981). Unfortunately, in case of social media analysis, the assessment of spatial autocorrelation is strongly affected by the problems depicted in the previous section. Therefore, one recommendation in terms of future research is to work on appropriate adaptations of corresponding measures and techniques in order to account for multi-scale (or rather: "mixed-scale") and multi-categorical effects. As long as these are not available, one should carefully parameterize respective techniques. Another (aspatial) approach might be to decompose social media datasets a priori, probably based on some other characteristic such as the Tweets' semantics. The worst option of all, however, would be to neglect the specific spatial characteristics of social media data when conducting spatial analysis. That would lead to a wrong evaluation of spatial effects; and thus to wrong analysis results.

Another recommendation is related to one of the promising opportunities that come with social media datasets: their wealth of information. We can obtain an array of valuable and potentially interrelated properties from social media data. These include temporal, semantic and spatial information. Correspondingly, one should try to analyze all these dimensions simultaneously instead of considering them in a separated fashion. This might unveil a much deeper understanding of social phenomena that are reflected in such datasets. Recent research efforts like, e.g., Steiger *et al.* (2015) reflect this idea. However,

it yet remains a challenge to find measures to incorporate these different kinds of information in joint methodology in a reasonable way.

Conclusion and an outlook on future work

We outlined some potential pitfalls when analyzing social media data spatially. These are caused by the inherent characteristics of the data, i.e., the way in which the data is collected and what such services are used for. Potential problems include geometric mixtures of differently scaled data; semantic mixtures that get blurred in joint attributes derived from the data; and (more generally) spurious assessments of spatial correlations and thus pattern in the data.

The previous paragraphs are clearly biased towards the concept of spatial autocorrelation. On the one hand this focus is due to the research focus of the authors. On the other hand this is due to the central role which spatial autocorrelation plays throughout the entire field of spatial analysis. However, there are of course other important characteristics and pitfalls that might also influence the spatial analysis of social media data. The observations come, for instance, with considerable uncertainties with respect to relevant dimensions: The text snippets are colloquial and oftentimes difficult to interpret (semantics), the time stamp is sometimes not in line with real-world happenings (temporal) and the geographic coordinates are prone to positioning inaccuracies (spatial). The intensities of all these uncertainties appear to be varying across different users, devices, regions, etc. All these uncertainties indeed have impact on the results of spatial analysis.

Future methodological research should focus on the specific spatial characteristics of social media data (that are not yet known to a full extent). For now, across all disciplines and domains, it is common sense to apply established standard methodology to social media data. Relatively little emphasis is put on purely methodological research on the background of the special characteristics of these datasets. Thus, there is still plenty of room for improvement. The discipline of GIScience could play a vital role in these developments. Beyond purely empirical research, the impact of the spatial disciplines has been quite small so far. However, given that many research questions around social media are distinctive spatial ones, we should put much more emphasis on specialized spatial analysis techniques for social media.

Conclusion

On the one hand, social media data offers an array of new perspectives regarding many research questions and applications. On the other hand, however, these datasets also come with a set of issues that need to be taken into account, in particular when it comes to spatial analysis. GIScience can contribute to the

development of new spatial analysis methods for social media data. Current major issues from a GIScience perspective include:

- the need of spatial analysis methods to be adapted towards uncertain and unstructured data types from LBSN;
- the handling of geographic scale effects when analyzing social media data;
- the need for combining different methods across disciplinary boundaries (e.g. social network analysis, semantic analysis, spatiotemporal analysis), in order to better utilize all available information dimensions;
- the development of data fusion and information extraction methods that take several different data sources simultaneously into account.

This would support exploring latent patterns and sensing geographical processes from social media data in a more realistic manner. GIScience could thus contribute to answering these important geographic questions and may play a major role in the further exploration of social media data.

References

Ajao, O., Hong, J., & Liu, W., 2015. A survey of location inference techniques on Twitter. *Journal of Information Science, 1*: 1–10.

Albuquerque, J., de, Herfort, B., & Brenning, A. 2015. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographical Information Science*, pending (pending).

Andrienko, G., & Andrienko, N. 2013. Thematic Patterns in Georeferenced Tweets through Space-Time Visual Analytics. *Computing in Science & Engineering, 15*(3): 72–82.

Anselin, L., 1995. Local Indicators of Spatial Association-LISA. *Geographical Analysis, 27*(2): 93–115.

Azevedo, T., & Bezerra, R. 2009. An analysis of human mobility using real traces. *In*: *Wireless Communications and Networking Conference WCNC,* pp. 1–6.

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, pp. 61–70.

Bakillah, M., Li, R.-Y., & Liang, S. H. L. 2014. Geo-located community detection in Twitter with enhanced fast-greedy optimization of modularity: the case study of typhoon Haiyan. *International Journal of Geographical Information Science*.

Benjamini, Y., & Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J*ournal of the Royal Statistical Society. Series B (Methodological)*, 57 (1): 289–300.

Candia, J., & González, M. 2008. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical, 41*(22).

Chae, J., Thom, D., Bosch, H., Jang, Y., Maciejewski, R., Ebert, D. S., & Ertl, T. 2012. Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In: *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 143–152.

Cheng, Z., Caverlee, J., & Lee, K. 2010. You Are Where You Tweet : A Content-Based Approach to Geo-locating Twitter Users. In: *Proceedings of the 19th ACM international conference on Information and knowledge management,* pp. 759–768.

Cheng, Z., Caverlee, J., Lee, K., & Sui, D. Z. D. 2011. Exploring Millions of Footprints in Location Sharing Services. *ICWSM*: 81–88.

Cho, E., Myers, S. A., & Leskovec, J. 2011. Friendship and mobility. In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining – KDD '11*. New York, NY: ACM, pp. 1082–1090.

Cliff, A., & Ord, J. 1981. Spatial processes: models & applications.

Coffey, C., & Pozdnoukhov, A., 2013. Temporal decomposition and semantic enrichment of mobility flows. In: *Proceedings of the 6th ACM SIGSPATIAL International Workshop on Location-Based Social Networks – LBSN '13*. ACM, New York, NY: ACM, pp. 34–43.

Collier, N., Son, N. T., & Nguyen, N. M. 2011. OMG U got flu? Analysis of shared health messages for bio-surveillance. *Journal of biomedical semantics, 2*(Suppl 5): S9.

Conover, M. D., Davis, C., Ferrara, E., McKelvey, K., Menczer, F., & Flammini, A. 2013. The geospatial characteristics of a social movement communication network. *PloS one, 8*(3): e55957.

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

Crooks, A., Croitoru, A., Stefanidis, A., & Radzikowski, J. 2013. #Earthquake: Twitter as a Distributed Sensor System. *Transactions in GIS, 17*(1): 124–147.

Culotta, A., 2010. Towards detecting influenza epidemics by analyzing Twitter messages. In: *Proceedings of the first workshop on social media analytics.* ACM, pp. 115-122.

Fischer, M. M., & Getis, A. 2010. Introduction. In: Fischer, M. M., & Getis, A. (Eds.) *Handbook of Applied Spatial Analysis*. Heidelberg: Springer, pp. 1–26.

Fotheringham, A., Brunsdon, C., & Charlton, M. 2003. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.

Gallagher, A., Joshi, D., Yu, J., & Luo, J. 2009. Geo-location inference from image content and user tags. *Computer Vision and Pattern Recognition Workshops, 2009.* CVPR Workshops 2009, IEEE, pp. 55–62.

Gao, S., 2014. Spatio-Temporal Analytics for Exploring Human Mobility Patterns and Urban Dynamics in the Mobile Age. *Spatial Cognition & Computation, 15*(2): 86–114.

Getis, A., & Ord, J. K. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis, 24*(3): 189–206.

Goodchild, M. 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal, 69*(4): 211–221.

Grinberg, N., Naaman, M., Shaw, B., & Lotan, G. 2013. Extracting Diurnal Patterns of Real World Activity from Social Media. *ICWSM*.

Haklay, M., Singleton, A., & Parker, C. 2008. Web mapping 2.0: The neogeography of the GeoWeb. *Geography Compass, 2*(6): 2011–2039.

Hasan, S., Zhan, X., & Ukkusuri, S. V. 2013. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In: *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing – UrbComp '13*. ACM, New York, NY: ACM Press, p. 1.

Hiruta, S., Yonezawa, T., Jurmu, M., & Tokuda, H. 2012. Detection , Classification and Visualization of Place-triggerd Geotagged Tweets. In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM.

Hong, L., Ahmed, A., Gurumurthy, S., Smola, A.J., & Tsioutsiouliklis, K. 2012. Discovering geographical topics in the twitter stream. In: *Proceedings of the 21st international conference on World Wide Web – WWW '12*. New York, USA: ACM, p. 769.

International Telecommunication Union (ITU). 2014. *Measuring the Information Society. Report*.

Jatowt, A., Antoine, É., Kawai, Y., & Akiyama, T. 2015. Mapping Temporal Horizons: Analysis of Collective Future and Past related Attention in Twitter, pp. 484–494.

Jiang, B., Yin, J., & Zhao, S. 2009. Characterizing the human mobility pattern in a large street network. *Physical Review E, 80*(2).

Jin, L., Long, X., Zhang, K., Lin, Y., & Joshi, J. 2014. Characterizing users' checkin activities using their scores in a location-based social network. *Multimedia Systems*.

Kafsi, M., & Cramer, H. 2015. Describing and Understanding Neighborhood Characteristics through Online Social Media. In: *Proceedings of the 24th International Conference on World Wide Web.* International World Wide Web Conferences Steering Committee, pp. 549–559.

Kelm, P., Murdock, V., & Schmiedeke, S. 2013. Georeferencing in social networks. *Social Media Retrieval.* Springer London, pp. 115–141.

Kinsella, S., Murdock, V., & Hare, N. O. 2011. 'I' m Eating a Sandwich in Glasgow: Modeling Locations with Tweets. In: *Proceedings of the 3rd international workshop on Search and mining user-generated contents*. ACM, pp. 61–68.

Kling, F., Kildare, C., & Pozdnoukhov, A. 2012. When a City Tells a Story: Urban Topic Analysis. *In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems*. New York, USA: ACM, pp. 482–485.

Kosala, R., & Adi, E. 2012. Harvesting Real Time Traffic Information from Twitter. *Procedia Engineering, 50* (Icasce), pp. 1–11.

Krumm, J., Caruana, R., & Counts, S. 2011. Learning Likely Locations. In: *User Modeling, Adaptation, and Personalization*. Springer Berlin Heidelberg, pp. 64–76.

Van Laere, O., Schockaert, S., & Dhoedt, B. 2010. Towards automated georeferencing of Flickr photos. *In*: *Proceedings of the 6th Workshop on Geographic Information Retrieval – GIR '10*. New York, USA: ACM Press, p. 1.

Lampos, V., & Cristianini, N. 2010. Tracking the flu pandemic by monitoring the Social Web. *Cognitive Information Processing (CIP).* 2010 2nd International Workshop on IEEE, pp. 411–416.

Lamprianidis, G., & Pfoser, D. 2011. Jeocrowd – Collaborative Searching of User-Generated Point Datasets. *ACM*.

Lee, R., & Sumiya, K. 2010. Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. In: *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks – LBSN '10*. New York, USA: ACM, p. 1.

Lenormand, M., Tugores, A., Colet, P., & Ramasco, J. J. 2014. Tweets on the road. *PloS one, 9*(8).

Li, C., Zhao, Z., Luo, J., Yin, L., & Zhou, Q. 2014. A spatial-temporal analysis of users' geographical patterns in social media: a case study on microblogs. *Database Systems for Advanced Applications.* Springer Berlin Heidelberg, pp. 296–307.

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

Li, W., Serdyukov, P., de Vries, A. P., Eickhoff, C., & Larson, M. 2011. The where in the Tweet. *In*: *Proceedings of the 20th ACM international conference on Information and knowledge management – CIKM '11*. New York, USA: ACM, p. 2473.

Liang, X., Zheng, X., Lv, W., Zhu, T., & Xu, K., 2012. The scaling of human mobility by taxis is exponential. *Physica A: Statistical Mechanics and its Applications, 391*(5): 2135–2144.

Longley, P. A., & Adnan, M. 2015. Geo-temporal Twitter demographics. *International Journal of Geographical Information Science*: 1–21.

Longueville, B., De Annoni, A., Schade, S., Ostlaender, N., & Whitmore, C. 2010. Digital earth's nervous system for crisis events: real-time sensor web enablement of volunteered geographic information. *International Journal of Digital Earth, 3*(3): 242–259.

Lovelace, R., Birkin, M., Cross, P., & Clarke, M. 2016. From big noise to big data: Toward the verification of large data sets for understanding regional retail flows. *Geographical Analysis, 48*(1): 59–81.

Miller, H J., & Michael, F. 2015. Goodchild. Data-driven geography. *GeoJournal*, 80(4): 449–461.

Murthy, D., & Longwell, S. a. 2013. Twitter and Disasters. *Information, Communication & Society, 16*(6): 837–855.

Nelson, T. 2012. Trends in spatial statistics. *The Professional Geographer, 64*(1): 83–94.

Noulas, A., Scellato, S., Mascolo, C., & Pontil, M. 2011. Exploiting Semantic Annotations for Clustering Geographic Areas and Users in Location-based Social Networks. In: *The Social Mobile Web 11*. AAAI.

Ord, J., & Getis, A. 1995. Local spatial autocorrelation statistics: distributional issues and an application. *Geographical analysis, 27*(4): 286–306.

Quercia, D., Capra, L., & Crowcroft, J. 2012. The Social World of Twitter: Topics, Geography , and Emotions. In: *ICWSM 12*. AAAI, pp. 298–305.

Ritzer, G., & Jurgenson, N. 2010. Production, Consumption, Prosumption The nature of capitalism in the age of the digital 'prosumer'. *Journal of Consumer Culture, 10*(1): 13–36.

Roick, O., & Heuser, S. 2013. Location Based Social Networks – Definition, Current State of the Art and Research Agenda. *Transactions in GIS, 17*(5): 763–784.

Sakaki, T., Okazaki, M., & 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*. ACM, New York, NY, pp. 851–860.

Scellato, S., & Mascolo, C. 2010. Distance matters: geo-social metrics for online social networks. In: *Proceedings of the 3rd conference on Online social networks.*

Sengstock, C., & Gertz, M. 2012. Latent geographic feature extraction from social media. In: *Proceedings of the 20th International Conference on Advances in Geographic Information Systems – SIGSPATIAL '12*. New York, New York, USA: ACM Press, p. 149.

Serdyukov, P., Murdock, V., & van Zwol, R. 2009. Placing flickr photos on a map. In: *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval – SIGIR '09*. New York, New York, USA: ACM Press, p. 484.

Shimatani, K. 2002. Point processes for fine-scale spatial genetics and molecular ecology. *Biometrical Journal, 44*(3): 325–352.

Sofean, M., & Smith, M. 2012. A Real-Time Architecture for Detection of Diseases using Social Networks: Design , Implementation and Evaluation. In: *Proceedings of the 23rd ACM conference on Hypertext and social media*. ACM, pp. 309–310.

Stefanidis, A., Crooks, A., & Radzikowski, J. 2013. Harvesting ambient geospatial information from social media feeds. *GeoJournal, 78*(2): 319–338.

Steiger, E., De Albuquerque, J. P., & Zipf, A. 2015. An advanced systematic literature review on spatiotemporal analyses of Twitter data. *Transactions in GIS* (pending).

Steiger, E., Ellersiek, T., & Zipf, A. 2014. Explorative public transport flow analysis from uncertain social media data. In: *Third ACM SIGSPATIAL Interna-*

tional Workshop on Crowdsourced and Volunteered Geographic Information (GEOCROWD).

Steiger, E., Lauer, J., Ellersiek, T., & Zipf, A. 2014. Towards a framework for automatic geographic feature extraction from Twitter. In: *Eighth International Conference on Geographic Information Science*.

Steiger, E., Resch, B., & Zipf, A. 2015. Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. *International Journal of Geographical Information Science*, X(X): xx-xx.

Steiger, E., Westerholt, R., Resch, B., & Zipf, A. 2015. Twitter as an indicator for whereabouts of people ? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems, 54*: 255–265.

Stoyan, D., & Stoyan, H. 1994. Fractals, random shapes and point fields: methods of geomtetrical statistics. Hoboken: John Wiley & Sons.

Tapia, A. H., Bajpai, K., Jansen, B. J., & Yen, J., 2011. Seeking the Trustworthy Tweet : Can Microblogged Data Fit the Information Needs of Disaster Response and Humanitarian Relief Organizations. *ISCRAM*(May): 1–10.

Thomson, R., Ito, N., Suda, H., Lin, F., Liu, Y., Hayasaka, R., Isochi, R., & Wang, Z. 2012. Trusting Tweets: The Fukushima Disaster and Information Source Credibility on Twitter. *In*: *Proceedings of the 9th International ISCRAM Conference*, pp. 1–10.

Veloso, A., & Ferraz, F. 2011. Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. In: *Proceedings of the 3rd International Web Science Conference*. ACM, 3.

Vieweg, S., Hughes, A., Starbird, K., & Palen, L. 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1079–1088.

Wakamiya, S., & Lee, R. 2012. Crowd-sourced Urban Life Monitoring: Urban Area Characterization based Crowd Behavioral Patterns from Twitter Categories and Subject Descriptors. In: *Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication*. ACM, 26.

Wakamiya, S., Lee, R., & Sumiya, K. 2011. Crowd-based Urban Characterization: Extracting Crowd Behavioral Patterns in Urban Areas from Twitter. In: *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*. ACM, New York, NY: ACM, pp. 77–84.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S., 2012. A System for Real-time Twitter Sentiment Analysis of 2012 U . S . Presidential Election Cycle. In: *Proceedings of the Association for Computational Linguistics 2012 System Demonstrations*. ACM, pp. 115–120.

Westerholt, R., Resch, B., & Zipf, A., 2015. A local scale-sensitive indicator of spatial autocorrelation for assessing high- and low-value clusters in

multiscale datasets. *International Journal of Geographical Information Science* (pending).

Wu, F., Li, Z., Lee, W., Wang, H., & Huang, Z. 2015. Semantic Annotaion of Mobility Data using Social Media. *Proceedings of the 24th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee,* pp. 1253–1263.

Wu, S., Hofman, J. M., Watts, D. J., & Mason, W. A. 2011. Who Says What to Whom on Twitter. In: *Proceedings of the 20th international conference on World wide web*. ACM, pp. 705–714.

Yardi, S., & Boyd, D. 2010. Tweeting from the Town Square: Measuring Geographic Local Networks. In: *ICWSM 10*. AAAI.

Zandbergen, P. a., & Barbeau, S. J. 2011. Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones. *Journal of Navigation, 64(*03): 381–399.

Zook, M., Graham, M., Shelton, T., & Gorman, S. 2010. Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake. *World Medical & Health Policy, 2*(2): 6–32.