Chapter 26: Geospatial Analysis

Olga Buchel, Diane Rasmussen Pennington

Abstract

This chapter is about geospatial analysis of social media. It summarizes major issues with retrieving, sampling, geocoding, and analyzing social media data. The chapter discusses geospatial analysis from the perspectives of different domains of knowledge, including information science, geographic information science, geovisualization, information visualization and visual analytics by presenting numerous illustrative examples and case studies. It also shows benefits and shortcomings of these methods and defines existing gaps in geospatial analysis.

Dr. Olga Buchel is an interdisciplinary researcher specializing in geovisualization, information visualization, visual analytics, and geographic information retrieval. She holds a PhD in Library and Information Science from Western University. In the past she has worked at the Alexandria Digital Library in Santa Barbara, CA, the forerunner of Google Maps. Olga's most recent publications are about geospatial analysis of user interactions with Flickr image collections, and the role of interactive exploratory visualizations in research and knowledge discovery in public health and international business. She has presented her research at a number of international conferences, including GEOMED, iConference, Annual Meetings of the Association for Information Science and Technology, visualization workshop at the American Academy of Management, workshop on Provenance of Sensemaking at IEEE VIS, and other conferences. She currently works with Big Data at SiTechnologyGroup, Inc., where she develops health applications and visualizations from clinical health records.

Dr. Diane Rasmussen Pennington is a Lecturer in Information Science in the Department of Computer and Information Sciences at the University of Strathclyde in Glasgow, Scotland, where she is a member of the iLab and the Digital Health and Wellness research groups. She is also the Social Media Manager of the Association for Information Science & Technology (ASIS&T). Dr. Rasmussen Pennington has taught classes on research methods, social media, knowledge organisation, and a range of information technology topics. Her diverse research areas encompass non-text information indexing and retrieval, Emotional Information Retrieval (EmIR), user behaviours on social media, and online health information preferences. She is the editor of *Indexing and Retrieval of Non-Text Information*(2012) and *Social Media for Academics: A Practical Guide* (2012). She is currently editing a book series entitled *Computing for Information Professionals*.

Keywords: geospatial analysis, spatio-temporal analysis, spatial networks, geo-social analytics, visual analytics

Introduction

Blogs, tweets, comments, images, videos, RSS feeds, online games, accounts in social media and clouds are inundated with references to geographic locations due to the proliferation of location-aware devices. On the one hand, social media systems track locations implicitly where people go, where they search or share information from. On the other hand, social media users voluntarily share the location of their travel routes, destinations, hotels, and restaurants as well as images enhanced with geospatial coordinates. As Goodchild (2007) noted, citizens have become sensors who actively collect and contribute geospatial information. This phenomenon gave rise to the convergence of geographic information science and social media (Sui and Goodchild, 2011). These volunteered crowd-sourced data reduce the burden of data collection (Stefanidis et al., 2013), and they open up exciting opportunities to study human movement from the perspective of their socio-spatial behaviour (De Longuevilleet al., 2009). This chapter summarizes a variety of techniques used in the investigation of information flows and social networks on human-defined landscapes.

This chapter takes a multidisciplinary approach as techniques are now being developed not only in geography, but many other research domains, particularly visual analytics, geovisualization, social sciences, and information science. In addition, geographic analyses are now integrated with other methods such as semantic analysis, machine learning, network analysis, econometrics, and human-computer interaction, and are being used for enhancing the understanding of spatio-temporal contexts of various phenomena. In other words, researchers use geographic locations shared in social media not only for understanding locations, but also to get better insights about phenomena under investigation (e.g., communities, economic and political impacts of events, disease outbreaks, communication patterns, emergency situations, and many others).

The section *Background Information* presents key properties of geographic locations which are crucial for understanding how geospatial analyses should be carried out. We also explain how geospatial analyses are complicated by semantic properties of information spaces and social networks and where references to geographic locations can be found in social media systems. In the *Analyzing Spatial Locations* section, we first give examples of research questions social media can be used for, a summary of how researchers prepare data for spatial analysis and how they assign coordinates to locations, and we highlight difficulties with extraction and disambiguation of place names. We then proceed to the discussion of pros and cons of the techniques used in geography, visual analytics and other research areas in order to extract insights about phenomena under investigation. We focus on *Exploratory Analysis*, I, *Standard Deviational Ellipse*s, and *Spatio-temporal Analysis*. The next section on *Geo-social Visual Analytics* addresses current limitations of geospatial analysis and describes new techniques that attempt to bridge network and map representations. The *Spatial Data Mining* section presents techniques for automated pattern extraction. We conclude our chapter with a debate about potential dangers of geospatial analysis associated with the breach of users' privacy, and give recommendations on how to protect privacy of social media users in geospatial analysis.

Background Information

Spatial information in social media is recorded in two forms: geospatial footprints and text (i.e., references to place names such as "Toronto" or "Paris"). A footprint is a representation of the spatial location or extent of a geographic object expressed in terms of geospatial coordinates (Hill, 2006). It can take many different forms: a dot, a line, a polygon, a set of dots, a boundary box, an image, or pixels. Footprints are required for creating a visualization on a map. Textual representations can take many forms too; they can be expressed in different languages or as codes, tags, ZIP Codes, mailing addresses, postal codes, time zones, IP addresses, or other notations. On the one hand, assigning a footprint no longer constitutes a difficulty due to an abundance of geocoding services. On the other hand, geocoding services are not always able to recognize location names in texts, or to match them with proper footprints due to place name changes, variations in transliterations, homonyms, variant spellings, or other semantic variations.

Back in the 1990s, when research on online geospatial systems was just starting, researchers in geodigital libraries, who were at the forefront of modeling geospatial descriptions, talked about feature types as important attributes of geospatial descriptions. Feature types are natural and cultural categories of geospatial locations (Hill, 2006). Natural features include continents, mountains, lakes, seas, forests, grasslands, and so on. Cultural features include types of businesses, man-made constructions, and places. These types help bridge place names with coordinates. They improve accuracy and precision for information retrieval, and they help people interpret the context in which communication takes place. For example, consider a medical instructor who often visits military conflict zones. Comments and messages from her on social media might differ in tonality and content according to the type of the location. In her native city she might look relaxed and happy, and involved in volunteering. In zones of military conflict her behaviour will change; she will share fewer comments and post pictures in a military uniform after she returns from the zone. Actions and comments of her close friends would also differ depending on her location. If she is getting ready to go to the zone of military conflict, friends would try to help her right away because they understand that she might need help urgently before she leaves for the zone. Without understanding location types, her comments are difficult to interpret. Designers at Facebook realize the drawback of not having location types in their social media platform; for this reason, they now offer a location check-in service. This service mostly provides places of interest as well as businesses that allow data scientists at Facebook to conduct research on checking in to certain location types (e.g., Chang and Sun, 2011). Using check-in data Chang and Sun (2011) built a model that helped them predict where people will check into next. Such predictions can improve ranking of places of interest and create better advertisement targeting. How to capture or extract other types of locations is a question that has yet to be addressed by researchers, social media developers, and analysts.

Besides place names and footprints, geographic references can also be classified in terms of accuracy, precision, scale, and uncertainty. Accuracy is “the degree to which the recorded value represents the ‘correct’ value” (Hill, 2006, p. 227). Accuracy of reference points collected with cell phones varies from 500 meters to 20 meters (Ramdani, 2011). Accuracy of IP geocoding services are only good enough to locate a particular city, or a country depending on the location in the world (Ramdani, 2011). In some countries accuracy is higher than in others. For example, the accuracy of detecting IP addresses in the Philippines or Croatia is lower than 60%, while in the USA and Canada, it is 84% (MaxMind, 2015). Accuracy of geocoding services will also differ due to different methods used for calculating coordinates (Whitsel et al., 2006). Precision refers to the potential amount of geographic extent represented by the locality. Coordinates are more precise and accurate than textual references because when textual references in the form of place names (not street addresses) are translated to coordinates, they are commonly represented either as a pair of coordinates that correspond to the central point of the location, or as a bounding box (i.e., a rectangle drawn around the place).

Scale is a primary property of maps. It is the ratio between the linear distance on the map and the corresponding linear distance on the Earth's surface (Longley et al., 2005). However, researchers in geographic information science as well as in library and information science have argued that geographic references form a semantic space of their own (Buchel, 2013; Fabrikant, 2001a; Fabrikant, 2001b; Fabrikant and Skupin, 2005; Fabrikant and Buttenfield, 2001). As such, the space also has scale. The semantic space is defined by implicit relationships among geographic references, specifically relationships among countries and provinces as well as provinces and smaller geographic locations. Together they form a semantic hierarchy, different levels of which correspond to different geospatial scales. Accuracy, precision, and scale at which geographic locations are reported in social media affect spatial uncertainty of geospatial representations. Uncertainty is defined as the difference between a real geographic phenomenon and the user's understanding of the geographic phenomenon (Longley et al., 2005). Things become even more complicated when we add semantic uncertainty to this mix (Bordogna et al., 2012). Semantic uncertainty implies that social media users may give different meanings to the same term, phrase and/or actions, which may lead to false conclusions. For instance, a phrase such as "close to the northeast of Milan" can be interpreted differently as people's perceptions of distance vary.

Research in geographic information retrieval suggests that feature types may reduce geospatial uncertainty and increase geospatial accuracy and precision (Bo and Baldwin, 2012).

Last but not least, it is important to take into account that references to geographic locations in social media are not the primary information objects. In other words, social media systems are not about geographic concepts or geography, but they are rather about people, information objects (images or videos), relationships among people and objects, communities of people that engage in communication, document sharing, information flows, and other activities. This suggests that geospatial analysis in such complex systems should take into account all these processes and relationships. Some researchers have already identified geospatial properties of social networks. Scellato et al. (2011) found that networks in Foursquare have different characteristic spatial lengths of interaction across both their social ties and social triads. Doytsher et al. (2010) found that human movements generate life patterns that include social connections and places a person visited.

Where can geospatial locations be found, and how often are they reported? Geographic references can be found in different contexts of social media systems. They are used to denote users' home locations, events, and spatial coverage of microblog entries or images. Geographic references may refer to static states, dynamic states, or events. For example, references to home locations can be considered more or less static whereas references to tweets are regarded ambient, because they represent momentary social hotspots (Stefanidis et al., 2013).

Studies report different numbers of users who provided geospatial locations in their data samples. Java et al.(2007) reported that 52% of Twitter users (39,000 out of 76,000) included in their study had geospatial locations. Hecht et al. (2011) reported that two out of three users in their study had information about their geospatial locations. Cheng et al. (2010) reported that 5% of users in their study listed locations in the form of coordinates, and 21% reported locations at the city level. Several studies have shown that locational information can be inferred from the content that users post in social media (Cheng et al., 2010; Popescu and Grefenstette, 2010; Backstrom et al., 2010). MacEachren et al. (2011) studied people who were trying to use Twitter for crisis management, and reported that the proportion of users with geolocation turned on is probably still in the single digits. Stefanidis et al. (2013) suggested that such variations in location reporting can be attributed to an uneven distribution of the latest mobile devices.

Analyzing spatial locations

In this chapter, we emphasize that spatial analysis is not only about linking microblogs to a map, but it is also about understanding how they relate, what they mean, and what should be done about them. It is about measuring geospatial footprints of online and offline communities, determining their volume, finding their proximity to other communities, overlaps and intersections, identifying spatial clusters, locating hot spots of activities, and making predictions about possible outcomes of events (ESRI, 2015).The Geospatial Analysis pipeline reminds us of the standard data science pipeline, shown in Figure 26.1.

[Figure 26.1 about here] Geospatial Analysis pipeline. Source: Authors.

Geospatial analysis of social media requires a wide array of interdisciplinary skills (not limited to knowledge of geographic information systems) including the ability to do semantic analysis, network analysis, retrieval analysis, and statistical analysis. It also requires good programming skills in order to be able to retrieve data from application programming interfaces (APIs), and to scrape data, do data cleaning and preprocessing, and perform data transformations.

Research questions

Researchers use social media and geospatial analysis for forecasting political opinions on the web (Sobkowicz et al., 2012), identifying and mapping global virtual communities (Stefanidis et al., 2013), making meteorological observations (Hyvärinen and Saltikoff, 2010), studying structure and dynamics of natural cities (Jiang and Miao, 2015), tracking infectious diseases (Padmanabhan et al., 2013), managing crisis situations (MacEachren et al., 2011a), capturing human movement patterns across political borders (Blanford et al., 2015), discovering significant events and patterns (Andrienko et al. 2010b), understanding protest movements (Gleason, 2013), finding geographic patterns of communication networks (Conover et al., 2013), and answering many other questions related to human movements and communication. The general trend is the following: researchers use maps 1) to report their findings, 2) to verify whether social media is more reliable than other techniques, 3) to discover new patterns and insights about phenomena, 4) to generate hypotheses about phenomena, and 5) to understand laws that can explain how networks work. For example, in a recent paper, Krings et al. (2009) investigated the network of mobile phone customers and analyzed the geographical patterns of the customers. After aggregating by city, these authors found that the inter-city communication intensity follows a gravity law. In other studies researchers compare patterns observed in empirical networks with patterns in spatial network models (Barthélemy, 2010).

Throughout its existence, the geographic information science community has developed a number of tools and techniques that can help with geospatial analysis. These techniques focus on identifying clusters in space and time, predictive modeling, exploratory analysis, and others. It is important to pay attention to what questions these techniques can answer and how. This may affect how the analysis will proceed. It may help expand the original set of questions social scientists are coming up with and may help them better understand social phenomena in terms of space and time. Using geographic information shared through social media is likely to produce a larger and more accurate dataset compared to geographic information collected in other ways, because it is created both automatically by users' mobile devices as well as by the users themselves. They are motivated to share it because they want their social networks to see it, whether it is for political organizing purposes or simply for staying in contact with friends and family.

Sampling

Most of the studies that we review in this chapter have used some sort of statistical sampling technique. Given the overwhelmingly large size of information spaces in each social media platform, researchers develop testbed collections by using queries. Queries can be geospatial and/or textual. For example, Purcell and de Beurs (2013) used only textual queries to collect queries about weather, but they had to eliminate many noisy tweets that their queries retrieved (e.g., "hot girls"). Stefanidis (2013) retrieved tweet hours within a 10-kilometre radius from

Tahrir Square during the Egyptian revolution. It is unclear whether all their tweets were relevant, as their study focused on analysis of networks, not analysis of semantics.

Noise, multiple variables, and uncertainty may potentially hinder exploration, hypothesis generation/exploration, and decision making. Researchers could improve their samples by using some of the query expansion techniques from information science and natural language processing. In information retrieval, the most common technique for query expansion is using a thesaurus, defined as a dictionary of terms related to the words in a query. Each word in a query can be automatically expanded with synonyms and related words from the thesaurus. This technique can be enhanced with term weighting, depending on how distant these terms are from the words in the query. Massoudi et al. (2011), for example, suggested using quality indicators to model retrieval of microblog posts. In their retrieval model, they assigned weights to emoticons, post length, shouting, capitalization, the existence of hyperlinks, reposts, followers, and recency of tweets. Other studies in information retrieval have also looked at more contextual searches, including event detection (Sayyadi et al., 2009), and mining consumer and political opinions (Sobkowicz et al., 2012).Using such techniques in the context of geographic analysis would improve the accuracy of geospatial analysis.

Also, it is important to keep in mind that even after all proper techniques of information retrieval are in place, it is possible to make mistakes in geospatial analysis. For example, in 2013, Google Flu Trends, which claimed to be an efficient public health tool for accurately monitoring the flu, far overstated its predictions. This overstatement can be seen in Figure 26.2. In all previous years it indeed provided very accurate results, but in 2013 their predictions went wrong. Butler (2013) reported on the huge discrepancy between Google Flu Trend's estimated peak flu levels and data collected by the U.S. Centers for Disease Control and Prevention (CDC) earlier in winter 2013. More precisely, Google doubled the numbers. The problem is that Google relies on searches related to flu symptoms, but in 2013 increased media attention to the flu season skewed Google's search engine traffic and consequently their geospatial predictions (Wagner, 2013).

[Figure 26.2 about here]–A visualization of three different methods for measuring flu prevalence in the US. Sources: Nature; Google Flu Trends (www.google.org/flutrends); CDC; Flu Near You

Geoparsing, geocoding and disambiguation of geographic place names

Extraction of geographic place names (also referred to as geoparsing) from short comments and tweets is different from extraction of place names from long and grammatically correct texts (Lingad et al., 2013). Whereas in grammatically correct texts geographical locations are usually capitalized, in short microblog posts such as tweets and comments, they are often placed in limited context and are not capitalized (e.g., British Columbia versus bc). Gelernter and Mushegian's (2011) analysis of Twitter messages from the February 2011 earthquake in Christchurch, Canterbury, New Zealand, showed that named entity recognition software recognizes places as proper nouns when locations are capitalized, but does not identify locations that are: not capitalized, local streets and buildings, non-standard place abbreviations, or misspellings. Karimzadeh et al. (2013) claimed that they solved this issue; they designed a tool calledGeoTxt API that extracts, disambiguates, and geocodes place names in short microblog posts. They did not mention, however, whether they used feature types.

Geotagging can be defined as “the process of adding geographical identification metadata to resources (websites, RSS feed, images or videos). The metadata usually consists of latitude and longitude coordinates, but they may also include altitude, camera heading direction and place names” (Torniai et al., 2007). Many newer cameras contain GPS receivers that add geographic coordinates to a photograph at the time it is taken. Also, social media users sometimes add semantic geotags to photographs such as “Paris” when they upload them.

A small but growing body of research involving geotagged photographs is developing. Rorissa et al. (2012) extracted geotags from Flickr images in order to examine their level of abstraction, and found no statistical significance with these levels, but suggested ways in which geotags can help people find photographs online more easily. O’Hare and Murdock (2013) used photographs’ tags and geotags to predict the locations of where photographs were taken with varying levels of accuracy. Sevillano et al. (2015) performed a similar experiment using only audio and visual information in a set of videos.

Kipp et al. (2014) extracted location information from the accounts of commenters, image descriptions (location field), and image titles. They investigated images and postcards posted on Flickr by the Library of Congress. All titles and descriptions of images in this collection have correct grammar and full texts. The extraction from image titles was complicated because all words in the titles were capitalized, and geographic names were often separated by other words. In the end, they assigned locations manually.

Besides adding coordinates to places extracted from unstructured texts or geotags, researchers often have to geocode from address-like strings (this process is also known as forward geocoding). Forward geocoding is a process of assigning coordinates for a full or partial address. A large arrays of services can be used for geocoding: Google Maps, Bing Maps, Twofishes, MapQuest, and other. A bit more challenging is the process of reverse geocoding, in which coordinates are mapped to addresses and toponyms. Uncertainty of the queried location complicates this process. For examples, it is unclear how to map GPS coordinates to one of many possible stores in a shopping mall (McKenzie and Janowicz, 2015).

Static versus interactive maps

Web 2.0 introduced us to not only social media platforms, but also interactive maps like Google Maps (https://www.google.com/maps/) andOpenStreetMap (https://www.openstreetmap.org). These maps have APIs which can be used for mashups, or custom maps that can be linked to databases on users' servers. These new technologies directly affect geospatial analysis. Many programmers who know how to link maps to data sprang into action and developed mashups for already existing databases (e.g., real estate databases, image repositories, public health databases and many others). However, many of these maps are not maps in a conventional sense, but rather retrieval tools; their purpose is to show users where information objects are located and allow them to click on the map to retrieve information about the objects. See the Flickr map in Figure 26.3 for an example of a map as a retrieval tool.

[Figure 26.3 about here] Example of a map as a retrieval tool. Source: http://www.flickr.com.

Researchers in academia welcomed these new developments too as this opened up an opportunity for crowdsourced mapping (Goodchild, 2007), ambient information mapping (Weidemann, 2013), real-time crisis maps (Middleton et al., 2014) and for moving geospatial analysis from proprietary mapping applications to the web. Some cartographers started referring to maps as visualizations (MacEachren, 1995). The emphasis of visualizations is "not on storing knowledge but on knowledge construction" (MacEachren et al., 1997, p. 336), which implies that maps have to be interactive. Web 2.0 gave rise to new techniques for representing information (e.g., linking social networks to maps and embedding other visualizations in maps), and for interacting with geospatial information (Buchel and Sedig, 2014) which was not possible to achieve with static maps. However, at the same time, analytical techniques that were developed in previous-generation geographic information systems (especially ESRI products; see http://www.esri.com/products) are hardly available for conducting geospatial analytics in web applications. For instance, it is still difficult to analyze hot spots or do grouping analysis without ESRI ArcGIS. ESRI is making attempts to bring their analytical techniques to the web; the company now offers an API that allows developers to embed them in their maps (see a list of available analysis tools from ESRI's ArcGIS API for JavaScript athttps://developers.arcgis.com/javascript/). In addition, geospatial analysis can now be carried out in cloud-based mapping applications such as Fusion Tables (https://sites.google.com/site/fusiontablestalks/stories), GIS Cloud (http://www.giscloud.com), CartoDB(https://cartodb.com), visualization frameworks such as D3.js (http://d3js.org), and R (https://www.r-project.org). QGIS (http://www.qgis.org) is a free, open source alternative to

ArcGIS. These tools offer a wide range of new and traditional analytical techniques. At first glance, cloud computing applications might look simple, but with additional programming, they can be used for fairly complex analysis.

Whereas static maps allowed carrying out analysis at the level of ‘flat’ snapshots, interactive maps opened many new options for analysis. It became possible to carry out dynamic analysis, such as tracking data related to human movement, and following the dynamics of social movements. Analysis in interactive visualizations is complicated by the motion of people and network dynamics in time and space. In interactive maps, great attention is paid to the conceptualization of time. Time has a complex structure; that is, it has a hierarchical system of units, including seconds, minutes, hours, days, weeks, months, years, decades, centuries and so on, which can be grouped into different calendar systems, cycles, etc. Time can be represented as a line, a cycle, a branch, or a hierarchy (Andrienko et al., 2010a).For example, Figure 26.4 shows health data over time represented as a linear plot graph on the left and as a cyclic spiral on the right. It can also be represented and parsed as sound in the form of data sonification (see Boase and Jamieson, this volume).

[Figure 26.4 about here] Health data over time as a linear plot graph and as a cyclic spiral. Source: Andrienko et al., 2010a

In the next section, we review both analytical and interactive techniques, how they can be used for analysis, what questions they can answer, and what the pros and cons of each technique are.

Our examples are drawn from several disciplines including geography, social sciences, visual analytics, health sciences, and others.

Data classification techniques

The starting point for any analysis usually involves proportional symbology or a choropleth map. Proportional symbology or choropleth maps rely on data classification. Researchers and analysts must decide which values should be associated with each bubble size on a symbol map or which values should be used for each class on a choropleth map. In other words, which units should be in the lowest class, which units should be in the highest class, and how should the rest of the units be distributed among the remaining classes? Most classification schemes in geographic information science provide a range of techniques for classifying univariate attributes; for example, mapping the number of tweets by county or state/province. Classification is an interesting topic on its own. There is a large number of classification schemes in geographic information science, including classifications based on unique values, manual classification, defined interval, exponential interval, equal count or quantile, percentile, natural breaks/Jenks, standard deviation, and box. Due to the size restrictions on this chapter, we are not able to explain the pros and cons of these methods; rather we refer readers to Mitchell (1999) and Longley et al. (2005).

In this chapter, we will explain the quantile method, as it is one of the most frequently used (Brewer and Pickle, 2002). Suppose we retrieved the following number of images per each location from Flickr. The frequency distribution of these counts is shown in Figure 26.5 below.

Frequencies range from 1 to 1,056. As you can see in the histogram, locations have very uneven counts of images linked to them, and the majority of locations have fewer than 20 images.

[Figure 26.5 about here]. Distribution of counts of images linked to different locations. Source: Authors.

The quantile classification method distributes equal numbers of observations into each class. The advantage to this method is that it often excels at emphasizing the relative position of the data values (i.e., which locations, or counties, or provinces/regions/states contain the top 20% of visualized objects). The major shortcoming of this classification is that locations placed within the same class can have wildly differing values, particularly if the data are not evenly distributed across its range (see the bars at the end of Figure 26.5). In addition, values with small range differences can be placed into different classes, suggesting a wider difference in the dataset than actually exists. For instance, in our case, locations with values of 1 can be placed in different classes.

Note the possible pitfalls of this method. With a four-category quantile classification, there is an equal number of cities in each class, and some locations with identical attribute values are placed in different classes. This suggests that this method may lead to a misleading visualization. This problem with classifications is not only true for quantile classification, but also for other classifications. They all may have problems with inclusions and exclusions. Discussion about how maps can lie can be found in Monmonier (2005;2014). A technique that can be useful for avoiding potential pitfalls with visualizations is to have controls for classification schemas in maps (see Figure 26.6). With such controls, the effects of classifications can be easily explored. Such controls can be added for interactive maps/mashups.

One critical issue to understand about data classification is that they enable map designers and analysts to layout all information on a flatland -- a two-dimensional surface (Tufte, 1991), which is the major drawback of this approach. Social media data is multivariate in nature, not limited to longitude and latitude variables. For this reason, data classifications can be regarded both useful and harmful for geospatial data analysis. They are useful because they create nice overviews and summaries, but they are harmful because they hide many details about the data. They display only one or two columns from a dataset, but we live in a world of very complex datasets. Each social media API has hundreds of descriptors that are commonly ignored during the analysis as they hardly fit into a classification schema in a visualization tool.

[Figure 26.6 about here]. Classification control. Source: OECD Regional Explorer, http://stats.oecd.org/oecdregionalstatistics/.

Exploratory analysis

Geographers suggest beginning geospatial analysis with exploratory analyses (Anselin,1999). The problem is that a collection of geographic data from a particular region may have many latent relationships which are difficult to display and communicate without explicit explanation about what is going on in the region. Exploratory analysis in professional GIS tools is often supported by additional statistical charts and graphs that show statistical distribution of data (e.g. scatterplots, scatterplot matrices, and box plots). These additional graphs and charts provide insight into the complex and subtle relationships that occur in geographic space (Gahegan, 1998).

To get started with exploratory analysis, we recommend novice researchers first study tasks that can be accomplished with such analysis (Andrienko and Andrienko, 2006). Andrienko and Andrienko (2006) give a comprehensive typology of the possible data analysis questions that "need to be answered by means of data analysis" (p. 8). They divide tasks into elementary and synoptic. Elementary tasks deal with individual elements of data and their properties; synoptic tasks deal with the datasets as the wholes and the patterns in the wholes. The main purpose of exploratory spatio-temporal analysis is to understand the overall behaviour captured by data. For example, in the context of social media analysis, research may be able to answer the following questions: How can the behaviour of social media users be characterized in terms of time and space? Is it changing over time? How does it change in terms of spatial dispersion? How does it change in terms of temporal aspects? Is there any periodicity in the behaviours?

Exploratory analysis of social media, however, should not be conducted solely in terms of space and time. Social media researchers should investigate not only the entire datasets, but also pay great attention to subsets (i.e., specific communities). This will allow the identification of differences in the behaviour of individual groups. Compare groups in terms of spatial relationships (proximity, intersections, overlaps, and so on). How do social structures differ in communities from different geospatial areas?

In the next two sections we give two examples of analytical techniques which we think might be useful for analysis of social media datasets. Both techniques have long history and have been well established in geospatial analysis.

Voronoi diagrams

A Voronoi diagram is an analytical technique to divide a map into geographic regions that are not equivalent to geographic regions. Its idea is simple: given a set of isolated points, points are associated with the closest member of the point set. The result is the partitioning of the space into a set of regions (Okabe et al., 2009). The Voronoi diagram implies that all possible points inside a polygon are closest to its centroid than to any other polygon (Manni et al.,2004). The external part of a Voronoi tessellation tends to infinity. Commonly, Voronoi diagrams are used to find the largest empty circle amid a set of points (e.g., to build a new pharmacy as far as possible from all the existing ones). Voronoi diagrams have been used in various knowledge domains, such as astronomy, business analytics, and soil analysis.

The earliest use of Voronoi diagrams in the context of geospatial analysis is attributed to John Snow (Brody et al., 2000). He used this technique in his second map of cholera investigation. At that time, he compiled the map and the diagram for illustrative rather than investigative purposes. The technique can be used for investigative purposes as well. For example, he could have drawn polygons around the pumps and then calculated the number of deaths in each polygon. That would have immediately shown that the pump that was at the center of the epidemics was an outlier, because the polygon that the pump was in had the largest number of deaths. Imagine using such techniques with health clinics providing flu shots and tweets about flu symptoms. If clinics posted the dates and times of their flu shot clinics in late autumn, along with their geographic locations, it might be easier for people to access the needed vaccinations. This technique can also be used for visualizing categorical data. For example, Manni et al. (2004) visualized genetic, morphologic, and linguistic patterns with Voronoi diagrams. Voronoi diagrams have also been used for enhancing spatial browsing and exploration of images on Flickr (Peca et al., 2011;Andrienko et al., 2010b).

Figure 26.7 below shows a Voronoi diagram of Craigslist which relays how the site redirects its users to local subdirectories based on their IP address (Nelson, 2011). The map approximates geographic coverage of Craigslist to Voronoi polygons. It is at least a start at visualizing the geographic coverage and distribution of the community-driven instances of Craigslist. Voronoi polygons might provide some useful context for other data, demographic or market information.

[Figure 26.7 about here] Map showing Craigslist market territories. Source: John Nelson, uxblog.idvsolutions.com

Standard deviational ellipse

This technique was first introduced by Lefever (1926). What is remarkable about this analytical approach is that it shows the orientation of the distribution. Human eyes can see the distribution on a map, but they cannot determine a trend in the data. The ellipse shows a spatial trend, so to speak; see Figure 26.8. Specifically, the ellipse shows the degree to which a distribution of features is concentrated or dispersed around its center (Wade and Sommer, 2006). The major axis shows the orientation of dispersion, the minor axis shows the minimum dispersion and the area of the ellipse is indicative of the spread (Gong, 2002). This type of analysis can be used for determining the orientation of opinions in social media, such as in survey analysis. For instance, Orchard et al. (2012) used this technique to analyze qualitative data about the life histories of sex workers. With the help of this technique, the researchers were able to determine that the areas where the women go for health and social services overlap the areas that are most dangerous for them (Orchard et al., 2012).

[Figure 26.8 about here]. Standard Deviational Ellipse. Source: Orchard t al. (2012).

Visual analytics, geovisualization, and information visualization

Visual analytics, geovisualization, and information visualization chart new directions in map visualization. Although they have distinct research agendas, they take similar approaches to geospatial analysis. Geovisualization research focuses on web-based, multi-view geospatial interfaces that support foraging and sensemaking (MacEachren et al., 2011b). Representative examples of geovisualizations are described in MacEachren et al. (2010) and Peca et al. (2011). Information visualization deals with visualization of abstract data that unlike spatial data usually have no intrinsic representation (Fekete andPlaisant, 2002, p. 1). The purpose of visual analytics is to "provide technology that combines the strengths of human and electronic data processing" (Keim et al., 2008, p. 162). Its key goal is to make data and information processing transparent for an analytic discourse. Geovisualization researchers in visual analytics aim to combine the strengths of human and electronic data processing in analysing spatio-temporal data and solving spatio-temporal problems (Andrienko et al., 2010). All these research directions put a great emphasis on interactions with visualizations that aim to enhance analytical tasks and expand an array of questions that can be asked about data. Interactions in this context are actions that provide users with the ability to directly or indirectly manipulate and interpret visual representations (e.g., rotate, select, and filter) (Yi et al., 2007). Interactions enable users to not only look at maps, but to change them according to their research questions. For example, if a map shows tweets, analysts should be able to search for specific tweets, topics in tweets, sentiments and so on to narrow down the representation of a map only to patterns that are relevant. The website http://www.onemilliontweetmap.com allows people to visualize in real time how many tweets are coming from geographic locations worldwide, and they can also filter them by keyword, hashtag, and so on. For example, Figure 26.9 shows a map of the last 10 million tweets posted starting from 15:50 BST on 15th December 2015.

[Figure 26.9 about here] Real time tweet map. Source: http://www.onemilliontweetmap.com.

Many of these systems are not yet available as commercial products; they are models and prototypes that demonstrate how support for data analysis, problem solving, decision-making and knowledge discovery can enhance geospatial analysis. For example, MacEachren et al. (2011a) described a prototype that enables information foraging and sensemaking using "tweet" indexing and display based on place, time, and concept characteristics. Schreck and Keim (2013) presented a model of an epidemic outbreak in a fictitious metropolitan area. Andrienko et al. (2010b) demonstrated a suite of methods for reconstructing past events from the activity traces that people leave in social media. Their method combines geocomputation, interactive geovisualization and statistical methods. They exemplify the utility of their methodology on a collection of Flickr photos.

A commercially available product that supports visual analysis is GeoTime (http://www.geotime.com). GeoTime is capable of detecting geo-temporal patterns and integrating narration in analytical processes. It improves understanding of entity movements, events, relationships, and interactions over time within a geospatial context. It uses narratives, hypertext-linked visualizations, visual annotations, and pattern detection to create an environment for analytic exploration and communication, thereby assisting analysts in identifying, extracting, arranging, and presenting stories within the data (Eccles et al., 2008). A snapshot from a GeoTime story can be seen in Figure 26.10. It shows the life patterns of two people; these patterns are described in terms of space and time. GeoTime has been successfully used in many real life decision making tasks, including crime detection, analysis of telecommunication patterns, military, government and business analysis tasks. In 2011, the London Police purchased GeoTime to study the behaviour of users on Facebook (Gayle,2011).

[Figure 26.10 about here]. Sample GeoTime story snapshot. Source: http://nickmalleson.co.uk/wp-content/uploads/2012/01/geotime_movements.jpg

GeoTime brought to life the famous space-time model, first envisioned by Hägerstrand in 1970 who planned to use this model as "a socio economic web model" (Hägerstrand, 1970, p. 10) to analyze people's interactions across space and time. It also materializes Tufte's dream of escaping data flatlands. Unlike other tools that visualize space and time, GeoTime visualizes the complexity of spatio-temporal relationships in a single 3-D view. GeoTime has multilingual support, and its latest version also has support for network analysis that is combined with geospatial analysis.

A tool that can largely complement analysis of social media in GeoTime by powerful semantic analysis is nSpace2. Like GeoTime, nSpace2 has won numerous IEEE Visual Analytics Science and Technology (VAST) Conference Contests as well. It supports natural language processing and innovative visual analytic techniques. It helps analysts efficiently and collaboratively produce insightful evidence-based reports. GeoTime and nSpace have been tested at VAST with an epidemiology analysis scenario using new reports. Tools enabled analysts to carry out trend analysis using the Country and Time dimensions as well as develop and validate hypotheses about why disease events and patterns occur (Proulx et al., 2006). This can be useful in the context of social media analysis too.

Geo-social visual analytics

A new trend is growing within the science of visual analytics known as geo-social visual analytics. In a research agenda for geo-social visual analytics, Luo and MacEachren (2015) explained that geo-social visual analytics differs from visual analytics by “explicit integration of social network perspectives and methods into the approach and tools” (p. 29), while still focusing on integration of interactive visual interfaces and computational analytical methods that can facilitate scientific reasoning. Another goal of geo-social visual analytics is to bridge the gap between methods used in geography, social sciences and network analysis. At the moment, spatial analysis and social network analysis solve social processes in their own contexts. While geographers look at movements, sociologists consider the relationships among people. Geographers also treat networks much more simply than physicists, who emphasize the power of networks, but ignore properties of geographic space (Curtin, 2007) and spatial constraints (Barthélemy, 2011). Geo-social visual analytics intends to pay more attention to overlaps between these two approaches and hopefully find better solutions for merging these two spaces. In the rest of this section, we give a few examples of how networks are analyzed in the framework of geo-social analytics.

Some geo-social analytics studies use integrated approaches that demonstrate how topology and geography interact with each other. For example, Luo et al. (2011) designed GeoSocial App, which reveals how groups identified in networks are positioned in geographic space over time. It enables capturing the dynamics of social relationships which is hard to understand from static graphs. Discovering interaction patterns between geographic space and network topology is useful for understanding individual- or group-level patterns that may have unique characteristics. Many studies suggest that certain processes in networks have structural and spatial constraints and that they have yet to be understood in empirical networks (Onnela et al., 2011; Barthélemy, 2011).

Such integrated approaches, however, do not summarize spatial or temporal as well as relational aspects of such networks. They allow researchers to examine how changes and processes in topologies affect geography and vice versa, but they do not provide an overview of what is going on in networks in space over time. Recently, an interesting solution for overcoming this limitation has been proposed by Koylu et al. (2014). Koylu et al.'s method takes into account distance, time (duration of interaction between individuals), and type of social relationship between each pair of individuals which are represented on a kernel density map (see Figure 26.10). Although Koylu et al. (2014) demonstrated the utility of this approach in the context of genealogical data, not social media data, their approach is well-suited for social media research. The results of Koylu et al.'s study reveal that family connectedness patterns in genealogical data are similar to migration and population growth patterns.

[Figure 26.10 about here]. Network topology represented on a map. Source: Koylu et al. (2014).

Spatial data mining

Spatial data mining is concerned with "the extraction of useful information and knowledge from massive and complex spatial databases" (Mennis and Guo, 2009, p.403). Spatial data mining techniques include spatial regression, spatial clustering, spatial autocorrelation, point pattern analysis, spatial classification and prediction, regionalization, and other. These techniques are useful for algorithmic detection of spatially-dependent patterns in data. They are not limited to traditional data types but also to newly emerged data types such as trajectories of individuals and groups of individuals with similar trajectories, data flows and other. Spatial patterns detected by spatial data mining techniques may yield important insights about individual and crowd behaviours, preferences, sentiments, human mobility networks (Gonzalez, Hidalgo, and Barabasi, 2008), and even properties of locations (Weiler et al., 2015). Data mining techniques could also help define rules for discovery of patterns, the purpose of which is to model human mobility patterns and consequently make predictions about all phenomena associated with human mobility (e.g., epidemics, information diffusion). For example, the co-location rule discovery process finds the subsets of features whose instances are frequently located together in geographic space (Shekhar and Huang, 2001).

A suite of tools for spatial data mining is available from the Spatial Data Mining and Visual Analytics Lab (http://www.spatialdatamining.org/software). Among these tools the most relevant tools for social media analysis are EntropMap and Flow Mapping. EntropMap helps detect the existence of multivariate relationships without assuming a prior relationship form. It can be used for analysis of opinions and estimating the results of elections. Flow Mapping can be used to explore communication flow patterns. In addition, grouping analysis in ArcGIS which is based on K-means can used for clustering networked datasets.

When mining spatial patterns, it is important to know that mobile networks prevailing in social media are limited in their predictability (Song et al., 2010; Gonzalez, Hidalgo, and Barabasi, 2008). Although on the one hand, human mobility patterns are “characterized by a deep-rooted regularity” and can be predicted and systematized, they lack variability: they have no significant gender- , age-, language-, or population-based differences (Song et al., 2010).

Privacy issues

Protecting the privacy of social media users’ geographic data is an ethical concern for researchers. Not all use of geographic data is unethical, but its use should be considered before research or development is undertaken. Martin (2015) suggested that “beneficial uses” of GPS data include “location-based coupons; traffic predictions; directions on map” while “questionable uses” would be “location-based stalking; iPhone as a homing beacon” (p. 68). Vicente et al. (2011) reviewed the basic features of social networks utilizing geographic information, such as friend tracking and “check-ins” at points of interest, and outlined potential ways to protect users’ privacy. Puttaswamy et al. (2014) described a tool they developed that shields the actual location data from servers, but still allows users to share location information with each other as desired. Reverse geocoding, or the process of obtaining highly specific location information about users through maps that have points indicating their geographic positions, raises confidentiality and privacy concerns, especially in the case of sensitive topics such as health data (Brownstein et al., 2005; Krumm, 2007).

Very little research has been accomplished in this important area, and more is needed. For now, researchers should consider what levels of scale, accuracy, and precision are absolutely necessary in order to carry out the analysis and visualization. For example, MacEachren et al. (2011a) aggregated frequency counts for tweets to 2 degree grid cells in order not to show point locations.

Conclusion

In conclusion, we would like to emphasize that geospatial analysis of social media is overwhelmingly complex. It requires an interdisciplinary approach that is well-grounded in geospatial, social, linguistics, and information retrieval theories. It is not enough to know only geospatial or social networking methods. It is highly important to understand how social space interacts with geographic space and time. Not all tools can answer the whole spectrum of these questions. Some tools can provide only fragmentary snapshots of phenomena where explicit and implicit contextual differences in individual behaviours are flattened by data classifications. Therefore, they can hinder our understanding of phenomena, rather than enhance it. Emerging visual analytics technologies, however, can help researchers understand how social, temporal, and spatial properties converge.

References

Andrienko, N. and Andrienko, G. (2006),*Exploratory Analysis of Spatial and Temporal Data*: *A Systematic Approach*, Springer Verlag, Berlin.

Andrienko, G., Andrienko, N., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S. I., Jern, M., Kraak, M-J, Schumann, H. and Tominski, C. (2010a),"Space, time and visual analytics", *International Journal of Geographical Information Science*, Vol. 24 No. 10, pp. 1577-1600.

Andrienko, G., Andrienko, N., Mladenov, M., Mock, M. and Pölitz, C. (2010b),"Discovering bits of place histories from people's activity traces", paper presented at the 2010 IEEE Symposium onVisual Analytics Science and Technology (VAST), Salt Lake City, UT.

Anselin, L. (1999), "Interactive techniques and exploratory spatial data analysis", in Longley, P., Goodchild, M.F., Maguire, D. and Rhind, D., *Geographical Information Systems: Principles, Techniques, Management and Applications,*John Wiley and Sons, New York, pp. 253-266.

Backstrom, L., Sun, E. and Marlow, C. (2010), "Find me if you can: improving geographical prediction with social and spatial proximity", paper presented at WWW '10, Raleigh, NC.

Barthélemy, M. (2011),"Spatial networks",*Physics Reports,*Vol. 499 No. 1, pp. 1-101.

Blanford, J. I., Huang, Z., Savelyev, A. andMacEachren, A. M. (2015),"Geo-located tweets: enhancing mobility maps and capturing cross-border movement", *PLoS One*, Vol. 10 No. 6, e0129202, Available: doi:10.1371/journal.pone.0129202

Bo, H. and Baldwin, P C.T. (2012), "Geolocation prediction in social media data by finding location indicative words", paper presented at COLING 2012, Mumbai, India.

Bordogna, G., Ghisalberti, G. and Psaila, G. (2012), "Geographic information retrieval: modeling uncertainty of user's context," *Fuzzy Sets and Systems*, Vol. 196, pp. 105-124.

Brody, H., Rip, M.R., Vinten-Johansen, P.,Paneth, N. and Rachman, S. (2000), "Map-making and myth-making in Broad Street: the London cholera epidemic, 1854", *The Lancet*, Vol. 356, pp. 64-68.

Brownstein, J. S., Cassa, C., Kohane, I. S. and Mandl, K. D. (2005), "Reverse geocoding: concerns about patient confidentiality in the display of geospatial health data, paper presented at the AMIA Annual Symposium, Washington, DC.

Buchel, O. (2013), "Redefining geobrowsing", paper presented at the Annual Meeting of the Association for Information Science and Technology, Montreal, Canada.

Buchel, O. and Sedig, K. (2014), “Making sense of document collections with map-based visualisations: the role of interaction with representations”, *Information Research,* Vol.19 No. 3, paper 631. Available:http://InformationR.net/ir/19-3/paper631.html

Butler, D. (2013, February 13),“When Google got flu wrong”, *Nature*, Vol. 494, pp. 155-156. Available: http://www.nature.com/news/when-google-got-flu-wrong-1.12413

Chang, J. and Sun, E. (2011), “Location3: How users share and respond to location-based data on social networking sites”, paper presented at the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain.

Cheng, Z., Caverlee, J. and Lee, K. (2010), “You are where you tweet: a content-based approach to geo-locating twitter users”, paper presented at the ACM Conference on Information and Knowledge Management (CIKM’10), Toronto, Canada.

Conover, M.D., Davis, C., Ferrara, E., McKelvey, K., Menczer, F. andFlammini, A. (2013), “The geospatial characteristics of a social movement communication network”, *PloS One*, Vol. 8 No. 3, e55957. Available:doi:10.1371/journal.pone.0055957.

Curtin, K. (2007), “Network analysis in geographic information science: review, assessment, and projections”, *Cartography and Geographic Information Science*, Vol. 34 No. 2, pp. 103-111.

DeLongueville, B., Smith, R. S., and Luraschi, G. (2009),“’Omg, from here, i can see the flames!’: a use case of mining location based social networks to acquire spatio-temporal data on forest fires”, paper presented at the 2009 ACMInternational Workshop on Location Based Social Networks, Seattle, WA.

Doytsher, Y., Galon, B. and Kanza, Y. (2010),“Querying geo-social data by bridging spatial networks and social networks”, paper presented at the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, San Jose, CA.

Eccles, R., Kapler, T., Harper, R. and Wright, W. (2008), “Stories inGeoTime”, *Information Visualization*, Vol. 7 No. 1, pp. 3-17.

ESRI (2015), “Spatial analysis” [Online]. Available:http://www.esri.com/products/arcgis-capabilities/spatial-analysis [15 December 2015].

Fabrikant, S. I. (2001a), “Evaluating the usability of the scale metaphor for querying semantic information spaces”, in Montello, D. R., *Spatial Information Theory: Foundations of Geographic Information Science*, Springer, Berlin, pp. 156-171.

Fabrikant, S. I. (2001b), “Visualizing region and scale in semantic spaces”, paper presented at the 20th International Cartographic Conference (ICC 2001), Beijing, China.

Fabrikant, S. and Buttenfield, B. (2001), “Formalizing semantic spaces for information access”, *Annals of the Association of American Geographers*, Vol. 91, pp. 263-280.

Fabrikant, S. and Skupin, A. (2005), “Cognitively plausible information visualization”, in MacEachren, A.M., Kraak, M-J.andDykes, J., *Exploring Geovisualization*, Elsevier, New York, pp. 667-690.

Fekete, J.D. and Plaisant, C. (2002),“Interactive information visualization of a million items”, paper presented at the IEEE Symposium on Information Visualization (INFOVIS 2002), Boston, MA.

Gahegan, M. (1998), “Scatterplots and scenes: visualisation techniques for exploratory spatial analysis”, *Computers, Environment and Urban Systems*, Vol. 22 No. 1, pp. 43-56.

Gahegan, M. (2005), “Beyond tools: visual support for the entire process of GIScience”, in Dykes, J., MacEachren, A.M. and M-J. Kraak, *Exploring Geovisualization,* Elsevier, New York, pp. 83-89.

Gayle,D. (2011), “Privacy storm after police buy software that maps suspects' digital movements” [Online]. Available:http://www.dailymail.co.uk/sciencetech/article-1386191/Privacy-storm-police-buy-Geotime-software-maps-suspects-digital-movements.html [15 December 2015].

Gelernter, J. and Mushegian, N. (2011), “Geo-parsing crisis messages from microtext”, *Transactions in GIS*, Vol. 15 No. 6, pp. 753-773.

Gleason, B. (2013), “#Occupy Wall Street: exploring informal learning about a social movement on Twitter”, *American Behavioral Scientist*, Vol. 57 No. 7, pp. 966-982.

Goodchild, M.F. (2007), “Citizens as sensors: the world of volunteered geography”, *GeoJournal*, Vol. 69 No. 4, pp. 211-221.

Gong, J. (2002), “Clarifying the standard deviational ellipse”, *Geographical Analysis,* Vol. 34 No. 2, pp. 155-167.

Gonzalez, M. C., Hidalgo, C.A. and Barabasi, A.L. (2008), “Understanding individual human mobility patterns”, *Nature*, Vol. 453 No. 7196, 779-782.

Hägerstrand, T. (1970), “What about people in regional science?”, paper presented at the Ninth European Congress ofthe Regional Science Association.

Hecht, B., Hong, L., Suh, B. and Chi, E. (2011), “Tweets from Justin Bieber’s heart: the dynamics of the ‘location’ field in user profiles”, paper presented at the ACM CHI Conference on Human Factors in Computing Systems, Vancouver, Canada.

Hill, L.L. (2006), *Georeferencing: The Geographic Associationsof Information*, MIT Press, Cambridge, MA.

Hyvärinen, O. andSaltikoff, E. (2010), “Social media as a source of meteorological observations”, *Monthly Weather Review*, Vol. 138 No. 8, pp. 3175-3184.

Java, A., Song, X., Finin, T. and Tseng, B. (2007), “Why we twitter: understanding microblogging usage and communities”, paper presented at the Joint 9th WEBKDD and 1st SNA-KDD Workshop, San Jose, CA.

Jiang, B. and Miao, Y. (2015),“The evolution of natural cities from the perspective of location-based social media”, *The Professional Geographer*, Vol. 67No. 2, pp. 295-306.

Karimzadeh, M., Huang, W., Banerjee, S., Wallgrün, J.O., Hardisty, F., Pezanowski, S., Mitra, P. and MacEachren, A.M. (2013,), “GeoTxt: A web API to leverage place references in text”, paper presented at the ACM 7th Workshop on Geographic Information Retrieval, New York.

Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J. and Melançon, G. (2008), “Visual analytics: Definition, process, and challenges”, *Lecture Notes in Computer Science: Information Visualization*, Vol. 4950, pp. 154-175.

Kipp, M.E.I., Choi, I., Beak, J., Buchel, O. and Rasmussen, D. (2014). “User motivations for contributing tags and local knowledge to the Library of Congress Flickr Collection”, paper presented at the Annual Conference of the Canadian Association for Information Science, Victoria, Canada.

Krings, G., Calabrese, F., Ratti, C. and Blondel, V.D. (2009), “Urban gravity: a model for inter-city telephone communication networks”, *Journal of Statistical Mechanics*, L07003.

Koylu, C., Guo, D., Kasakoff, A. and Adams, J. W. (2014), “Mapping family connectedness across space and time”, *Cartography and Geographic Information Science*, Vol. 41 No. 1, pp. 14-26.

Krumm, J. (2007), “Inference attacks on location tracks”, paper presented at the 5th International Conference on Pervasive Computing (PERVASIVE’07), Toronto, Canada.

Lefever, D. W. (1926), “Measuring geographic concentration by means of the standard deviational ellipse”, *American Journal of Sociology*, Vol. 32 No. 1, pp. 88-94.

Longley, P.A.,Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (2005), *Geographic Information Systems and Science*, John Wiley and Sons, New York.

Luo, W. andMacEachren, A. M. (2015),“Geo-social visual analytics”, *Journal of Spatial Information Science*, Vol. 8, pp. 27-66.

MacEachren, A. M. (2004). *How maps work: representation, visualization, and design*. Guilford Press.

MacEachren, A.M. and Kraak, M.J. (1997), "Exploratory cartographic visualization: advancing the agenda", *Computers & Geosciences*, Vol. 23 No. 4, pp. 335-343.

MacEachren, A.M., Stryker, M.S., Turton, I. J. andPezanowski, S. (2010), "HEALTH GeoJunction: place-time-concept browsing of health publications*", International Journal of Health Geographics*, Vol. 9, http://www.ij-healthgeographics.com/content/9/1/23.

MacEachren, A.M., Jaiswal, A., Robinson, A.C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X. and Blanford, J. (2011a), "SensePlace2: GeoTwitter analytics support for situational awareness", paper presented at the 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), Providence, RI.

MacEachren, A.M., Robinson, A. C., Jaiswal, A., Pezanowski, S., Savelyev, A., Blanford, J. and Mitra, P. (2011b),"Geo-twitter analytics: applications in crisis management", paper presented at the25th International Cartographic Conference, Paris, France.

Manni, F., Guerard, E. andHeyer, E. (2004),"Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm", *Human Biology*, Vol. 76 No. 2, pp. 173-190.

Martin, K.E. (2015), "Ethical issues in the Big Data industry", *MIS Quarterly Executive*, Vol. 14 No. 2, pp. 67-85.

Massoudi, K., Tsagkias, M., de Rijke, M. andWeerkamp, W. (2011), "Incorporating query expansion and quality indicators in searching microblog posts", *Lecture Notes in Computer Science: Advances in Information Retrieval*, Vol. 6611, pp. 362-367.

MaxMind. (2015), "GeoIP2 city accuracy", https://www.maxmind.com/en/geoip2-city-database-accuracy.

McKenzie, G. and Janowicz, K. (2015), "Where is also about time: a location-distortion model to improve reverse geocoding using behavior-driven temporal semantic signatures",*Computers, Environment and Urban Systems*, Vol. 54, pp. 1-13.

Mennis, J. and Guo, D. (2009), "Spatial data mining and geographic knowledge discovery – an introduction",*Computers, Environment and Urban Systems*, Vol. 33 No. 6, pp. 403-408.

Middleton, S.E., Middleton, L. and Modafferi, S. (2014), "Real-time crisis mapping of natural disasters using social media", *IEEE Intelligent Systems*, Vol. 29 No. 2, pp. 9-17.

Mitchell, A. (1999), *The ESRI Guide to GIS Analysis, Volume 1: Geographic Patterns and Relationships*, ESRI Press, Redlands, CA.

Monmonier, M. (2005), “Lying with maps”, *Statistical Science*, Vol. 20 No. 3, pp. 215-222.

Monmonier, M. (2014), *How to Lie with Maps*, University of Chicago Press, Chicago.

Nelson, J. (2011),“Data visualization at IDV Solutions.Chalkboard Maps: United States of Craigslist”, http://uxblog.idvsolutions.com/2011/07/chalkboard-maps-united-states-of.html

O’Hare, N. and Murdock, V. (2013), “Modeling locations with social media”, *Information Retrieval*, Vol. 16, pp. 30-62.

Okabe, A., Boots, B., Sugihara, K. and Chiu, S. N. (2009), *Spatial Tessellations: Concepts and Applications of Voronoi diagrams,* John Wiley and Sons, New York.

Orchard, T., Farr, S., Macphail, S., Wender. C. andYoung,D. (2012), “Sex work in the Forest City: sex work beginnings, types, and clientele among women in London, Ontario”, *Sexuality Research and Social Policy*, Vol. 9 No. 4, pp. 350-362.

Padmanabhan, A., Wang, S., Cao, G., Hwang, M., Zhao, Y., Zhang, Z. andGao, Y. (2013), “FluMapper: an interactive CyberGIS environment for massive location-based social media data analysis”, paper presented at the ACM Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery, San Diego, CA.

Peca, I., Zhi, H., Vrotsou, K., Andrienko, N. and Andrienko, G. (2011), “KD-photomap: exploring photographs in space and time”, paper presented at the 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), Providence, RI.

Proulx, P., Tandon, S., Bodnar, A., Schroh, D., Harper, R. and Wright, W. (2006), “Avian flu case study with nSpace and GeoTime”, paper presented at the 2006 IEEE Symposium on Visual Analytics Science and Technology, Baltimore, MD.

Popescu, A. and Grefenstette, G. (2010),“Mining user home location and gender from Flickr tags”, paper presented at the International Conference on Weblogs and Social Media (ICWSM’10), Washington, DC.

Purcell,D. and de Beurs, K. (2013) “It’s hot in here: Twitter as data source of understanding perceptions of heat and drought hazards”, paper presented at the Social Media & Society Conference, Halifax, Canada.

Puttaswamy, K.P.N., Wang, S., Steinbauer, T., Agrawal, D., El Abbadi, A., Kruegel, C. and Zhao, B.Y. (2014), “Preserving location privacy in geosocial locations”, *IEEE Transactions on Mobile Computing*, Vol. 13 No. 1, pp. 159-173.

Ramdani,D. (2011),“GPS applications on cellular phone with geoid addition to height” [Online]. Available: http://mycoordinates.org/gps-applications-on-cellular-phone-with-geoid-addition-to-height/ [15 December 2015].

Rorissa, A., Rasmussen Neal, D., Muckell, J. and Chaucer, A. (2012), “An exploration of tags assigned to still and moving images on Flickr”, in Rasmussen Neal, D., *Indexing and Retrieval of Non-Text Information*, De Gruyter Saur, Berlin, pp. 185-211.

Sayyadi, H., Hurst, M. and Maykov, A. (2009),“Event detection and tracking in social streams”, paper presented at the International Conference on Weblogs and Social Media (ICWSM), San Jose, CA.

Scellato, S., Noulas, A., Lambiotte, R. andMascolo, C. (2011), “Socio-spatial properties of online location-based social networks”, paper presented at ICWSM-11, Barcelona, Spain.

Schreck, T. andKeim, D. (2013), “Visual analysis of social media data”, *Computer*, Vol. 46 No. 5, pp. 68-75.

Sevillano, X., Valero, X. and Alías, F. (2015), “Look, listen and find: a purely audiovisual approach to online videos geotagging”, *Information Sciences*, Vol. 295, pp. 558-572.

Shekhar, S. and Huang, Y. (2001),“Discovering spatial co-location patterns: a summary of results”, *Lecture Notes in Computer Science: Advances in Spatial and Temporal Databases*, Vol. 212, pp. 236-256.

Sobkowicz, P., Kaschesky, M. and Bouchard, G. (2012),“Opinion mining in social media: modeling, simulating, and forecasting political opinions in the web”, *Government Information Quarterly*, Vol. 29 No. 4, pp. 470-479.

Song, C., Qu, Z., Blumm, N. and Barabási, A.L. (2010), “Limits of predictability in human mobility”, *Science*, Vol. 327 No. 5968, pp. 1018-1021.

Stefanidis, A., Cotnoir, A., Croitoru, A., Crooks, A., Rice, M. andRadzikowski, J. (2013), “Demarcating new boundaries: mapping virtual polycentric communities through social media content”, *Cartography and Geographic Information Science*, Vol. 40 No. 2, pp. 116-129.

Sui, D. andGoodchild, M. (2011),“The convergence of GIS and social media: challenges for GIScience”, *International Journal of Geographical Information Science*, Vol. 25 No. 11, pp. 1737-1748.

Torniai, C., Battle, S. and Cayzer, S. (2007), “Sharing, discovering and browsing geotagged pictures on the web” [Online]. Available:http://www.hpl.hp.com/techreports/2007/HPL-2007-73.pdf [15 December 2015].

Tufte, E.R. (1991), “Envisioning information”, *Optometry & Vision Science*, Vol. 68 No. 4, pp. 322-324.

Vicente, C.R., Freni, D., Bettini, C. and Jensen, C.S. (2011), “Location-related privacy in geo-social networks”, *IEEE Internet Computing*, Vol. 15 No. 3, pp. 20-27.

Wade, T. andSommer, S. (2006), *A to Z GIS: An Illustrated Dictionary of Geographic Information Systems*, ESRI Press, Redlands, CA.

Wagner,D.(2013), "Google flu trends wildly overestimated this year's flu outbreak" [Online]. Available:http://www.thewire.com/technology/2013/02/google-flu-trends-wildly-overestimated-years-flu-outbreak/62113/[15 December 2015]

Weidemann, C. (2013),"Social media location intelligence: the next privacy battle – an ArcGIS add-in and analysis of geospatial data collected from twitter.com", *International Journal of Geoinformatics*, Vol. 9 No. 2. Available: http://journals.sfu.ca/ijg/index.php/journal/article/view/139 [15 December 2015].

Weiler, M., Schmid, K.A., Mamoulis, N. and Renz, M. (2015), "Geo-social co-location mining", paper presented at the Second International ACM Workshop on Managing and Mining Enriched Geo-Spatial Data (GeoRich '15), Melbourne, Australia.

Whitsel, E.A., Quibrera, P.M., Smith, R.L., Catellier, D.J., Liao, D., Henley, A.C. and Heiss, G.(2006), "Accuracy of commercial geocoding: assessment and implications",*Epidemiological Perspectives & Innovations*, Vol. 3 No. 8.Available: doi: 10.1186/1742-5573-3-8.

Yi, J.S., ah Kang, Y., Stasko, J.T. and Jacko, J.A. (2007), "Toward a deeper understanding of the role of interaction in information visualization", *IEEE Transactions onVisualization and Computer Graphics,* Vol. 13 No. 6, pp. 1224-1231.