

Location Based Social Networks – Definition, Current State of the Art and Research Agenda

Oliver Roick* and Susanne Heuser

Abstract

The growing penetration of GPS equipped smartphones allowing users to constantly share geographic information on their current whereabouts has lead the way to an augmentation of existing *Social Network Sites* with location based features or the creation of new ones exclusively around geographic information. Within these *Location Based Social Networks* vast amounts of geographic information are allocated, which attracted the attention of researches with various scientific backgrounds. This paper presents a comprehensive definition of this special type of *Social Network Sites* and an overview of research activities, which are currently conducted using the data. Finally, a research agenda for GIScience is drawn from existing research directions.

I Introduction

The recent ten years have seen a tremendous paradigm shift of how information on the web is created, maintained and used. Based on technological developments – such as the vast penetration of broadband internet access coinciding with new web-based software, allowing users to participate in the creation of content for the Web – the role of the user changed from a solely consumer perspective to that of a consumer and producer; a mantle that has been

*Heidelberg University, Chair of GIScience, Berliner Straße 48, 69120 Heidelberg, E-Mail: roick@uni-heidelberg.de

coined as *prosumer* (TABSCOTT 2001). Since then, users of the Web have produced enormous amounts of *User Generated Content* including different types of media like photos on Flickr and videos on Youtube, blog posts or product descriptions and ratings. These developments have been subsumed under the buzzword *Web 2.0* (O'REILLY 2005).

These developments further changed the way geographic information is acquired, maintained and distributed through the Web. The growing availability of affordable GPS devices enabled amateur users without any formal qualifications to survey geographic data. Easy-to-use and mostly free open-source software further allowed these users to distribute these data through the Web. This new type of user-generated content drives successful collaborative projects – such as OpenStreetMap or WikiMapia. Big companies – e.g.; Google, Apple or Nokia – further employ user-generated geographic content to improve their services; for instance, by updating their base maps based on information retrieved from their users or to extract real-time traffic information. This often happens without explicit knowledge of the user as the algorithms are integrated in the phone's operating system by default.

However, the rise of the Web 2.0 also yielded new means for humans to communicate, interact, share information and maintain relationships with acquaintances over the Web. *Social Network Sites* play a key role in this evolution (BOYD & ELLISON 2008) and have therefore attracted millions of users worldwide. Facebook – for instance – had a daily average of more than 584 million¹ active users in September 2012.

The advent of smartphones, equipped with sensors allowing users to permanently locate themselves geographically, marks another milestone in this development. Online content gets more and more enriched with geographic information that constitutes a new layer of context and is employed to organize and display information (GORDON & DE SOUZA E SILVA 2011). This development led the way to a conversion of GIS and social media resulting in an augmentation of existing *Social Network Sites* with new location-based capabilities; e.g., Facebook or Twitter, and the creation of new ones exclusively around location-based information, like Foursquare (SUI & GOODCHILD 2011).

In these special type of *Social Network Sites* vast amounts of geographic information are allocated over time, which can be accessed through public *Application Programming Interfaces (API)* and have attracted the attention of researchers with varying scientific backgrounds. This paper attempts to summarize scientific research efforts around user generated geographic information in *Social Network Sites*.

¹According to Facebook Fact Sheet (<http://newsroom.fb.com/content/default.aspx?NewsAreaId=22>)

First, the nature of this new type of user generated content is outlined by introducing the concept of *Location Based Social Networks (LBSN)*. Then, an overview of the current state of the art in research relating to LBSN is given. It can be shown that research up to now focuses mostly on topics in social sciences, such as the formation of social networks and human behavior in space and time. Despite the fact that data from LBSN provide an interesting source for research in GIScience there are still few studies focussing on the exploitation and application of such information using methods of GIScience. We conclude by deriving open research questions for the GIScience community.

2 What are Location Based Social Networks?

An early description of web-based services that allow internet users to construct (semi-)public profiles, articulate lists of acquaintances and provide means of communication among users through comments, private or instant messaging as well as sharing digital media, such as photos or videos, has been given by BOYD & ELLISON (2008) and termed *Social Network Sites*. This definition encloses the disclosure of personal information and the contribution of content to the web. However, in order to embrace the new developments regarding location-aware features in *Social Network Sites* the definition given before has been extended as follows:

Social Network Sites that include location information into shared contents are called *Location Based Social Networks (LBSN)*. These online networks display geographic information on a map or as a list of status updates ordered by geographic proximity as opposed to the traditional concept reverse chronological ordering (GORDON & DE SOUZA E SILVA 2011). This phenomenon has also been named *Locative Mobile Social Networks*. It is defined as commercial applications, available through a cell phone, that help building networks of mobile nodes (in this case moving people) by displaying the geographic position of users on a map. Users are enabled to locate each other in physical space and interact with one another depending on the relative distance (DE SOUZA E SILVA & FRITH 2010).

In general, there are two ways of how geographic information can be shared in *Location Based Social Networks* (ELWOOD ET AL. 2011): First, the annotation of location information to digital artifacts, also known as *geotagging* (TURNER 2006). This turns photographs, videos, blog posts or tweets on Twitter² into geographic information. A prominent example for *geotagging* is the photo sharing site Flickr³, which – if available – retrieves geographical information from

²<http://twitter.com>

³<http://flickr.com>

the image's EXIF data or otherwise allows users to attach geographic coordinates by selecting the respective position by simply clicking on a map (A more detailed discussion of different ways to annotate geographic information can be found in the following section). This geographic information can then be used in a local search or by putting popular photos on a map (see figure 2).

Second, the sharing of activities along with current whereabouts. For instance, users on Foursquare⁴ may check-in at a certain place and thus share that information with friends. This is referred to as *geosocial networking*. Such *geosocial networking* applications are often equipped with gaming features to motivate users to constantly share their location. Foursquare awards every check-in with a certain amount of points, which are allocated over time. Users further earn badges when they check-in at several similar venues such as train stations or coffee shops (see figure 1). The number of points and badges allow for a comparison of a user's activity and experience with his or her friends on the network site (GORDON & DE SOUZA E SILVA 2011). Information – such as the number of check-ins at a certain venue, the number of individual visitors, likes and tips – further indicate the popularity of that specific place and can be employed to improve a user's experience of the city.

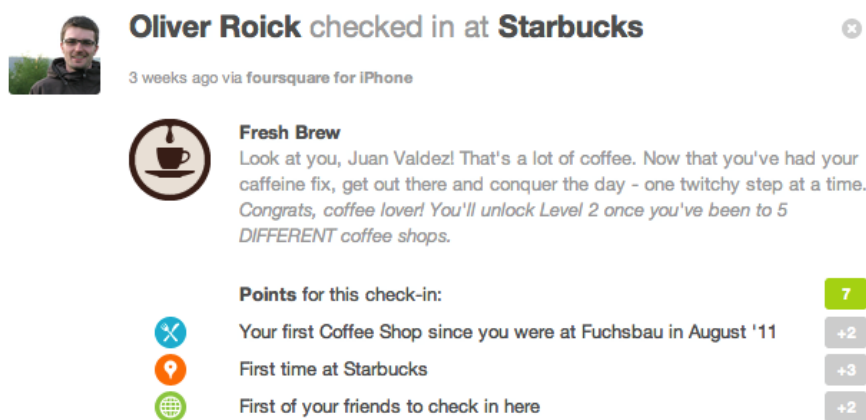


Figure 1: Example of a checkin on Foursquare. The user earned points for different reasons and unlocked a badge for visiting coffee shops.

Scholars in media geography also approached the topic, but from a distinct perspective. They differentiate how media and location context are interrelated. The term *locative media* is associated with the annotation of geo-spatial

⁴<http://foursquare.com>



Figure 2: A map depicts the locations of georeferenced images on the photo sharing site Flickr.

information to digital artifacts; whereas *mediated localities* refer to traced actions of a subject in the world, for instance applications that display user's geographic positions on a map. Both are subsumed under the umbrella of *geo-media* (THIELMANN 2010).

The concepts discussed here are interrelated as depicted in figure 3. Within LBSN users share geo-referenced information in different ways. *Geotagging* results in digital artifacts that are referenced to a location in the real world and thus can be seen as *locative media*. On the other hand, people, who share their location on a *geosocial networking* site, can be located at a certain place at a certain time. Hence, they are considered as *mediated localities*.

As with classic *Social Network Sites*, LBSN focus on a multitude of different user interests and activities resulting in a wide variety of applications and usage types. In order to provide an overview on different usage forms and to structure the market of LBSN applications, several authors strived to categorize existing LBSN. Based on the primary motivation to share a location, LINDQUAIST ET AL. (2011) and TANG ET AL. (2010) distinguish between *social-driven* and *purpose-driven* LBSN. In *social-driven* LBSN the sharing of the current whereabouts is mainly motivated by self-expression and socializing. Users share their location mostly out of fun and to build and maintain their social network. The present location is usually shared with a large group of people. Contrary, users of *purpose-driven* LBSN share their location with a selected group of a few acquaintances or sometimes only one person, because location

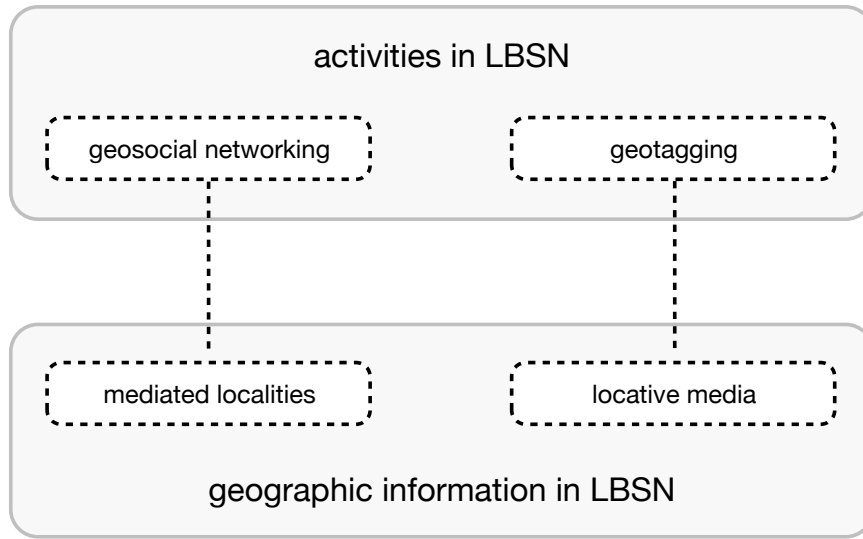


Figure 3: Activities in LBSN and corresponding type of geographic information.

sharing always follows a specific reason; e.g., telling a family member that one arrived safely after a travel. Another classification is based on the primary usage purpose. SUI & GOODCHILD (2011) differentiate between *Social check-in sites*, *Social review sites* and *Social scheduling/events sites*. However, a further description of the categories and the distinguishing characteristics are not given by the authors.

3 Assigning geographic information to digital artifacts

There are several different ways of how position information can be captured in order to assign this information to content on the Web. Today's phones are almost always equipped with GPS receivers that allow to track a user's position with a relatively good accuracy. However, in some cases a satisfactory GPS signal might not be available. This especially holds true for urban canyons and inside buildings. So alternative positioning methods might be applied: Triangulation of wireless signal strength or the current cell of the mobile network. Both methods have advantages over GPS when the latter falls short: The density of cellular network antennas in highly populated areas reduces the potential area in which a user might be location and thus increases the accuracy of

the location detection (MOUNTAIN & RAPER 2001).

In certain online networks – such as the photo sharing site Flickr – a geo-location can be attached to a document by tagging the correct position on the map. The accuracy of the position is largely dependent on completeness and accuracy of the underlying base map and the user’s ability to read and interpret that map as well as her familiarity of the area.

Then, geo-coding of addresses is the third option to retrieve the geo-position. The quality of the geocoding algorithms and the quality of the underlying database are factors influencing the result and therefore the quality of the tag.

The different methods discussed here determine the quality – especially the accuracy – of the location information to the digital artifact. Information on the method applied to tag a position have to be taken into account when assessing the quality of a data set or when dealing with user-generated information and applying it to application scenarios.

4 Current Research on Location Based Social Networks

Location Based Social Networks are a relatively young phenomenon and yet have gained attention of scholars from different scientific backgrounds. Topics, which have been addressed, predominantly include social relationships and the influence of geographic proximity, the structure of social networks and privacy issues. Investigations on locating users and georeferencing digital content to close gaps in availability of geographic information, issues in data quality and information retrieval have been further conducted towards making data in LBSN exploitable and usable.

The following section provides an overview of the research approaches related to *Location Based Social Networks*.

4.1 Location Based Social Networks and Social Sciences

For the first time ever, a comprehensive collection of information on humans their social ties and behavior in space and time, compiled from data shared in *Location Based Social Networks*, is available for investigation. This has been referred to as *deep data about the many*, as opposed to *deep data about the few* versus *surface data about the many* (SUI & GOODCHILD 2011). This massive amount of information opens new possibilities for social sciences to research and understand the emergence of social networks in a spatial context. Further, this spatio-temporal data allows for studying human behavior and social dynamics

and draw insights on human mobility as well as perception and shape of urban areas (CRANSHAW ET AL. 2012).

4.1.1 Analyzing and Predicting Social Ties

The invention of new communication technologies, such as the telephone or the World Wide Web, influenced the way humans communicate over large distances. These *mediated relationships* (WERLEN 1995) lead to a perceived compression of space and time (HARVEY 1989) since social interaction is no more limited to face-to-face interactions. The advancement and wide penetration of broadband internet in general and *Social Network Sites* in special have further influenced the way humans communicate to each other. *Social Network Sites* enable humans to form and maintain social ties even over long spatial distances. Based on geographic information from online social networks, several scholars attempted to investigate these effects. It has been found that social networks that are formed online mostly spread over a geographically small area and form local clusters. However, long-range ties occur more often than in offline networks (SCELLATO ET AL. 2011a) and are even more likely in content sharing communities, such as LiveJournal or Twitter (SCELLATO ET AL. 2010). Other studies demonstrated the importance of a common language for the formation of social networks: English-speaking members of social network tend to have larger networks than those members with other native languages (GONZALEZ ET AL. 2011). TAKHTEYEV ET AL. (2012) present similar results regarding the structure of social networks and the influence of language on social ties and further find – while frequent airplane connections between locations may stimulate its initiation – social ties are still predominantly constrained to national borders. These local constraints may also be the result of the fact that a small number of countries account for most of the activities on social networks, while others are mostly consuming information (KULSHRESTHA ET AL. 2012).

Drawing from the conclusion that spatial proximity is still a dominating factor for the structure of social networks, several authors attempted to predict social ties based on location information of users. The main assumption here is that users, who often can be found at places close to each other, have a higher probability of being friends. This probability is even larger when two individuals visit the places at the same time. CRANDALL ET AL. (2010) therefore analyzed geo-referenced Flickr images and the time the respective photograph was taken and develop a probabilistic model that predicts the probability of two persons being friends based on the number of co-occurrences within a predefined geographic area and a given time-frame. Co-location is also a fundamental concept of another study by CRANSHAW ET AL. (2010), which employs

location information from the online social network Locaccino⁵. The authors derive several model variables describing the intensity and duration of the observed co-locations, location diversity and the mobility of the co-located users. Based on these variables the friendship of two co-located persons is predicted using different trained classifiers. In a similar study, SCELLATO ET AL. (2011b), used data of users and their check-ins from the now retired LBSN Gowalla of four consecutive points in time allowing for evaluation and improvement of the methodology at each time step. The authors also derived a set of model variables describing both visited places and social ties the friendship candidates have in common. Further, different machine learning algorithms to predict the friendship between both persons are applied.

The discussion has shown that there are currently two main research directions regarding the emergence of social ties in *Social Network Sites*: First, the analysis of dominating factors on the formation of ties, predominantly geographic distance and language. Second, the prediction of social ties between members of an online social network community based on relative geographic distance on the one hand and further based on the number of co-location occurrences.

4.1.2 Analyzing Human Behavior in Space and Time

Continuously shared location information can be further employed to analyze human behavior and derive consolidated findings on human mobility patterns and the structure of urban areas. An investigation of spatio-temporal activity patterns using Foursquare data reveals expected findings: Temporal check-in patterns differ between weekdays and weekends. Further, the predominant venue categories vary throughout the day with traffic related venues in the morning shifting to *home* as the main check-in category in the evening (NOULAS ET AL. 2011). The according movement patterns have also been investigated: Movements follow the *Lévy Flight Pattern* with many sequential short distance moves and interrupted by occasional long distances travels (CHENG ET AL. 2011). Movement patterns have also been derived using sequences of photographs taken and uploaded to Flickr. Using geovisualization techniques GIRARDIN ET AL. (2007) analyze activity patterns of Tourists in the area of Florence, Italy.

Spatio-temporal information from social networks has also been applied to investigate on the structure of cities by analyzing the distribution of human activities. Based on a sample of Foursquare check-ins in New York, London and Paris, BAWA-CAVIA (2011) exemplify emerging *social hubs* as areas of a city

⁵<http://locaccino.org/>

with a remarkably high user activity. Whereas in New York most activity focusses on Manhattan and the east of Brooklyn, in London and Paris these *hubs* are more dispersed throughout the city. The distribution on London shows polycentric processes while in Paris the activity pattern is more equally distributed over the city area evidencing fragmentation of the city. CRANSHAW ET AL. (2012) further employed checkins on Foursquare to derive local clusters representing characteristic neighborhoods of a city. Using a spectral clustering algorithm the approach takes both spatial and social proximity into account and therefore finds areas that are distinct from municipal district borders but rather represent so called *Livehoods*. The results have been examined using semi-structured interviews with residents and mostly confirmed shape and size of the *Livehoods*. Similar, HOLLENSTEIN & PURVES (2010) derive colloquial names of urban areas from *tags* of georeferenced photos on Flickr and therefore demonstrate that cities are perceived and structured by humans independent from municipal borders. Using Flickr data, SUN ET AL. (2012) investigate the spatial distribution of tourist accommodations in the city of Vienna. By applying point pattern analysis and spatial cluster detection, the authors identify a seasonal variation of tourist accommodation distribution across the city. A comparable approach is presented by SCHLIEDER & MATYAS (2009) who apply clustering of geo-referenced tourist images from Panoramio to identify Points of Interest within a city and propose different measures to define the popularity of these POIs by taking the number of images and the number of contributors per cluster into account.

4.2 Inferring Location from Implicit Information

4.2.1 Geo-locating Pieces of Content

In order to geographically exploit information from LBSN precise geo-referenced information is a crucial basis. However, especially in such *Social Network Sites* that are focused on the exchange of media an explicit geographic reference is sometimes not available. In some cases, these information can be extracted from explicit location provided by latitude-/longitude-coordinates or inferred from textual contents using gazetteer services such as Geonames⁶ (BOUILLOT ET AL. 2012).

Rather often this explicit information might not be available or is given in an imprecise or blurred way. Methods for inferring the location of the referenced artifact are therefore required. Scholars have mostly focused on the LBSN Flickr and Twitter to infer the location for a piece of digital content

⁶<http://geonames.org>

from implicit information applying different methods from content analysis over machine learning to clustering.

For the case of Twitter, GELERNTER & MUSHEGIAN (2011) applied Named Entity Recognition to extract location information such as country, city or street names as well as names of buildings to automatically derive a location reference to the respective tweet. The method located only 34% of the investigated tweets correctly, which is mostly due to spelling errors or the use abbreviations.

Other studies apply different approaches to locate images posted to Flickr based on keywords attached to the images. SERDYUKOV ET AL. (2009) compose a language-based model from the user annotated *tags*, which computes a list of possible locations for a given tag set ranked by likelihood. A more sophisticated method for geo-referencing Flickr images employs clustering. VAN LAERE ET AL. (2010) use k-medoids clustering to derive a certain number of areas with a typical set of tags in selected cities with notable tourist attractions in Europe. Using a Naive Bayes classifier the authors assign photos from a test data set to predefined areas. According to the authors the results are reasonable and can be improved by applying smoothing. This work has been extended and applied by JOSHI ET AL. (2010). The authors further consider tag entropy with respect to location and find that tags with low entropy clearly indicate place names such as cities or nations that can be also utilized for georeferencing images based on tags. In another study visual image analysis methods are employed additionally (GALLAGHER ET AL. 2009).

4.2.2 Locating Users

Knowledge on a user's location can be applied to provide contextual information tied to the specific location or to add a location reference to a piece of content shared in the online network. Two fundamentally different approaches to infer a user's location have been presented: Applying known locations of a user's friends and inferring the location through analysis of the contents posted to the web.

DAVIS JR. ET AL. (2011) developed a method for inferring the home location of a Twitter user from known locations of the user's followers by selecting the most popular one among the friends locations. In order to achieve reasonable results the authors constrained their approach to users with a certain minimum and maximum number of friends and a minimum number of occurrences of a certain location to be considered as valid. BACKSTROM ET AL. (2010) try to locate users on Facebook with an unknown home town by taking home addresses of the user's friends into account. Based on an analysis of the relationship between distance and friendship the authors calculate the likelihood

for a given location by calculating the friendship probability of a given edge as a function of distance and then weighting the known location accordingly.

Contrary, CHENG ET AL. (2010) analyzed the content of tweets of single users for words that are characteristic for certain places and apply maximum likelihood methods in order to anticipate the city the respective user lives in.

The approaches discussed here all show moderately good results. BACKSTROM ET AL. (2010) were able to locate 67,5% users within an 25 mile radius of their home town. This number is largely depending on the number of friends a user has as the accuracy of the localization drops with decreasing number of friends. CHENG ET AL. (2010) located about 50% of tested users correctly with an average error distance of about 535 miles. These results show that the proposed approaches provide a good starting point for predicting user locations but have to be improved in order to provide more reliable results.

4.3 Quality Assessment

Notwithstanding its importance, quality assessment of crowdsourced geo-information in the context of LBSN is one of the topics that have been approached least yet. The positional accuracy of geo-tagged photographs on Flickr and Panoramio has been investigated by HOCHMAIR & ZIELSTRA (2012). It has been found that Panoramia is more accurate when compared to Flickr. Panoramio further shows a larger homogeneity, while the accuracy on Flickr seems to be dependent on the object type being pictured. However, the approach has a major methodological drawback: References are provided by an analyst, who estimates the position of the photographer based on the pictured frame and focal length. The quality of this estimation is highly dependent on the local knowledge of the analyst and the base maps used to geo-reference the estimated position. Hence, the results of this study have to be treated carefully. The study also raises questions regarding the concept of quality assessment as fitness for use investigation. The authors identify *footprint mismatch error* that occurs when users geotag the object being pictured instead of the photographers position as a potential cause of error. Whether this actually is an error depends on the use case. If one wants to find images of a certain landmark, then the position of the photographed object will provide a more suitable context for location search than the photographer's position.

Since there is only one study regarding the quality of data from LBSN it is obvious that there are manifold open research question regarding this topic. In order to draw a more exhaustive picture of data quality in LBSN data sets other than Flickr and Panoramio have to be investigated. Especially the venue databases of Facebook and Foursquare provide an interesting foundation for further research or applications.

For other crowdsourced geodata beyond LBSN there are several studies. In particular the quality of OSM has been investigated recently (HAKLAY 2010; GIRRES & TOUYA 2010; NEIS ET AL. 2012). An in-depth review of these studies is out of scope of this paper but the methods proposed in these studies may be suitable for investigation on the quality of geographic data in LBSN. Therefore, investigations on the applicability and appropriate potential improvements of these methods are required.

4.4 Extracting Information from Geo-referenced Digital Content

Social networks further pose a valuable data source in different application contexts. Especially, in time critical scenarios such as disaster response and the monitoring of diseases geographic information from social networking sites can provide useful data. Both topics have emerged as the most researched and discussed in scientific literature.

4.4.1 Disaster response

Monitoring the stream of Twitter messages allows for the extraction of tweets related to certain events. Comparisons between media coverage of the Sichuan Earthquake in 2008 and Twitter messages related to the event revealed, that reports on Twitter appear within seconds and a large number of tweets is posted temporally close to the event, while media coverage usually starts a day after. In fact, the earthquake was even reported before the USGS detected the incident (LI & RAO 2010).

These advantages have been adopted in a study by EARLE ET AL. (2011) who scan the contents of tweets for the keyword earthquake and try to detect occurrence and magnitude of an earthquake event by comparing long term keyword frequency with the short term frequency. Depending of the sensitivity of the trigger threshold they were able to detect a significant number of earthquakes while still having a large number of false positives. According to the authors, false positives can be prevented by utilizing natural language processing algorithms, which can identify the context of the tweet. A more sophisticated approach is proposed by SAKAKI ET AL. (2010), who apply probabilistic models to detect the occurrence of earthquakes and estimate the center of origin using Kalman or particle filtering. Compared to EARLE ET AL. (2011) the number of correctly detected earthquakes is increased dramatically. The approach is further extended to detect typhoons and estimate their most probable trajectory of movement.

In the aftermath of a natural hazard near-time situational information is useful for command and control operations to coordinate disaster response. In an empirical study, VIEWEG ET AL. (2010) analyze contents of tweets with respect to geographical location and useful information in hazard aftermath. They show that these messages provide reasonable geo-coded information, which can be used to guide disaster response. However, an approach for automatic extraction of this information is missing. The main problem here is how to distinguish useful information from the stream of incoming messages related to the event. One option would be to introduce a standardized structure using *hashtags* within a message (STARBIRD & STAMBERGER 2010). While this approach allows for automatic processing and information extraction of the data, it can be questioned if this scheme will be adopted by the public.

Strategies of how relevant information can be extracted from the stream of Twitter messages are thus required. It further needs to be researched how the relevant information can be presented to assist disaster response operators. This includes how these informations can be aggregated spatially and thematically as well as how the information can be visualized.

4.4.2 Monitoring Diseases and Health

Twitter users tend to share personal information, such as health conditions. Crawling the stream of Twitter messages and analyzing their content for appropriate keywords can extract information on local occurrences of certain ailments and their spread in space and time.

Two studies reveal a correlation of the number of influenza related messages to statistics from official agencies. The methods presented here are based on sets of classified Twitter messages providing an indication of influenza. COLLIER ET AL. (2011) detect influenza break-outs in a similar fashion to Google's Flu Trends. The results show a moderate correlation to the observations from the Center for Disease Control and Prevention (CDC) but the authors also state that further research is required in order to achieve as precise results as Google Flu Trends. CULOTTA (2010) demonstrated that multiple regression analysis against the number of twitter messages is a suitable method to detect influenza break-outs.

Based on these findings scholars are able to predict disease break-outs using Twitter data. LAMPOS ET AL. (2010) introduce an influenza score for selected cities in the United Kingdom, which is ground truthed with data from Health Protection Agency. RITTERMAN ET AL. (2009) on the other hand, use prediction market models to forecast the beginning of the swine flu pandemic in 2009. However, all approaches discussed here do not include geographical information. Twitter most often provides some sort of geo-referencing to

the tweets or the location reference can be inferred using methods discussed in section 4.2. Considering geographic information would allow for complementing the studies to detect the places of origin as well as monitor the spread of a disease throughout space and time. First steps in this direction have been accomplished by PAUL & DREDZE (2011) and GOMIDE ET AL. (2011). Based on the probabilistic *Ailment Topic Aspect Model* (ATAM) PAUL & DREDZE (2011) analyzed the occurrence of messages related to allergies and found that known patterns of allergy spread throughout a year can be reproduced using Twitter data. Using a linear regression model to predict the intensity of dengue fever at certain locations and an ST-DBScan algorithm to identify areas with large amount of dengue occurrences, GOMIDE ET AL. (2011) propose an application that allows for monitoring the disease in Brazil in both space and time.

The section showed that information mining from LBSN is still in its early stages. Event detection for earthquakes and diseases through data from LBSN is possible and has to be further refined. Open issues regarding monitoring and extraction of relevant information remain.

It has been demonstrated that Twitter is a potential data source for retrieving hazard related information but the problem of extracting relevant information has not been solved. Further, monitoring of diseases has been done in temporal context, but is far from being comparable to results of Google Flu Trends or authoritative statistics. However, its potential has already been demonstrated. The geographic realm, mainly the spread of diseases, has been studied only marginally. Therefore, further research especially coming from a GIScience perspective is required.

4.5 Privacy

By introducing a new layer of context to information shared in online social networks, LBSN offer new possibilities of how urban spaces are experienced and social relationships are formed and retained. However, the disclosure of personalized geographic information on the web is often perceived as a threat each individual's privacy (GORDON & DE SOUZA E SILVA 2011). Privacy related issues have therefore been subject to intense media coverage and also gained the attention of scholars identifying and classifying privacy threats and proposing approaches how privacy can be protected while the functionality of LBSN is ensured.

Disclosing of a person's location at a certain time can happen on different ways. First, a user intentionally shares her location to inform acquaintances on her whereabouts and availability. Second option, which currently grows in popularity among Facebook users, is the disclosure of a person's current location by a friend. The second one is often perceived as a threat to privacy because

the person loses control over the information on herself shared online. Based on this assumption RUIZ VICENTE ET AL. (2011) identify four specific classes of privacy threats: *Location privacy*, which refers to a user knowing that another person is located at a certain place at a certain time. *Absence privacy* refers to the opposite situation of knowing that a person is not at a certain place; e.g., not being at home. Inferring from the locations of other users that a person is currently with somebody else is considered as an infringement to *co-location privacy* and *identity privacy* denotes to a threat of privacy where the otherwise disguised identity (e.g. by using a pseudonym) of a user can be inferred from the identity of other co-located users. Other concerns are related to the problem of boundary preservation, e.g. others being able to constantly track a user's position as it is possible in Google Latitude (PAGE ET AL. 2012).

Despite privacy being a topic many people are concerned about – though this clearly a phenomena that depends on provenance, social background and age – it has been found that it seems difficult for users of LBSN to articulate their privacy preservation rules. SADEH ET AL. (2009) therefore develop a system, which allows users to audit the disclosed location in response to a location request and helps adjusting their personal privacy policies accordingly. The study demonstrated that being able to review and adjust personal privacy policies users are able to better understand in which cases their current location is requested by others and how their privacy settings work.

According to user defined privacy settings, approaches have been developed in response to privacy threats discussed above. FRENI ET AL. (2010) formalize the problem of *location privacy* and *absence privacy* infringement and implement a cloaking algorithm, which is considered to be placed between the social network service and the user. The algorithm generalizes the current location of a user, for instance at a city wide level, or delays the publication of the user's location. These aspects are – on a more theoretical level – further discussed by RUIZ VICENTE ET AL. (2011). The suggested privacy preservation concepts are complemented by an extension of spatial and temporal scope in a way that the number of persons within a certain area and time span is large enough to prevent others from identifying possible co-located users. This can be applied to protect *co-location* and *identity privacy*. The problem of preserving co-location privacy in LBSN has been addressed by CAMILLI (2012). Similar to FRENI ET AL. (2010) the problem of *co-location* privacy is formalized and an algorithm for its preservation is proposed.

Users are further concerned with the trustworthiness of service providers. A problem, which is tackled in the work of MASCETTI ET AL. (2011), who introduce two protocols for preserving privacy that encrypt location information in order to prevent friends and service providers from gaining knowledge on user related geographic information.

To summarize, there are studies that investigate on current privacy settings in LBSN and how users can adjust these settings according to their personal privacy policies and attempts to technically protect privacy based on the requirements provided by users. Despite the promising approaches discussed here, there are still questions that have not been answered. It is not clear to which extend spatial and temporal cloaking will limit the benefit of a service. Further, questions such as who operates a cloaking service and how save encryption are methods need to be addressed.

5 Towards a GIScience Research Agenda Regarding LBSN

The previous section has demonstrated that there is already an impressive amount of research activities regarding *Location Based Social Networks*. Despite the huge potential of user-generated data from LBSN, only few studies demonstrated how these data could be exploited and used in certain applications and use cases. CRAGLIA ET AL. (2012) argue that social media providing access to crowdsourced geographic knowledge may play a key role when implementing Al Gore's vision of Digital Earth (GORE 1999). SUI & GOODCHILD (2011) therefore propose the idea of a global data service that retrieves, stores and provides access to geospatial data from LBSN and assists users in the discovery of relevant information.

However, before such a global data service can be implemented and before data from LBSN can be fully exploited and applied to different use case scenarios, several open issues need to be addressed. Possible research directions regarding crowdsourced geographic information from LBSN have been identified by SUI & GOODCHILD (2011) and ELWOOD ET AL. (2011). From a GIScience perspective these objectives include questions of how to access the datasets, find relevant information and disseminate these information to the user. This requires research into methods for conflating and synthesizing geographic data between sources of user generated content and authoritative data sets to fill gaps in data availability. As with all crowdsourced information, questions regarding data quality are of utmost importance. Methods to assess the data quality with respect to application scenarios need to be developed. Further, approaches of how to present these informations to the user are required, including new ways of visualizing different types of media with respect to geographic location a needed. Beyond that, social implications of these new type of Location Based Services need to be researched. This includes privacy issues, the debate in who owns and controls the data and how to cope with the

problem of digital divide. SUI & GOODCHILD (2011) additionally ask for new GIScience theories on ontologies and visualization techniques for networks and formalization of place in geographic information systems.

Drawing from these suggestions three main research directions can be identified as depicted in table 1: New theories for GIScience, social sciences and information extraction. By relating objectives defined by SUI & GOODCHILD (2011) as well as ELWOOD ET AL. (2011) open research questions can be derived.

Table 1: research-Comparison of objectives as defined by SUI & GOODCHILD (2011) and the research objective already addressed as discussed in section 3.

Research objective	SUI & GOODCHILD (2011)	Objectives addressed
Social aspects		
Spatial dynamics of human behavior	•	•
Privacy	•	•
Digital divide	•	
Analysis of social ties		•
Prediction of social ties		•
Urban structures		•
Information extraction		
Data conflation	•	
Extraction of relevant information	•	•
Quality	•	•
Location inference		•
Event detection		•
GIScience theories		
Formalization of place	•	
Network-based ontologies	•	
Representation of mulitmedia	•	

Section 4 has shown that research based on data from LBSN is predominantly focussed on questions related to social sciences including the analysis and prediction of social ties and the investigating human behavior in space and time. Further, issues of privacy have been addressed extensively, though we have shown that there are still open issues.

Still a huge part of the objectives proposed by SUI & GOODCHILD (2011)

and ELWOOD ET AL. (2011) have not been investigated yet. Issues of extraction of relevant information and data quality have been approached to a very small extent whereas the objective of conflation of data from LBSN has not been studied at all. Further, issues on control of data, the representativeness of data as well as new concepts for GIScience have not been discussed as well.

Therefore, the remainder of the section discusses open research issues regarding these objectives.

5.1 Understanding types of LBSN and their data

As the previous section 4 shows, only a few LBSN with geographic information have been subject to scientific research. Studies are predominantly focused on Twitter and Flickr, whose APIs are publicly accessible and not restricted by request limits.

However, there is a large number of *Location Based Social Networks* that focus on use cases different from content sharing and therefore provide different data and information that can be exploited. The types of these online social networks range from check-in services, such as Foursquare, over networks for sportive activities (e.g.; MapMyTracks⁷ or runtastic⁸) to mobile social games (Gbang⁹).

The multitude of different networks needs to be investigated and categorized to provide an overview on existing types of LBSN. Further, it needs to be examined which data is available in the online network and how we can access and use that data. Depending on the nature of the specific online network, the data types will vary accordingly. According to our preliminary investigations, Foursquare and Facebook Places primarily provide information on Point of Interest with the according attributes, such as address or opening hours, as well as information on the number of visiting users, whereas applications such as MapMyTracks are focussed on sportive activities and will therefore provide linear data on hiking or biking routes with difficulty levels or estimated time duration attached as attribute data.

Another critical concern in data availability is the distribution of users activity across space. Figure 4 depicts the distribution messages on Twitter and photo posted to Flickr and clearly indicates different distribution patterns of user activity. Data availability depends on the area being investigated. Therefore, global studies on user behavior and its relation to the spatial distribution of information, which is available online, are needed.

⁷<http://www.mapmytracks.com/>

⁸<http://www.runtastic.com>

⁹<http://gbanga.com/>



Figure 4: Distribution of user activity in online networks Twitter (blue) and Flickr (red) across Europe. (FISCHER 2011)

Once we know about the spatial data provided by LBSN and how to access these, other objectives need to be addressed when exploiting the information. These objectives include questions regarding the quality of these user-generated spatial data sets and the way the data can be integrated with exiting data and into applications. These objectives will be discussed in the following subsections.

5.2 Data Quality

Geographic data collected by untrained and often inexperienced volunteers always should be subject to critical quality control before employing these data in research projects and applications. Therefore, the assessment of data quality is one of the most crucial open research problems. In order to describe the quality of geo-spatial data, several elements of geo data quality have been defined: Lineage, positional and attribute accuracy, logical consistency, completeness and temporal quality (VAN OORT 2006). In terms of *Volunteered Geographic Information* other authors suggest to extend these elements and emphasize the experience of the contributing user (*user quality*) and the lineage of the respective feature (*feature quality*) as well as the interrelation between both aspects (VAN EXEL ET AL. 2010). These elements may provide a good starting point for investigations of data quality from LBSN.

The most pressing questions are how good positional and attribute accuracy as well as data set completeness are compared to traditional governmental

and commercial data sets or data from crowdsourcing activities such as OpenStreetMap. Just as previous studies on *VGI* data quality revealed (NEIS ET AL. 2012; HAKLAY 2010; GIRRES & TOUYA 2010), highly populated areas may unveil a better overall data quality, which is related to a superior user activity in these areas. Therefore, investigations on the correlation of user activity (e.g. the number of check-ins or visiting people in certain areas) and data quality may provide insights, which can be employed for quality estimations in unknown areas. Further, studies on the temporal quality of the data set may provide insights on how a growing (or probably shrinking) user community in LBSN influences its data quality over time. LBSN such as Foursquare run on top of large data bases of highly dynamic data such as *Point of Interests* with businesses opening, closing and moving to new locations on a daily basis. Answers to the question of how user's updating activities keep up with these dynamics can further provide a starting point for extrapolations on how the data sets will evolve in the future. To investigate these objectives methods already applied in studies of geo-data quality in general and especially *VGI* have to be evaluated for suitability and improved if necessary.

ELWOOD ET AL. (2013) further propose approaches for quality assurance as an emerging research objective. The authors suggest a review system with trustworthy users acting as moderators who review the contributions. A second strategy is to employ a syntactic rule system that allows for cross-checking user-contributed observations. However, research is required on how to formulate and organize such a rule system and what kinds of analysis are required to apply such a system.

5.3 Data Integration

When combined with other data sets from official or volunteered sources and other Social Network Sites, data from Location Based Social Networks will attain its full potential when the single data sets complement each other. A crucial objective is therefore to develop, implement and evaluate new methods for fusing user generated data from LBSN with other sources.

The problem of data matching and conflation has been studied extensively throughout the recent years. A significant number of methods for matching and conflating geo-spatial data, which employ geographic proximity and semantic similarity in different ways, has been proposed before (RUIZ ET AL. 2011).

However, most methods discussed in the literature have been developed for and tested against artificial or authoritative data with predefined and predictable data quality standards. As discussed in the previous section, quality standards and control are not given by definition in LBSN, therefore we

cannot rely on positional accuracy for geometric matching or attribute completeness and accuracy to derive semantic relatedness alone. The problem is amplified when considering the high density of data especially in urban areas. Figure 5 demonstrates that this is even the case for small cities. A matching algorithm primarily relying on geometric proximity is not feasible, because corresponding features are not easily identifiable. Another problem is that attribute information is mostly given by *tags* attached by users. The resulting corpus of keywords follows no formal taxonomy or wordlist and is thus also known as folksonomy (GUY & TONKIN 2006). Hence, in order to fuse geographic data from social networks matching methods are required that employ geometric proximity and semantic distances (JANOWICZ ET AL. 2011) in a combined manner.

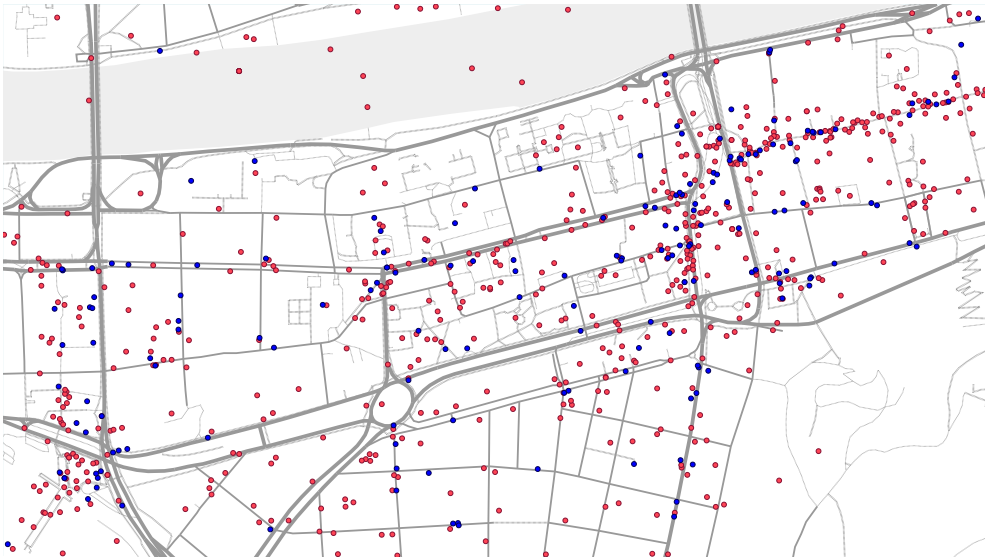


Figure 5: Distribution of Points of Interests from Facebook (blue) and Foursquare (red) Heidelberg's city centre.

When conflating the different data sets these uncertainties have to be taken into account as well. SUI & GOODCHILD (2011) already stated the problem of conflating geometries with different unknown quality characteristics and generalizations. Considering user activity as a proxy for data quality, the weight of a feature for the final conflation result can be adjusted accordingly. This can further be taken into account when conflating attribute values. The values of the data set with the best expected data quality can be taken as foundation and complemented with information from the corresponding sources.

5.4 Space vs. Place

The concept of space, which emphasizes a specific position on earth has long been the main way to link information to a location on earth. However, humans tend to refer to specific locations by name, without knowing its exact position or extent (ELWOOD ET AL. 2013). Contributions of HOLLENSTEIN & PURVES (2010) and CRANSHAW ET AL. (2012) have demonstrated the potential of user-generated data for the derivation of place representations. Gazateers have been used to link place names to locations on earth. Open research questions are thus how these fuzzy information can be represented in traditional space-oriented GIS or how they have to be extended. It further needs to be investigated what data from online social networks can be used to populate these gazateers. ELWOOD ET AL. (2013) raise additional questions towards a platial GIS, its required functionality, data models, cognitive models to successfully operate such a system.

5.5 Representativeness of data

Another critical issue of user-generated information is the representativeness of data. It can be questioned if data shared in online social networks can be considered representative of the whole society since the user group can be expected to be narrowed towards *digital natives*; e.g., young people who grew up having constant internet access and use the Web constantly throughout their life (BOYD & CRAWFORD 2012). This also includes the problem of digital divide: Most activity in online social networks is concentrated to Northern America and Europe, whereas in other parts of the world access to digital devices and the internet is reserved to a few people. These facts have to be considered when dealing with data from social networks. We further need research into these aspects, especially on the spatial and social constraints of user activity and how the growing penetration of cell phones in developing countries influences these developments.

5.6 Data Ownership & Privacy

Closely related to issues of privacy are questions of how to store data and how the data is exploited. Users have to trade off between the amount and depth of personal information they share online and an increase of information relevance. Sharing of data is essential for online social networks to work out, so question of data storage and ownership are essential. Currently three approaches to these questions exist:

Big players – like Facebook – require the user to sign their Terms of Service

that often include passages that allow the company to use personal information to make money; for instance, by revealing that information to advertisers. This leads to serious threats of privacy as users lose control over their data. The new online social network App.net¹⁰ addresses these privacy issues by requiring the user to pay a fee. In return, users are guaranteed that no personal information is shared with third parties. A third alternative is Diaspora¹¹. The underlying software is completely open-source; every user may run a personal instance and therefore keep control of the data. However, running one's own service is relatively difficult and only technically qualified users may be able to run such a service.

All approaches have clear drawbacks, be it advertisements based on personal data, monetary aspects or required technical abilities. Thus, novel concepts have to be developed that allow to keep the user's personal data secure, while still allowing companies to run their services.

6 Conclusion

This review paper provides an overview of the nature and current research on *Location Based Social Networks*. Drawing from different attempts to define and structure the nature of LBSN a synoptical definition on the topic as well as the data produced by the users has been given.

It has been demonstrated that despite the young age of LBSN there is already a considerable number of research efforts on this topic. Scholars working with data from LBSN are predominantly related to social sciences so most activities are focussed on investigating social networks, how geography influences the formation and evolution of these networks and the spatial dimension of human behavior. Privacy is – due to the nature of personal location information – another issue that has been studied extensively. Potential threats of privacy have been identified and methods preserving privacy have been proposed.

However, several open questions remain unanswered. First, so far only a few LBSN have been studied at all. Most investigations focus on Twitter, Flickr and to a lesser extent on Facebook. The multitude of other LBSN remains left behind yet. It has been further exemplified, that the exploitation of geographic information and the research on information retrieval from these LBSN has just started. Only a few studies investigated how data from social networks can be applied to use case in natural hazard response and monitoring diseases.

¹⁰<http://app.net/>

¹¹<http://diasporaproject.org/>

Open research objectives remain in the fields on clarifying and categorizing the multitude of available LBSN as well as investigating on the provided data and information. We conclude that crucial part in research should be the assessment of data quality, both the investigation of specific elements of data quality and the assessment of quality from other proxy characteristics. The fusion of complementary data from different source may further provide a solution for data inaccuracies. Here, the development of matching methods combining spatial and semantic proximity are required. Ethical questions on privacy and commercial data exploitation need to be addressed as well. Finally, new GIScience concepts are required that take account for place centric way of thinking.

References

- 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, WWW '10*, 61–70. ACM, Raleigh, North Carolina, USA.
- BAWA-CAVIA, A. (2011): Sensing The Urban: Using location-based social network data in urban analysis. In: *First Workshop on Pervasive Urban Applications (PURBA)*. URL <http://purba.mit.edu/sites/default/files/Pervasive2011.Workshop03.Paper06.pdf>.
- BOUILLOT, F.; PONCELET, P. & ROCHE, M. (2012): How and why exploit tweet's location information? In: J. GENSEL; D. JOSSELIN & D. VANDENBROUCKE (Hrsg.), *Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27*. Avignon, France.
- BOYD, D. & CRAWFORD, K. (2012): Critical Questions for Big Data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5): 662–679.
- BOYD, D. & ELLISON, N. (2008): Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13: 210–230.
- CAMILLI, M. (2012): Preserving Co-Location Privacy in Geo-Social Networks. *CoRR*, abs/1203.3946.
- 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, CIKM '10*, 759–768. Toronto, ON, Canada.

- CHENG, Z.; CAVERLEE, J.; LEE, K. & SUI, D. (2011): Exploring Millions of Footprints in Location Sharing Services. In: *Fifth International AAAI Conference on Weblogs and Social Media*. Barcelona, Spain. URL <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2783>.
- COLLIER, N.; SON, N. & N.M., N. (2011): OMG U got flu? Analysis of shared health messages for bio-surveillance. *Journal of Biomedical Semantics*, 2(5): 9–19.
- Craglia, M.; de Bie, K.; Jackson, D.; Pesaresi, M.; Remete-Fülöpp, G.; Wang, C.; Annoni, A.; Bian, L.; Campbell, F.; Ehlers, M.; van Genderen, J.; Goodchild, M.; Guo, H.; Lewis, A.; Simpson, R.; Skidmore, A. & Woodgate, P. (2012): Digital Earth 2020: Towards the vision for the next decade. *International Journal of Digital Earth*, 5(1): 4–21.
- CRANDALL, D.J.; BACKSTROM, L.; COSLEY, D.; SURI, S.; HUTTENLOCHER, D. & KLEINBERG, J. (2010): Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences*. URL <http://www.pnas.org/content/early/2010/12/02/1006155107.abstract>.
- CRANSHAW, J.; SCHWARTZ, R.; HONG, J. & SADEH, N. (2012): The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In: *International AAAI Conference on Weblogs and Social Media*. URL <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4682>.
- CRANSHAW, J.; TOCH, E.; HONG, J.; KITTUR, A. & SADEH, N. (2010): Brindging the Gap Between Physical Location and Online Social Networks. In: *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubi-comp '10, 119–128. Copenhagen, Denmark.
- CULOTTA, A. (2010): Towards detecting influenza epidemics by analyzing Twitter messages. In: *Proceedings of the First Workshop on Social Media Analytics*, SOMA '10, 115–122. ACM, Washington D.C., USA.
- DAVIS JR., C.A.; PAPP, G.L.; DE OLIVEIRA, D.R.R. & DE L. ARCANJO, F. (2011): Inferring the Location of Twitter Messages Based on User Relationships. *Transactions in GIS*, 15(6): 735–751.
- DE SOUZA E SILVA, A. & FRITH, J. (2010): Locative Mobile Social Networks: Mapping Communication and Location in Urban Spaces. *Mobilities*, 5(4): 485–505.

- EARLE, P.; BOWDEN, D. & GUY, M. (2011): Twitter earthquake detection: earthquake monitoring in a social world. *Annals of Geophysics*, 24(6): 708–715.
- ELWOOD, S.; GOODCHILD, M. & SUI, D. (2011): Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers*. URL <http://rsa.informaworld.com/10.1080/00045608.2011.595657>.
- ELWOOD, S.; GOODCHILD, M. & SUI, D. (2013): Prospects for VGI Research and the Emerging Fourth Paradigm. In: D. SUI; S. ELWOOD & M. GOODCHILD (Hrsg.), *Crowdsourcing Geographic Knowledge. Volunteered Geographic Information (VGI) in Theory and Practice*. Springer Berlin / Heidelberg.
- FISCHER, E. (2011): European detail map of Flickr and Twitter locations. URL <http://www.flickr.com/photos/walkingsf/5912946760/>.
- FRENI, D.; RUIZ VICENTE, C.; MASCETTI, S.; BETTINI, C. & JENSEN, C.S. (2010): Preserving location and absence privacy in geo-social networks. In: *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM '10, 309–318. Toronto, ON, Canada.
- GALLAGHER, A.; JOSHI, D.; YU, J. & LUO, J. (2009): Geo-location inference from image content and user tags. *Computer Vision and Pattern Recognition Workshop*, 55–62.
- GELERNTER, J. & MUSHEGIAN, N. (2011): Geo-parsing Messages from Micro-text. *Transactions in GIS*, 15(6): 753–773.
- GIRARDIN, F.; DAL FIORE, F.; BLAT, J. & RATTI, C. (2007): Understanding of Tourist Dynamics from Explicitly Disclosed Location Information. In: *4th International Symposium on LBS and TeleCartography*. Hong Kong.
- GIRRES, J.F. & TOUYA, G. (2010): Quality Assessment of the French Open-StreetMap Dataset. *Transactions in GIS*, 14(4): 435–459.
- GOMIDE, J.; VELOSO, A.; MEIRA, W.; ALMEIDA, V.; BENEVENUTO, F.; FERRAZ, F. & TEIXEIRA, M. (2011): Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. In: *Proceedings of the ACM WebSci'11*, 1–8. Koblenz, Germany.
- GONZALEZ, R.; RUMÍN, R.C.; CUEVAS, Á. & GUERRERO, C. (2011): Where are my followers? Understanding the Locality Effect in Twitter. In: *Proc. of CoRR Workshop on Service Oriented Computing*. Paphos, Greece.

- GORDON, E. & DE SOUZA E SILVA, A. (2011): *NetLocality. Why Location Matters in a Networked World*. Wiley-Blackwell, West Sussex, UK.
- GORE, A. (1999): The Digital Earth: Understanding our planet in the 21st century. *Photogrammetric Engineering and Remote Sensing*, 65(5): 528–530.
- GUY, M. & TONKIN, E. (2006): Folksonomies. Tidying up Tags? *D-Lib Magazine*, 12(1).
- HAKLAY, M. (2010): How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, 37: 682–703.
- HARVEY, D. (1989): *The Condition of Postmodernity*. Blackwell Publishers, Oxford.
- HOCHMAIR, H. & ZIELSTRA, D. (2012): Positional Accuracy of Flickr and Panoramio Images in Europe. In: A. CAR; G. GRIESEBNER & J. STROBL (Hrsg.), *Geospatial Crossroads @ GI Forum '12. Proceedings of the Geoinformatics Forum*, 14–23. Wichman, Heidelberg.
- HOLLENSTEIN, L. & PURVES, R. (2010): Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, (1): 21–48.
- JANOWICZ, K.; RAUBAL, M. & KUHN, W. (2011): The semantics of similarity in geographic information retrieval. *Journal of Spatial Information Science*, (2): 29–57.
- JOSHI, D.; GALLAGHER, A.; YU, J. & LUO, J. (2010): Exploring user image tags for geo-location inference. In: *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 5598–5601.
- KULSHRESTHA, J.; KOOTI, F.; NIKRAVESH, A. & GUMMADI, K. (2012): Geographic Dissection of the Twitter Network. In: *International AAAI Conference on Weblogs and Social Media*. URL <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4685/4984>.
- LAMPOS, V.; DE BIE, T. & CRISTIANINI, N. (2010): Flu Detector - Tracking Epidemics on Twitter. In: J. BALCÁZAR; F. BONCHI; A. GIONIS & M. SEBAG (Hrsg.), *Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science*, vol. 6323, 599–602. Springer Berlin / Heidelberg.

- LI, J. & RAO, H. (2010): Twitter as a Rapid Response News Service: An Exploration in the Context of the 2008 China Earthquake. *The Electronic Journal on Information Systems in Developing Countries*, 42(4): 1–22.
- LINDQUAIST, J.; CRANSHAW, J.; WIESE, J.; HONG, J. & ZIMMERMANN, J. (2011): I'm the Mayor of My House: Examining Why People Use Foursquare - a Social-Driven Location Sharing Application. In: *Proc. of the 2011 annual conf. on Human factors in computing systems*, CHI '11, 2409–2418. ACM, Vancouver, BC, Canada.
- MASCETTI, S.; FRENI, D.; BETTINI, C.; WANG, X.S. & JAJODIA, S. (2011): Privacy in geo-social networks: proximity notification with untrusted service providers and curious buddies. *The VLDB Journal*, 20(4): 541–566.
- MOUNTAIN, D. & RAPER, J. (2001): Positioning techniques for location-based services (LBS): characteristics and limitations of proposed solutions. *Aslib Proceedings*, 53(10): 404–412.
- NEIS, P.; ZIELSTRA, D. & ZIPF, A. (2012): The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011. *Future Internet*, 4(1): 1–21.
- NOULAS, A.; SCELLATO, S.; MASCOLO, C. & PONTIL, M. (2011): An Empirical Study of Geographic User Activity Patterns in Foursquare. In: *Fifth International AAAI Conference on Weblogs and Social Media*. Barcelona, Spain. URL <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2831>.
- O'REILLY, T. (2005): What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software. URL <http://oreilly.com/web2/archive/what-is-web-20.html>.
- PAGE, X.; KOBZA, A. & KNIJNENBURG, B. (2012): Don't Disturb My Circles! Boundary Preservation Is at the Center of Location-Sharing Concerns. In: *International AAAI Conference on Weblogs and Social Media*. URL <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4679>.
- PAUL, M. & DREDZE, M. (2011): You Are What You Tweet: Analyzing Twitter for Public Health. In: *Fifth International AAAI Conference on Weblogs and Social Media*. Barcelona, Spain. URL <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2880>.
- RITTERMAN, J.; OSBORNE, M. & KLEIN, E. (2009): Using prediction Markets and Twitter to predict a Swine Flu Pandemic. In: *Proceedings of the 1st International Workshop on Mining Social Media*.

- RUIZ, J.J.; ARIZA, F.J.; UREÑA, M.A. & BLÁZQUEZ, E.B. (2011): Digital map conflation: a review of the process and a proposal for classification. *Int. Journal of Geographical Information Science*, 25(9): 1439–1466.
- RUIZ VICENTE, C.; FRENI, D.; BETTINI, C. & JENSEN, C. (2011): Location-Related Privacy in Geo-Social Networks. *IEEE Internet Computing*, 15(3): 20–27.
- SADEH, N.; HONG, J.; CRANOR, L.; FETTE, I.; KELLEY, P.; PRABAKER, M. & RAO, J. (2009): Understanding and capturing people’s privacy policies in a mobile social networking application. *Personal Ubiquitous Comput.*, 13(6): 401–412.
- 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*, WWW ’10, 851–860. ACM, New York, NY, USA. URL <http://doi.acm.org/10.1145/1772690.1772777>.
- SCCELLATO, S.; MASCOLO, C.; MUSOLESI, M. & LATORA, V. (2010): Distance Matters: Geo-social Metrics for Online Social Networks. In: *Proceedings of the 3rd conference on Online social networks*, WOSN’10. Boston, Massachusetts.
- SCCELLATO, S.; NOULAS, A.; LAMBIOTTE, R. & MASCOLO, C. (2011a): Socio-spatial Properties of Online Location-based Social Networks. In: *Fifth International AAAI Conference on Weblogs and Social Media*. Barcelona, Spain. URL <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2751>.
- SCCELLATO, S.; NOULAS, A. & MASCOLO, C. (2011b): Exploiting Place Features in Link Prediction on Location-based Social Networks. In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD ’11, 1046–1054. San Diego, California, USA.
- SCHLIEDER, C. & MATYAS, C. (2009): Photographing a City: An Analysis of Place Concepts Based on Spatial Choices. *Spatial Cognition & Computation: An Interdisciplinary Journal*, 9(3): 212–228.
- 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, 484–491. ACM, Boston, Massachusetts.
- STARBIRD, K. & STAMBERGER, J. (2010): Tweak the Tweet: Leveraging Microblogging Proliferation with a Prescriptive Syntax to Support Citizen Reporting. In: *Proceedings of the 7th International ISCRAM Conference*. Seattle, USA.

- SUI, D. & GOODCHILD, M. (2011): The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science*, 25(11): 1737–1748.
- SUN, Y.; HELBICH, M.; FAN, H. & ZIPE, A. (2012): Analyzing Human Activities through Volunteered Geographic Information: Using Flickr to Analyze Spatial and Temporal Pattern of Tourist Accommodation. In: *9th Symposium on Location Based Services*. Munich. Germany.
- TABSCOTT, D. (2001): *The Digital Economy: Promise and Peril in the Age of Networked Intelligence*. McGraw Hill, New York, USA.
- TAKHTEYEV, Y.; GRUZD, A. & WELLMAN, B. (2012): Geography of Twitter networks. *Social Networks*, 34(1): 73–81.
- TANG, K.; LIN, J.; HONG, J.; SIEWIOREK, D. & SADEH, N. (2010): Rethinking location sharing: Exploring the implications of social-driven vs. purpose-driven location sharing. In: *Proceedings of the 12th ACM international conference on Ubiquitous computing*, UbiComp '10, 85–94. ACM, Copenhagen, Denmark.
- THIELMANN, T. (2010): Locative Media and Mediated Localities. *Aether: The Journal of Media Geography*, 5(1): 1–17.
- TURNER, A. (2006): *Introduction to Neogeography*. O'Reilly Media.
- VAN EXEL, M.; DIAS, E. & FRUIJTIER, S. (2010): The impact of crowdsourcing on spatial data quality indicators. In: *GIScience 2010*. Zurich, Switzerland. URL http://www.giscience2010.org/pdfs/paper_213.pdf.
- VAN LAERE, O.; SCHOCKAERT, S. & DHOEDT, B. (2010): Towards automated georeferencing of Flickr photos. In: *Proceedings of the 6th Workshop on Geographic Information Retrieval*, 1–5. Zurich, Switzerland.
- VAN OORT, P. (2006): *Spatial data quality: from description to application*. Dissertation, Wageningen Universiteit.
- VIEWEG, S.; HUGHES, A.L.; STARBIRD, K. & PALEN, L. (2010): Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In: *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, 1079–1088. Atlanta, Georgia, USA.
- WERLEN, B. (1995): *Sozialgeographie alltäglicher Regionalisierungen 1: Zur Ontologie von Gesellschaft und Raum*. Franz Steiner Verlag, Stuttgart.