REPRESENTING AND ANALYZING LOCATION-BASED SOCIAL MEDIA ACTIVITY IN GIS

by

XUEBIN WEI

(Under the Direction of Xiaobai Yao)

ABSTRACT

Human activities are generated due to physiological, psychological and economic needs, and have spatial, temporal, and social elements. In the era of big data, geo-tagged social media are becoming new platforms that influence human behaviors in space and time, and are also serving as new channels for geographers to observe human interactions and social connections at fine scales. Traditional geographical representation and analysis methods in GIS are not sufficient to tackle the much more complex nature of the location-based social media activity. The convergence of GIS and social media has resulted in data avalanche and requires new theories in GIScience. This dissertation has developed methods and tools to represent and analyze location-based social media activates (LBSMA) in GIS in three perspectives: (1) this research has proposed a conceptual framework of location-based social media activity to model human activities in spatial-temporal-social dimension , and has implemented this data framework to organize LBSMA in GIS and produce practical tools to calculate useful measurements; (2) this research has developed a random walking algorithm to characterize urban road networks by calculating the possibility distribution of human locations over time; and (3) this research has introduced location-based social connections to visualize and quantify social connections in spatial-temporal dimension. In addition, this research has established a dedicated website (www.lbsoical.net) to extract and analyze the real-world social media data, and provide visualization and analysis function for various studies. The developed methods and tools in this research can organize, visualize, simulate and analyze human activities in spatial-social-temporal dimension. Those methods have added to our understanding of human interactions by providing innovative and applicable measures for places, social connections and human activities. The findings from this research have yielded new insights regarding human activities in virtual and physical space, and have enhanced technical capabilities for social media analysis in GIS. The developed methods can help identify place-based or people-based strategies, e.g., urban planning, traffic planning, commercial advertising or energy communicating. The proposed framework will pave new avenues for future research, such as public health, transportation, urban geography and social science.

INDEX WORDS: Location-Based Social Media, Geographic Information System, Random Walk, Location-Based Social Connection

REPRESENTING AND ANALYZING LOCATION-BASED SOCIAL MEDIA ACTIVITY IN GIS

by

XUEBIN WEI

B.S.E, Wuhan University, China, 2008

M.E, Wuhan University, China, 2010

M.S, University of Twente, The Netherlands, 2010

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2015

REPRESENTING AND ANALYZING LOCATION-BASED SOCIAL MEDIA ACTIVITY IN GIS

by

XUEBIN WEI

Major Professor:	Xiaobai Yao
Committee:	Marguerite Madden
	Lan Mu
	Shih-Lung Shaw

Electronic Version Approved:

Suzanne Barbour

Dean of the Graduate School

The University of Georgia

August 2015

ACKNOWLEDGEMENTS

I would like to express my sincerely appreciation to whom I met in the Department of Geography, the University of Georgia. In particular, I am deeply grateful to my major processor, Dr. Xiaobai Yao, for her valuable advices, tremendous supports, and strongly encourages during my four-year study and research. I also appreciate my committee members: Dr. Marguerite Madden, Dr. Lan Mu, and Dr. Shih-Lung Shaw, for their valuable comments, constructive feedback and enormous help. I would like to express my appreciation to all the faculties, staff members and students of the Geography Department who made me feel at home during my doctoral study.

Finally, this project is supported by the 2014 Innovative and Interdisciplinary Research Grant from the Graduate school of the University of Georgia.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION AND LITERATURE REVIEW

Background

Geography is the science of discovering, investigating and understanding nature, as well as humans and their interactions; it is the discipline that studies the social and environmental phenomena of geographic space (Goodchild 2010). Geographic Information Science (GIScience), a branch of geography, describes, analyzes, models, reasons about, and makes decisions on phenomena distributed on the surface of the earth (Wright, Goodchild, and Proctor 1997). Research on human activities plays an important role in geography. People in GIS are resource contributors and a study target, and the average citizen is both a consumer and a producer of geographic information (Goodchild 2010). Location awareness is embedded in the use of mobile phones, handheld devices and laptops (Licoppe and Inada 2008). Meanwhile, location-based and people-based data organization has been applied in time geography to model human activities in physical space (Miller 2003). Human activities in virtual space (the use of computers or telephone to communicate) also gains interests from the study of virtual geography (Batty 1997).

Human activities in physical space and virtual space can influence each other (Yu and Shaw 2008). With the popularization of location-based social networking, social media has moved from virtual space to physical place. The convergence of GIS and social media has resulted in data avalanche and requires new theories in GIScience, e.g., real-time monitoring, formalizing place and multimedia representation (Sui and Goodchild 2011). Cyber geography has been introduced to investigate interconnected spatial patterns and relationships between cyberspace and the real world. In the conceptual framework of cyber geography , place, which can also be considered as a spatial-temporal construct (Yao 2010), combines temporal information and social media content (Tsou and Leitner 2013); a message from social media is considered an extension of the human mind (Tsou and Leitner 2013).

Research Objectives

This research defines the location-based social media activity (LBSMA) as a subset of human activity, which occurs in the real world with the contents advertised on social media (Figure 1-1). The location-based social activities not only inhabit geographical and temporal positions, but also embody semantic messages and social associations of human activities. In addition to answering *where* and *when* in traditional human activities, the location-based social activities can reveal *what* and *who*, and potentially indicate *why* and *how* of human activities. This research has answered the question: how to represent and analyze the location-based social activities in GIS?

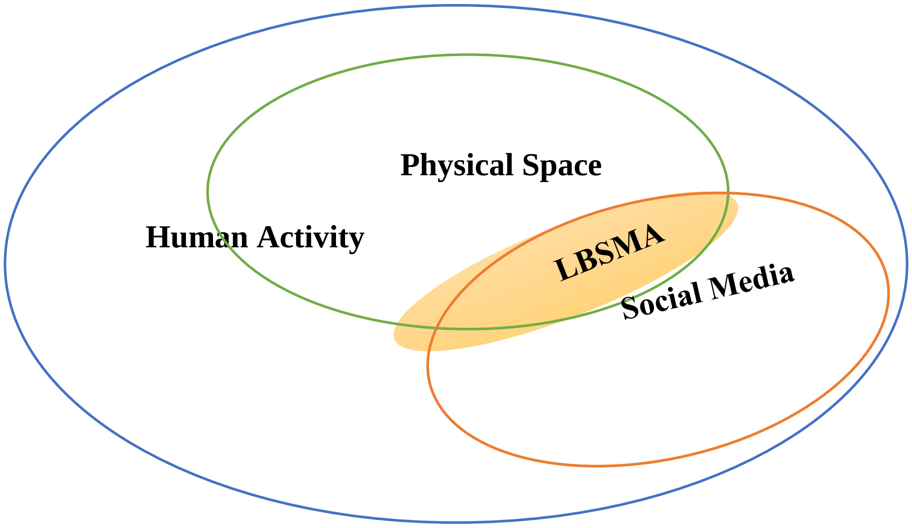


Figure 1-1 Definition of Location-Based Social Media Activities (LBSMA)

Overarching Research Goal: representing and analyzing location-based social media activity (LBSMA) in GIS

- Objecvite-1: This research has developed a framework of LBSMA data that provides theoretical and technical foundations for the representation of LBSMA in GIS.

Connected with social media data, LBSMA can provide additional information on human activities in a spatial-temporal-social dimension. A comprehensive data conceptualization of location-based social activity is required to effectively and efficiently organize such kind of data in GIS. A spatial-temporal-social dimensional data model of LBSMA has been developed in order to provide the theoretical and technical foundation for analysis of LBSMA.

- Objecvite-2: This research has analyzed human activity in spatial, temporal, and social dimension, and provided innovative and feasible measures of places, people and social connections.

Road networks are essential components of urban places. The relationships between road networks and human activities are mutually dependent. This research has developed a simulation algorithm that can mimic behaviors of human activities to calculate the possibility location of human activities in spatial-temporal dimension. The calculated values can serve as measurements of urban areas in terms of spatial importance indicators that are associated with social-economic characteristics of surround areas.

Substantial measures of human activities are also needed in GIS to handle the emerging social media data. The measurement of social association in traditional social network analysis, for instance, is only a simple indicator of a static network parameter. Based on the proposed LBSMA model, the social connections have been quantified in a spatial-temporal dimension where human activities can be visualized and analyzed based on time, location and features of social interaction.

Significance of Study

Innovative Model of LBSMA Data

This research has filled the blank of a suitable data representation model for human activity in a spatial-temporal-social dimension. This project has designed an innovative data model for LBSMA analysis by integrating the social networks model with the human activity model. Previous studies only focus on one aspect of human behavior, either physical movements in the spatial-temporal dimension or structure and interaction exploration in the social networks. However, human activities are generated due to physiological, psychological and economic needs and have spatial, temporal, and social elements (Ronald et al. 2012) . Human activities are constrained by physical and social conditions. Mobile positioning data cannot answer why people move that way (Torrens, Li, and Griffin 2011). Therefore, by connecting the spatial-temporal dimension of human activities to their corresponding social association, this research is able to provide a comprehensive LBSMA data model that can represent and analyze human activities in a spatial-temporal-social dimension.

Applicable Methods for LBSMA Analysis

People produce voluminous location-based data and valuable information in social media. Although several attempts were made to analyze the location-based social media data (ESRI 2013; Geofeedia 2014), systematic and substantial methodologies for social media data in GIS are absent. This research has developed ways to make use of LBSMA data for different purposes. The research has employed spatial-temporal analysis, semantic extraction and machine learning technics to create a taxonomy and classification system of LBSMA. New measurements of human activities and places are proposed by incorporating the social contents and social interactions from social media. Potential position of human activities under spatial-temporal-social constraints can be simulated to quantify spatial importance of road networks. Such simulation is valued in different scenario planning practices.

Literature Review

A set of word clouds of a 5-year interval in Figure 1-2 conceptually summarizes the related topics from the literature since the 1990s. In the first 5-year interval, spatial database structure and corresponding data mining/ knowledge discovery are the major topics. Network analysis gained concentration at the second stage where research centers on structure and relationship analysis of social network. Starting from the 2000s, mobile phones connected with GPS and the Internet became popular, which triggered the studies of real-time tracking and modeling of human activities. In the 2010s, the possibility of harvesting data from social media motivates the new research interests of social events, i.e., virtual activities in social media.

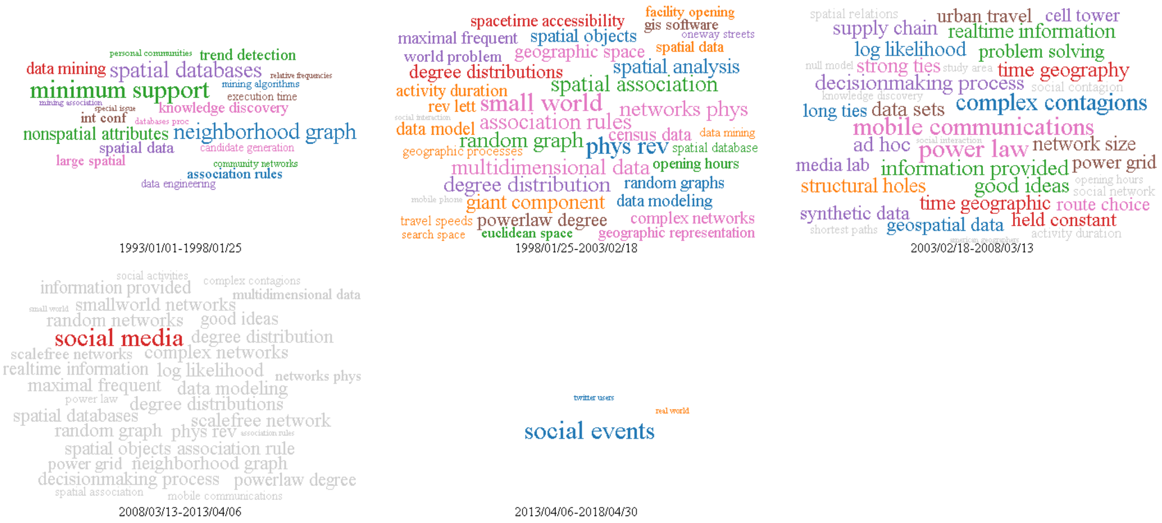


Figure 1-2 Word Cloud from Literature by 5-year Interval

Human Activity

Research of human behavior holds an important position in the geographic study. The study of human spatial behavior covers a wide range of topics, including travel and way-finding, migration and residential mobility, decision making and choice behavior, as well as spatial cognition and environmental perception (Kwan 2010). Space, time and space-time have been recognized as complex social and cultural constructions as well as geometric dimensions in geography (Merriman 2012). Time, space and social differentiation are frequently examined in critical geography (Schwanen and Kwan 2012).

The location-based organization of data of GIS fits with the place-based theories and models of transportation and urban form but ignores the basic spatial-temporal conditions of human activities. Time geography offers a people-oriented extension to place-based tools in GIS (Miller 2003).

- Time Geography

Time consciousness was discussed in human ecology (time space diagrams), Marxism (internalized sense of time discipline) and the ideology of everyday time practice (documents or texts/ devices or instruments/ disciplining or drilling people to routinize a set of practices) (N. J. Thrift 1996). All time is relative and rooted in the context in which people observe or experience it (Dodgshon 2008). However, traditional regional geography, a science of exploring the interaction of environment and human behavior, cannot answer the following question: What is the need of social distribution? Meanwhile, social organization and micro-level technology of individual behavior were also neglected (Hägerstrand 1970). There were only abstract parameters of time that were devoid of all differentiating content or sequence in the 1960s (Dodgshon 2008).

Hägerstrand, therefore, introduced the time coordination in regional geography to resolve those problems in the 1970s. He argued that an individual has only one role, occupies one location in a given time duration, and the time scale is continuous. Accordingly, a time-space model had been constructed in which three constraints were proposed: capability constraints limited by biological constructions or tools, coupling constraints limited by communication among people, and authority constraints consisted in a hierarchy of domains (Hägerstrand 1970). Anderson also proposed a concept of time budget in which the time-budget diary incorporates the timing, duration and location of an individual's activities (Anderson 1970). Individual movement implies a trade-off between the inseparability and scarce nature of space and time, and is conditioned by various constraints and opportunities. These ideas became known as time geography (Kuijpers et al. 2010). Time geography is not an attempt to predict exact travel behavior but indicates individual travel possibilities (Neutens et al. 2008).

Hägerstrand's space-time geographies had a significant impact on geographical thinking (Merriman 2012). However, the space-time analysis in which time is both experientially and spatially referenced didn't say how the space and time involved were constructed or experienced (Dodgshon 2008). Schwanen and Kwan also demand a comprehensive and systematic analysis of mutual implication among space, time and social differentiation, and a pluralistic approach in terms of theory and methodology. They argue that time, space and social differentiation should be coupled in the study of practices or phenomena in critical space-time geographies (Schwanen and Kwan 2012). For example, Information Communication Technology (ICT) uses leads for the relaxation of some of the space-time constraints that limit people's mobility and activity space, and creates new topologies of spatial interaction of human activities (Kwan 2007).

- Representation of Human Activity in GIS

GIS is considered as digital electronic data storage that produces spatial representations. Two central characteristics of GIS are the storage of digital data and the production of electronic spatial representations. Therefore, GIS can be regarded as a set of tools, technologies, approaches and ideas that are vitally embedded in broader transformations of science, society and culture (Pickles 1995).

Geographic data models must serve as an acceptable reflection of the real-world phenomena. GIS, therefore, requires a critical theory reflecting sustained interrogation of the ways in which the use of technology, and its products reconfigure broader patterns of culture or political relations (Pickles 1995). GIS visualization, for instance, can establish important connections between large-scale phenomena and the everyday lives of individuals. Critical agency of GIS users can play a significant role, for example in the subjectivities of GIS practitioners, and the ability of GIS to understand individual experiences (Kwan 2002).

The ontological issues of time in GIS include linear or cyclic views of time, multiple times (world time, valid time, user defined time), continuous or discrete change, branching time and alternative timelines, etc. (Peuquet 2001). The first in-depth analytical treatment of time-geography was conducted by Lenntorp in the late 1970s. The geo-computation method has been adopted in most time-geography research on individual accessibility. 3D geo-visualization is also widely used to facilitate the identification and interpretation of spatial patterns and relationships (Kwan 2004).

Time geography is a constraint-oriented approach for understanding human activities in a spatial and temporal dimension. Human activity was presented based on the necessary conditions of human participation in physical or virtual space (Miller and Bridwell 2009).

Three core entities are defined in GIS to represent human activities: space-time prisms, space-time path and potential path areas (Figure 1-3). Space-time path traces an individual's movements in space with respect to time. It reveals the varying constraints and activity space of individuals. Space-time prism delimits possible locations for the space-time path during a period of potential activity participation. Potential path areas, on the other hand, are the projection of a space-time prism to the geographical plane (Neutens et al. 2008). They are direct measures of individual accessibility to an environment and available resources (Miller and Bridwell 2009).

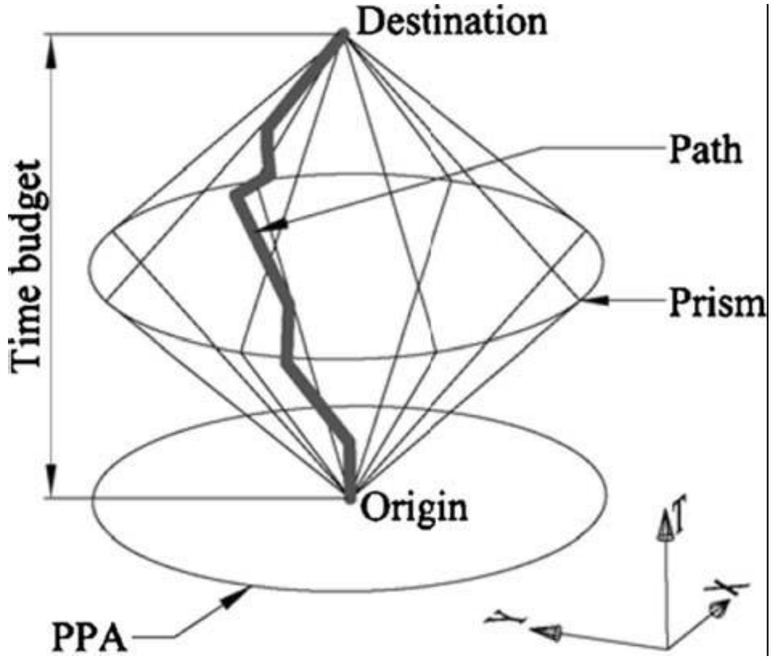


Figure 1-3 Space-Time Path, Prism and Potential Path Area (Neutens et al. 2008)

GIS is able to facilitate the exploration of spatiotemporal patterns in a large dataset. Rooted in Hägerstrand’s time geography, Shaw, Yu and Bombom presented a generalized space-time path approach for visualization and exploration of spatiotemporal changes of individuals by identifying spatial cluster centers of observations and connecting those centers according to the temporal sequence (Shaw, Yu, and Bombom 2008). Those cluster algorithms in temporal GIS include k-means, Clara and genetic algorithms (Adnan et al. 2010). Shaw and Yu also extended their work by incorporating the human activities in virtual spaces (Shaw and Yu 2009). Chen et al proposed an Activity Pattern Analyst Extension in ArcGIS which is able to generate, filter segment and query space-time path and perform spatial-temporal density and path clustering analysis (Chen et al. 2011).

The time-space prism/path has arbitrary spatial and temporal resolutions and is explicit with respect to informational assumptions (Tavakoli and Fakhraie 2011). Accordingly, some improved time-space methods have evolved. For example, by arguing that classic time geography only admits uniform travel velocities, Miller and Bridwell have developed an analytical time geographic model in which travel velocities vary continuously across space (Miller and Bridwell 2009). Kuijpers et al. also criticize classic space-time prism which assumes that the start and end points, i.e., anchors, of an individual are perfectly known or fixed. Accordingly, they have generalized the concept of anchor points to anchor regions to allow prism anchors to be uncertain or flexible, and hence increase the pliability of observations of anchor points (Kuijpers et al. 2010). Yu and Shaw have extended Hagerstand's spatial-temporal model by incorporating the concept of virtual space, namely a place where cyber-communication or interaction occurs. A prototype model of adjusted space-time prism has been implemented for analyzing potential human activities, e.g., available opportunities under specific space/time constraints (Yu and Shaw 2008).

In the 1990s, the emphasis of resource management shifted from inventory toward maintenance. How to represent geographical phenomena in time, as well as in space, became a problem in GIS. Several data models were built to represent temporal relationship and patterns of change, including snapshot images and variable-length lists (Peuquet and Duan 1995). All those models are location based. In addition, an event-based spatial data model was devised by

Peuquet and Duan in which time line had recorded the changes of a single thematic domain. The advantage of such a model was the increased ability to perform temporal manipulations (Peuquet and Duan 1995). A Spatiotemporal Triad Framework was also devised by Peuquet which adopts a triad representational framework including object based representation, location based representation and time based representation (Peuquet 1994).

Dragicevic and Marceau argue that time can only be observed through apparent changes of objects and their attributes occurring in space; the paradigm of temporal data representation in GIS is still unresolved. One possible means to study a dynamic phenomenon that happened during the past is to apply temporal interpolation between consecutive snapshots by simulating the change that occurred during the interval based on fuzzy set theory (Dragicevic and Marceau 2000).

Neutens et al. argue that classical time-geographic concepts are inadequate for dealing with real-world complexities due to the heterogeneous environment. They have advanced a hybrid CAD-system utilizing the robust 3D design tools in CAD for model construction and data storage. Such a hybrid system can capture the interpersonal variations in accessibility by accounting for the individual’s time budget (Neutens et al. 2007, 2008). Kim and Kwan also developed an operational method and GIS-based algorithm that better represents the space-time characteristics and human behavior by considering opportunities, possible activity duration and effect of the transport network (Kim and Kwan 2003).

Yi et al. focus on the social activities rather than the physical activities. They argue that the temporal consideration of a social network, especially the edge attribute (life cycle), is necessary. An adjacency matrix representation, namely TimeMatrix, was devised for the temporal social network analysis. In the TimeMatrix, two-dimensional cells were designed to represent both temporal and structural attributes of social networks. Different representation technologies, e.g., semantic zooming, aggregation and collapsing, overlays and filtering, were also adopted (Ji, Niklas, and Lee 2010). Mennis also devised a spatio-temporal raster data structure based on multidimensional map algebra (Mennis 2010).

Goodchild and Cova have introduced an innovative GIS representation model to handle some dilemmas in time geography, e.g., discrete vs. continuous or spatial vs. temporal. A geo-atom was defined as “an association between a point location in space–time and a property”. Other concepts, including geo-fields (aggregation of geo-atoms), geo-object (an aggregation of points in space–time whose geo-atoms meet certain requirements), geo-dipole (as a tuple connecting a property and value not to one location in space–time as in the case of the geo-atom but to two) and metamap were constructed based on the geo-atom (Goodchild, Yuan, and Cova 2007). Such hybrid representation has been implemented lately and is capable of answering the questions about relationships among location, time and attributes (Pultar et al. 2010).

Kwan suggested a hypertext metaphor for modeling human spatial interactions in which each individual has several nodes to different social networks while each link represents interactive coordination between two individuals (Kwan 2007).

Human behaviors can also be represented in terms of *rules* which is an implication of the form $A \rightarrow B$, where *A* and *B* are sets of attributes. Mining frequent item-sets and association rules in a spatial-temporal dimension are the major challenges (Gidofalvi and Pedersen 2005).

- Analysis of Human Activity

Accessibility measurement is one of the major topics in human activity analysis and transport planning. Accessibility measures the capacity of a location or object to be reached by, or to reach different, locations (Rodrigue, Comtois, and Slack 2013). Based on the time-space path representation of human activities, three complementary perspectives of measuring human accessibility in spatial-temporal dimension have been summarized by Miller: constraints-oriented approach, attraction-accessibility measures and benefit measures (Miller 1999). Kim and Kwan measured spatial-temporal accessibility by incorporating the effect of minimum activity participation time and the maximum travel time threshold. In their calculation, the individual accessibility measure equals the sum of weighted areas of opportunities multiplying possible activity participation time of all potential path areas (Kim and Kwan 2003).

With the prevalence of mobile devices, gathering and recording detailed individuals' activities has become available. Mobile phones are ideal vehicles to study human activities for both individuals and organizations (Eagle and Sandy Pentland 2006). The mobile phone data has been utilized to classify identifiable routines in people's daily life (Eagle and Sandy Pentland 2006), and understand road usage patterns in urban areas (Wang et al. 2012). Human movements can also be simulated and predicted by applying a machine learning scheme on collected individual behavioral data (Torrens, Li, and Griffin 2011).

Various simulation models have been developed to imitate human activities and responses for better understanding of the complex systems, such as cities (Benenson and Torrens 2004). Random Walk is a special type of simulation modeling. The concept was initially proposed by Prof. Karl Pearson, in 1905, to estimate the probability that after n steps of random walks a person is at a distance between r and r + d from his starting point (Pearson 1905). Some modified versions of random walks include self-avoiding walks, self-attracting walks and correlated random walks (de Smith 2010). This method has gained many interests in a variety of research, ranging from electric communication (Leng et al. 2007), social problems (Short et al.

2008) to complex networks (Yang 2005). In time geography, random walk simulation was used to predict the probability distribution of an agent's location over time (Winter and Yin 2011). In urban studies, the random walk algorithm has been applied to detect community structures (Pons and Latapy 2005) and to identify urban isolations (Volchenkov and Blanchard 2007).

Social Network

Milgram performed an interesting experiment which reveals a surprising fact: only five links are required on average to connect any two people in a social network (Milgram et al. 1965). Milgram's experiment invokes the concept of "small-world", and the model he applied to describe the social relationship has been widely adopted and improved in following research (Kleinfeld 2002). A social network is a collection of social ties among friends in which people are represented as points and acquaintance with others represents links. The Milgram model has a similar network structure in which some pairs of these objects are connected by links, and the cause-effect relationship can be subtle (Easley and Kleinberg 2010).

- Social Network Analysis

A social network is a typical network that traditional network measures from graph theory can also apply to social network analysis. The vertex connectivity of large networks normally follows a scale-free power-law distribution governed by a robust self-organizing phenomenon, e.g., continuous expansion and preferential attachment (Barabasi and Albert 1999). Statistical methods can handle millions of vertices and tell what the network looks like (Newman 2003). Traditional network analysis involves plenty of measures in terms of connectivity, centrality and vulnerability, etc. Degree of a vertex in a graph, for example, is the number of edges that have this vertex as an end vertex. Centrality indices quantify an intuitive feeling that some vertices or edges are more central than others in most networks. In centrality measurements, the betweenness is calculated based on the set of shortest paths in a graph, while the closeness is calculated based on the total distance from one vertex to the other entire vertex. Eigenvector centrality indicates the central importance of one vertex depending on its neighbors. The Connectivity, on the other hand, describes the strength of connections between vertices with respect to the number of vertex- or edge-disjoint paths (Brandes and Erlebach 2005). Modularity can be applied to distinguish communities of networks by considering the edges that fall within a community between a community and the rest of the network (Newman 2006). Strongly connected components refer to the nodes within the component that can be reached from every other node in the component by following directed links; and weakly connected components are the nodes that can be reached from every other node by following links in either direction. If the largest component encompasses a significant fraction of the graph, it is then called the giant component (Newman 2003). Social networks are highly dynamic objects, which grow and change quickly over time. Link-prediction provides such a solution that given a snapshot of a social network at a given time, one can predict the edges that will be added into the network during the interval from time to time (Liben-Nowell and Kleinberg 2007).

A message from social media adds a significant and important dimension of information wherein the qualitative semantic data and information embedded in the social media data needs further processing and normalization to allow quantitative analyses (Bahir and Peled 2013). The semantic message cannot be directly processed. For example, GIS still cannot unambiguously recognize and sufficiently perform spatial reasoning with place names in linguistic expressions (Vasardani, Winter, and Richter 2013). Additional semantic analysis is required. Mika introduced the *Flink* system to extract, aggregate and visualize social network content. The author argues that the first challenge is the extraction, representation and aggregation of social knowledge (Mika 2005). Lampos and Cristinini searched symptom-related statements from Twitter and turned the information into a flu-score generated by a machine learning technology to monitor the flu pandemic (Lampos and Cristianini 2010).

- Social Media Research

From an individual perspective, social media activities can be defined as interactions with others through cyber space. Therefore, social activities can be subtracted from the media where people interact with others. Different social media leads to different social relationships. For example, Twitter fosters an asymmetric network structure where people prefer to broadcast individual content, while LinkedIn and Facebook aim to capture pre-existing ties by focusing on social interactions among friends (Takhteyev, Gruzd, and Wellman 2012). Previous studies have investigated the content and friendship structure on Twitter (Takhteyev, Gruzd, and Wellman 2012), Facebook (Lewis et al. 2008) and Weibo (Li et al. 2013). The spatial distribution of location-based social activities from different social media has also been explored in recent studies (Jiang and Miao 2014; Liu et al. 2014). GIS is also considered as a media in terms of communicating and sharing knowledge and supporting location-based social networking (Sui and Goodchild 2011). For example, Geofeedia is such a website that provides online services of location-based streaming, search, monitoring, and analytics of social media content (Geofeedia 2014). ESRI also provides a GeoEvent Processor Extension which can connect to social media providers, process and filter real-time data, monitor assets and update maps (ESRI 2013).

Organization of This Dissertation

This dissertation consists three papers, an introduction chapter and a conclusion chapter. The introduction chapter describes the background of studies in human activities and location-based social media, and addresses the two research objectives.

Chapter 2 introduces the proposed conceptual data framework of location-based social media activity and implemented pilot prototype system in GIS. A dedicated website is established to extract and visualize real world location-based data. Those data are organized in the proposed data framework, and several tools of visualizing and analyzing location-based social activity is implemented in the pilot prototype. The usefulness of the pilot prototype is proved in a case study where Facebook data is extracted, organized, and analyzed in spatial-temporal-social dimension.

A random walking algorithm is developed in Chapter 3 to characterize the spatial importance of urban road networks by calculating the possibility distribution of human activity in temporal-spatial dimension. The proposed random walking algorithm can simulate human behaviors in traveling with the consideration of physical and topological characteristics of road networks and human preferences. The research found that the calculated random walking values of road networks are highly correlated to some important social-economic variables, and thus can serve as a spatial importance indicator of road networks in urban planning, traffic simulation and etc.

Chapter 4 analyzes location-based social media activity by focusing on the spatial-social dimension. With the data collection website, an extracting social connection methodology is developed, and a definition of location-based social connections is introduced. Based on the proposed location-based social connection, the social relationships can be spatialized and quantified in GIS that can visualize and identify spatial-social clusters in GIS.

Chapter 5 summarizes the work of this dissertation and addresses the future research of representing and analyzing location-based social media activity in GIS.

References

Adams, P. 1998. Network topologies and virtual place. *Annals of the Association of American Geographers* 88 (1):88.

Aderamo, A. J., and S. A. Magaji. 2010. Rural Transportation and the Distribution of Public Facilities in Nigeria: A Case of Edu Local Government Area of Kwara State. *Journal of Human Ecology* 29 (3):171–179.

Adnan, M., A. D. Singleton, P. A. Longley, and C. Brunsdon. 2010. Towards Real-Time Geodemographics: Clustering Algorithm Performance for Large Multidimensional Spatial Databases. *Transactions in GIS* 14 (3):283–297.

Anderson, J. 1970. Time-Budgets and Human Geography. *Area* 2 (1):50–51.

Atlanta Regional Commission. 2004. Commute Options. *www.atlantaregional.com*. http://www.atlantaregional.com/transportation/commute-options (last accessed 7 January 2014).

Backstrom, L., E. Sun, and C. Marlow. 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. New York, NY, USA: ACM http://doi.acm.org/10.1145/1772690.1772698 (last accessed 10 December 2014).

Bahir, E., and A. Peled. 2013. Identifying and Tracking Major Events Using Geo-Social Networks. *SOCIAL SCIENCE COMPUTER REVIEW* 31 (4):458–470.

Barabasi, A.-L., and R. Albert. 1999. Emergence of Scaling in Random Networks. *Science* 286 (5439):509.

Baran, P. K., D. A. Rodr ǵuez, and A. J. Khattak. 2008. Space Syntax and Walking in a New Urbanist and Suburban Neighbourhoods. *Journal of Urban Design* 13 (1):5–28.

Batty, M. 1997. Virtual geography. Time and Space Geographic Perspectives on the Future., eds. M. Batty and S. Cole, 337–352. Great Britain, BUTTERWORTH-HEINEMANN.

Behrens, R. B., and L. A. Kane. 2004. Road capacity change and its impact on traffic in congested networks: evidence and implications. *Development Southern Africa* 21 (4):587–602.

Benenson, I., and P. M. Torrens. 2004. *Geosimulation: automata-based modelling of urban phenomena*. Hoboken, NJ: John Wiley & Sons.

Blanchard, P., and D. Volchenkov. 2008. Exploring Urban Environments By Random Walks. *arXiv:0801.3216v1 [physics.soc-ph]* 1021:183–203.

———. 2010. Random Walks Estimate Land Value. *arXiv:1003.0384 [physics]* :1–15.

Bono, F., E. Gutiérrez, and K. Poljansek. 2010. Road traffic: A case study of flow and path-dependency in weighted directed networks. *Physica A: Statistical Mechanics and its Applications* 389 (22):5287–5297.

Brandes, U., and T. Erlebach. 2005. *Network analysis methodological foundations*. NewYork: Springer Berlin Heidelberg.

Brin, S., and L. Page. 2012. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer Networks* 56 (18):3825 – 3833.

Butts, C. T., R. M. Acton, J. R. Hipp, and N. N. Nagle. 2012. Geographical variability and network structure. *Social Networks* 34 (1):82–100.

Chan, S. H. Y., R. V. Donner, and S. Lämmer. 2011. Urban road networks — spatial networks with universal geometric features? *The European Physical Journal B - Condensed Matter and Complex Systems* 84 (4):563–577.

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

Chen, J., S.-L. Shaw, H. Yu, F. Lu, Y. Chai, and Q. Jia. 2011. Exploratory data analysis of activity diary data: a space–time GIS approach. *Journal of Transport Geography* 19 (3):394–404.

Chen, Q., and S. Chen. 2007. A highly clustered scale-free network evolved by random walking. *Physica A: Statistical Mechanics and its Applications* 383 (2):773–781.

Cho, E., S. A. Myers, and J. Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1082–1090. ACM.

Clauset, A., C. R. Shalizi, and M. E. Newman. 2009. Power-law distributions in empirical data. *SIAM review* 51 (4):661–703.

Crandall, D. J., L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. 2010. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences* 107 (52):22436–22441.

Cranshaw, J., R. Schwartz, J. I. Hong, and N. M. Sadeh. 2012. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In *ICWSM*. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/download/4682/4967 (last accessed 6 April 2014).

Croitoru, A., A. Crooks, J. Radzikowski, and A. Stefanidis. 2013. Geosocial gauge: a system prototype for knowledge discovery from social media. *International Journal of Geographical Information Science* 27 (12):2483 – 2508.

Croxford, B., A. Penn, and B. Hillier. 1996. Spatial distribution of urban pollution: civilizing urban traffic. *Science of The Total Environment* 189:3–9.

Dodgshon, R. A. 2008. GEOGRAPHY'S PLACE IN TIME. *Geografiska Annaler Series B: Human Geography* 90 (1):1–15.

Dragicevic, S., and D. J. Marceau. 2000. A fuzzy set approach for modelling time in GIS. *International Journal of Geographical Information Science* 14 (3):225–245.

Eagle, N., A. Pentland, and D. Lazer. 2009. Inferring friendship network structure by using mobile phone data. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* 106 (36):15274 – 15278.

Eagle, N., and A. Sandy Pentland. 2006. Reality mining: sensing complex social systems. *Personal & Ubiquitous Computing* 10 (4):255–268.

Easley, D., and J. Kleinberg. 2010. *Networks, Crowds, and Markets : Reasoning About a Highly Connected World*. Cambridge University Press.

Emch, M., E. D. Root, S. Giebultowicz, M. Ali, C. Perez-Heydrich, and M. Yunus. 2012. Integration of Spatial and Social Network Analysis in Disease Transmission Studies. *Annals of the Association of American Geographers* 102 (5):1004–1015.

ESRI. 2013. GeoEvent Processor Extension. *ArcGIS for Server*. http://www.esri.com/software/arcgis/arcgisserver/extensions/geoevent-extension (last accessed 18 January 2014).

Facebook. 2013. Facebook Developers. *Facebook Developers*. https://developers.facebook.com/ (last accessed 25 December 2013).

Geofeedia. 2014. Search & Monitor Social Media by Location. *Geofeedia*. http://corp.geofeedia.com/ (last accessed 18 January 2014).

Gidofalvi, G., and T. B. Pedersen. 2005. Spatio-temporal Rule Mining: Issues and Techniques. In *Data Warehousing and Knowledge Discovery*, Data warehousing and knowledge discovery; DaWaK 2005., eds. J. M. Morvan and J. M. Morvan, 275–284. Berlin, [Great Britain], Springer.

Goodchild, M. F. 2010. Twenty years of progress: GIScience in 2010. *Journal of Spatial Information Science* (1):3.

Goodchild, M. F., M. Yuan, and T. J. Cova. 2007. Towards a general theory of geographic representation in GIS. *International Journal of Geographical Information Science* 21 (3):239–260.

Google. 2013. Google App Engine — Google Developers. *Google App Engine: Platform as a Service*. https://developers.google.com/appengine/?csw=1 (last accessed 24 January 2014).

Hagberg, A. A., D. A. Schult, and P. J. Swart. 2008. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, 11–15. Pasadena, CA USA.

H ägerstrand, T. 1970. What About People in Regional Science? *Papers in Regional Science* 24 (1):7.

Hanneman, R. A., and M. Riddle. 2005. *Introduction to social network methods*. University of California Riverside.

Hillier, B. 1999. The common language of space A way of looking at the social, economic and environmental functioning of cities on a common basis. *Journal of Environmental Sciences (IOS Press)* 11 (3):344–349.

Hipp, J. R., R. W. Faris, and A. Boessen. 2012. Measuring “neighborhood”: Constructing network neighborhoods. *Social Networks* 34 (1):128–140.

Horner, M., B. Zook, and J. Downs. 2012. Where were you? Development of a time-geographic approach for activity destination re-construction. *COMPUTERS ENVIRONMENT AND URBAN SYSTEMS* 36 (6):488–499.

Hornsby, K., and M. J. Egenhofer. 2000. Identity-based change: a foundation for spatio-temporal knowledge representation. *International Journal of Geographical Information Science* 14 (3):207–224.

Humphreys, L. 2007. Mobile Social Networks and Social Practice: A Case Study of Dodgeball. *Journal of Computer-Mediated Communication* 13 (1):341–360.

Jiang, B. 2009. Ranking Spaces for Predicting Human Movement in an Urban Environment. *International Journal of Geographical Information Science* 23 (7):823–837.

Jiang, B., and C. Claramunt. 2002. Integration of Space Syntax into GIS: New Perspectives for Urban Morphology. *Transactions in GIS* 6 (3):295–309.

Jiang, B., and C. Claramunt. 2004. A Structural Approach to the Model Generalization of an Urban Street Network*. *GeoInformatica* 8 (2):157–171.

Jiang, B., and T. Jia. 2011. Agent-based simulation of human movement shaped by the underlying street structure. *International Journal of Geographical Information Science* 25 (1):51–64.

Jiang, B., and Y. Miao. 2014. The Evolution of Natural Cities from the Perspective of Location-Based Social Media. *arXiv:1401.6756 [nlin, physics:physics]* :1–14.

Ji, S. Y., E. Niklas, and S. Lee. 2010. TimeMatrix: Analyzing Temporal Social Networks Using Interactive Matrix-Based Visualizations. *International Journal of Human-Computer Interaction* 26 (11/12):1031–1051.

Kim, H.-M., and M.-P. Kwan. 2003. Space-time accessibility measures: A geocomputational algorithm with a focus on the feasible opportunity set and possible activity duration. *Journal of Geographical Systems* 5 (1):71.

Kleinfeld, J. S. 2002. The Small World Problem. *Society* 39 (2):61–66.

Knox, E. G., and M. S. Bartlett. 1964. The Detection of Space-Time Interactions. *Applied Statistics* 13 (1):25.

Kuijpers, B., H. J. Miller, T. Neutens, and W. Othman. 2010. Anchor uncertainty and space-time prisms on road networks. *International Journal of Geographical Information Science* 24 (8):1223.

Kwan, M.-P. 1998. Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-based Framework. *Geographical Analysis* 30 (3):191–216.

———. 2002. Feminist Visualization: Re-Envisioning GIS as a Method in Feminist Geographic Research. *Annals of the Association of American Geographers* (4):645.

———. 2004. GIS Methods in Time-Geographic Research: Geocomputation and Geovisualization of Human Activity Patterns. *Geografiska Annaler Series B: Human Geography* 86 (4):267.

———. 2007. Mobile Communications, Social Networks, and Urban Travel: Hypertext as a New Metaphor for Conceptualizing Spatial Interaction. *Professional Geographer* 59 (4):434–446.

———. 2010. A Century of Method-Oriented Scholarship in the Annals. *Annals of the Association of American Geographers* 100 (5):1060–1075.

Lampos, V., and N. Cristianini. 2010. Tracking the flu pandemic by monitoring the social web. *2010 2nd International Workshop on Cognitive Information Processing (CIP)* :411.

Leng, S., L. Zhang, H. Fu, and J. Yang. 2007. Mobility analysis of mobile hosts with random walking in ad hoc networks. *Computer Networks* 51 (10):2514–2528.

Lewis, K., J. Kaufman, G. A. Marco, W. B. Andreas, and C. A. Nicholas. 2008. Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks* 30:330–342.

Liben-Nowell, D., and J. Kleinberg. 2007. The link-prediction problem for social networks. *Journal of the American Society for Information Science & Technology* 58 (7):1019–1031.

Licoppe, C., and Y. Inada. 2008. Geolocalized Technologies, Location-Aware Communities, and Personal Territories: The Mogi Case. *Journal of Urban Technology* 15 (3):5–24.

Li, J., and Q. Guo. 2003. Quantitative research of urban spatial morphology based on syntactic analysis. *Enginering Journal of Wuhan University* 36 (2):69–73.

Lindsey, G., J. Wilson, E. Rubchinskaya, J. Yang, and Y. Han. 2007. Estimating urban trail traffic: Methods for existing and proposed trails. *Landscape and Urban Planning* 81:299–315.

Liu, X., and B. Jiang. 2012. Defining and Generating Axial Lines from Street Center Lines for better Understanding of Urban Morphologies. *International Journal of Geographical Information Science* 26 (8):1521–1532.

Liu, Y., Z. Sui, C. Kang, and Y. Gao. 2014. Uncovering Patterns of Inter-Urban Trip and Spatial Interaction from Social Media Check-In Data. *PLoS ONE* 9 (1):1–11.

Li, Y., H. Gao, M. Yang, W. Guan, H. Ma, W. Qian, Z. Cao, and X. Yang. 2013. What are Chinese Talking about in Hot Weibos? *arXiv preprint arXiv:1304.4682*. http://arxiv.org/abs/1304.4682 (last accessed 26 December 2013).

Mennis, J. 2010. Multidimensional Map Algebra: Design and Implementation of a Spatio-Temporal GIS Processing Language. *Transactions in GIS* 14 (1):1–21.

Mennis, J. L. P., Donna J.Qian, Liujian. 2000. A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science* 14 (6):501.

Merriman, P. 2012. Human geography without time-space. *Transactions of the Institute of British Geographers* 37 (1):13–27.

Mika, P. 2005. Flink: Semantic Web technology for the extraction and analysis of social networks. *Web Semantics: Science, Services and Agents on the World Wide Web* 3 (2–3):211–223.

Milgram, S., L. Mann, S. Harter, and B. Kass. 1965. THE LOST-LETTER TECHNIQUE: A TOOL OF SOCIAL RESEARCH. *Public Opinion Quarterly* 29 (3):437.

Miller, H. J. 1999. Measuring space-time accessibility benefits within transportation networks: basic theory and computational procedures. *Geographical analysis* 31 (1):1–26.

———. 2003. What about people in geographic information science? *Computers, Environment and Urban Systems* 27:447.

Miller, H. J., and S. A. Bridwell. 2009. A Field-Based Theory for Time Geography. *Annals of the Association of American Geographers* 99 (1):49–75.

Miyagawa, M. 2009. Optimal hierarchical system of a grid road network. *Annals of Operations Research* 172 (1):349–361.

Naaman, M., J. Boase, and C.-H. Lai. 2010. Is it really about me?: message content in social awareness streams. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 189–192. ACM http://dl.acm.org/citation.cfm?id=1718953 (last accessed 11 December 2014).

Neutens, T., N. Weghe, F. Witlox, and P. Maeyer. 2008. A three-dimensional network-based space–time prism. *Journal of Geographical Systems* 10 (1):89–107.

Neutens, T., F. Witlox, N. Van De Weghe, and P. H. De Maeyer. 2007. Space-time opportunities for multiple agents: a constraint-based approach. *International Journal of Geographical Information Science* 21 (10):1061–1076.

Newman, M. E. 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary physics* 46 (5):323–351.

Newman, M. E. J. 2003. The Structure and Function of Complex Networks. *SIAM Review* (2):167.

———. 2006. Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences of the United States of America* (23):8577.

Ozbay, K., D. Ozmen, and J. Berechman. 2006. Modeling and Analysis of the Link between Accessibility and Employment Growth. *Journal of Transportation Engineering* 132 (5):385–393.

Padmanabhan, A., S. Wang, G. Cao, M. Hwang, Z. Zhang, Y. Gao, K. Soltani, and Y. Liu. 2014. FluMapper: A cyberGIS application for interactive analysis of massive location-based social media. *Concurrency and Computation: Practice and Experience* 26 (13):2253–2265.

Pearson, K. 1905. The Problem of the Random Walk. *Nature* 72 (1865):294.

Peuquet, D., and N. Duan. 1995. An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographical Information Systems* 9 (1):7–24.

Peuquet, D. J. 1994. It’s about Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems. *Annals of the Association of American Geographers* (3):441.

———. 2001. Making Space for Time: Issues in Space-Time Data Representation. *GeoInformatica* 5 (1):11.

Pickles, J. C. 1995. Representations in an Electronic Age: Geography, GIS, and Democracy. In *Critical Geographies: A Collection of Readings*, eds. H. Bauder and S. E.-D. Mauro, 637–663. British Columbia, Canada.: Praxis (e)Press.

Pons, P., and M. Latapy. 2005. Computing communities in large networks using random walks. In *20th International Symposium on Computer and Information Sciences,* 284–293. Istanbul, Turkey: Springer.

Porras, R., T. Takeshita, M. Ikezoe, and R. Araya. 2002. A Study on the Pedestrian Space applying Space Syntax and the Segment Unit. *Journal of Asian Architecture and Building Engineering* 1 (1):197–203.

Porta, S., V. Latora, F. Wang, S. Rueda, E. Strano, S. Scellato, A. Cardillo, E. Belli, F. Cardenas, B. Cormenzana, and L. Laura. 2012. Street Centrality and the Location of Economic Activities in Barcelona. *Urban Studies* 49 (7):1471–1488.

Pultar, E., T. J. Cova, Y. May, and M. F. Goodchild. 2010. EDGIS: a dynamic GIS based on space time points. *International Journal of Geographical Information Science* 24 (3):329–346.

Raubal, M., H. J. Miller, and S. Bridwell. 2004. User-Centred Time Geography for Location-Based Services. *Geografiska Annaler Series B: Human Geography* 86 (4):245–265.

Rey, S. J., and L. Anselin. 2007. PySAL: A Python Library of Spatial Analytical Methods. *The Review of Regional Studies* 37 (1):5–27.

Rodrigue, J.-P., C. Comtois, and B. Slack. 2013. *The Geography of Transport Systems* 3 edition. NewYork: Routledge.

Ronald, N., V. Dignum, C. Jonker, T. Arentze, and H. Timmermans. 2012. On the engineering of agent-based simulations of social activities with social networks. *Information and Software Technology* 54 (6):625–638.

Russell, M. A. 2013. *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. O’Reilly Media, Inc.

Scellato, S., A. Noulas, R. Lambiotte, and C. Mascolo. 2011. Socio-Spatial Properties of Online Location-Based Social Networks. *ICWSM* 11:329–336.

Schwanen, T., and M.-P. Kwan. 2012. Critical Space-Time Geographies: Guest Editorial. *Environment and Planning A* 44 (9):2043–2048.

Shaw, S.-L., and H. Yu. 2009. A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. *Journal of Transport Geography* 17 (2):141–149.

Shaw, S.-L., H. Yu, and L. S. Bombom. 2008. A Space-Time GIS Approach to Exploring Large Individual-based Spatiotemporal Datasets. *Transactions in GIS* 12 (4):425–441.

Short, M. B., M. R. D'Orsogna, V. B. Pasour, G. E. Tita, P. J. Brantingham, A. L. Bertozzi, and L. B. Chayes. 2008. A statistical model of criminal behavior. *Mathematical Models & Methods in Applied Sciences* 18:1249–1267.

De Smith, M. 2010. *Statistical Analysis Handbook: Concepts*, *Techniques*, *Tools*. www.statsref.com/HTML/.

Song, X., X. Wang, A. Li, and L. Zhang. 2011. Node Importance Evaluation Method for Highway Network of Urban Agglomeration. *Journal of Transportation Systems Engineering and Information Technology* 11 (2):84–90.

Sui, D., and M. Goodchild. 2011. The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science* 25 (11):1737.

Taaffe, E. J., H. L. Gauthier, and M. E. O'Kelly. 1996. *Geography of transportation*. Upper Saddle River, N.J.: Prentice Hall.

Takhteyev, Y., A. Gruzd, and B. Wellman. 2012. Geography of Twitter networks. *Social Networks* 34 (1):73–81.

Tavakoli, M., and S. Fakhraie. 2011. THEORY OF TIME GEOGRAPHY (EXPLANATION OF ITS APPLICABLE QUANTITIES IN PLANNING). *International Journal of Academic Research* 3 (3):655–662.

Thrift, N. 1996. *Spatial Formations*. SAGE Publications. http://books.google.com/books?id=B2hw8X-yigIC.

Thrift, N. J. 1996. *Spatial formations*. Sage.

Torrens, P., X. Li, and W. Griffin. 2011. Building Agent-Based Walking Models by Machine-Learning on Diverse Databases of Space-Time Trajectory Samples. *Transactions in GIS* 15:67.

Train, K. 2009. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Tsou, M.-H., and M. Leitner. 2013. Visualization of social media: seeing a mirage or a message? *Cartography and Geographic Information Science* 40 (2):55.

Turner, A. 2009. The Role of Angularity in Route Choice. In *Spatial Information Theory*, Lecture Notes in Computer Science., eds. K. S. Hornsby, C. Claramunt, M. Denis, and G. Ligozat, 489–504. Springer Berlin Heidelberg.

Twitter. 2013. Exploring the Twitter API. *Twitter Developers*. https://dev.twitter.com/console (last accessed 25 December 2013).

Vance, C., and R. Hedel. 2008. On the Link Between Urban Form and Automobile Use: Evidence from German Survey Data. *Land Economics* 84 (1):51–65.

Vasardani, M., S. Winter, and K.-F. Richter. 2013. Locating place names from place descriptions. *International Journal of Geographical Information Science* 27 (12):2509–2532.

Volchenkov, D., and P. Blanchard. 2007. Random walks along the streets and canals in compact cities: Spectral analysis, dynamical modularity, information, and statistical mechanics. *Physical Review E* 75 (2):026104.

Wang, D., and T. Cheng. 2001. A spatio-temporal data model for activity-based transport demand modelling. *International Journal of Geographical Information Science* 15 (6):561–585.

Wang, F., A. Antipova, and S. Porta. 2011. Street centrality and land use intensity in Baton Rouge, Louisiana. *Journal of Transport Geography* 19 (2):285–293.

Wang, J., X. Wu, Y. Bo, and J. Guo. 2011. Improved Method of Node Importance Evaluation Based on Node Contraction in Complex Networks. *Procedia Engineering* 15:1600–1604.

Wang, P., T. Hunter, A. M. Bayen, K. Schechtner, and M. C. Gonz ález. 2012. Understanding road usage patterns in urban areas. *Scientific Reports* 2 (1001):1–6.

Watling, D., D. Milne, and S. Clark. 2012. Network impacts of a road capacity reduction: Empirical analysis and model predictions. *Transportation Research Part A: Policy and Practice* 46 (1):167–189.

Wewal, J., D. Wilkie, P. Merrell, and M. C. Lin. 2010. Continuum Traffic Simulation. *Computer Graphics Forum* 29 (2):439–448.

Winter, S., and Z.-C. Yin. 2011. The elements of probabilistic time geography. *GeoInformatica* 15 (3):417–434.

Worboys, M. 2005. Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science* 19 (1):1–28.

Wright, D. J., M. F. Goodchild, and J. D. Proctor. 1997. GIS: Tool or Science? Demystifying the Persistent Ambiguity of GIS as “Tool” Versus “Science.” *Annals of the Association of American Geographers* (2):346.

Xing, W., and A. Ghorbani. 2004. Weighted PageRank algorithm. In *Communication Networks and Services Research*, 305–314. IEEE.

Yang, S.-J. 2005. Exploring complex networks by walking on them. *Physical Review E* 71 (1):016107.

Yao, X. 2010. Modeling Cities as Spatio-Temporal Places. In *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, 311–328. Netherlands: Springer http://proxy-remote.galib.uga.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=76899530&site=eds-live.

Yuan, L., S. Chen, Y. Wang, Z. Yu, W. Luo, and G. L ü. 2010. CAUSTA: Clifford Algebra-based Unified Spatio-Temporal Analysis. *Transactions in GIS* 14:59–83.

Yuan, M. 1997. Use of knowledge acquisition to build wildfire representation in Geographical Information Systems. *International Journal of Geographical Information Science* 11 (8):723–746.

Yuan, M., A. Nara, and J. Bothwell. 2014. Space–time representation and analytics. *Annals of GIS* 20 (1):1–9.

Yu, H. 2006. Spatio-temporal GIS Design for Exploring Interactions of Human Activities. *Cartography and Geographic Information Science* 33 (1):3–19.

Yu, H. 2010. China's House Price: Affected by Economic Fundamentals or Real Estate Policy? *Frontiers of Economics in China* 5 (1):25–51.

Yu, H., and S.-L. Shaw. 2008. Exploring potential human activities in physical and virtual spaces: a spatio-temporal GIS approach. *International Journal of Geographical Information Science* 22 (4):409–430.

Zhang, H., and Z. Li. 2011. Weighted ego network for forming hierarchical structure of road networks. *International Journal of Geographical Information Science* 25 (2):255–272.

Zhang, S., and X. Yao. 2011. *Social-spatial structure of Beijing : a spatial-temporal analysis*. 2011.

CHAPTER 2 REPRESENTING LOCATION-BASED SOCIAL MEDIA ACTIVITY IN GIS[1]

[1] Xuebin Wei and Xiaobai Yao. Submitted to International Journal of Geographical Information Science, 07/15/ 2015.

Abstract

This research develops a conceptual framework and logical models for the representation and analysis of location-based social media activity (LBSMA) of human beings in GIS. With increasing popularity of location-based social networking, social media, such as Twitter and Facebook, have become new channels to observe human activities in physical and virtual world. At the same time, there is a shift of some human interactions from the physical space to the social space. Traditional geographical representation in GIS is not sufficient to handle the increased sophistication of human activities related to, or embedded in, location-based social media data. This research proposes a conceptual framework of location-based social media activity to model human activities in spatial-temporal-social dimension in GIS. This research designs a conceptual model for the representation of LBSMA data in a GIS environment and implemented a pilot computer system with application examples of analyzing human activities in spatial-temporal-social dimensions. The study develops strategies to collect real-world LBSMA data from Facebook and Twitter. A case study with the collected data tests the developed prototype. It is demonstrated that the proposed data model enables us to answer the types of questions that could not be answered in traditional GIS.

Introduction

Understanding human behavior through human activities has been an important geographic inquiry in the literature. Researchers have studied human behavior from various perspectives. For instance, behavior geography concerns the cognitive process of human behavior and draws on other fields as well because human activities are generated due to physiological, psychological and economic needs (Ronald et al. 2012). Another closely related thread of research examines human activities through visualization, analysis, and modeling of human dynamics in a GIS environment. Our research attempts to contribute to the latter, motivated by two fundamental issues. First, the growing popularity of network-based social media and the availability of such data provide us an unprecedented opportunity to study human activities in new lights. The new types of phenomena and new types of data require new conceptualization, new methodologies, and new tools to make the best out of them. Secondly, it has been well recognized that social connections play an important role in human behavior. However, social network has been ignored or oversimplified in current representations of human activities in current off-the-shelf GIS programs. Therefore, this research aims to develop a GIS conceptual framework and associated logical model to represent space, time, and social connections from location-based social media activity (LBSMA) data in GIS.

By the nature of the research issue, studying human activities typically requires data at individual level, or so called disaggregated data, with fine spatial and temporal granularities. However, commonly available data are usually aggregated, such as those from the census. Thus obtaining suitable data is difficult. The availability of location-based social media data provides an unprecedented opportunities for this type of research, as such data are inherently entered on individual basis and are of high granularities in space and time. In the era of big data, enormous attention has been attracted to extracting data from Internet-based platforms to study human societies. A message from social media is considered an extension of human mind (Tsou and Leitner 2013). Furthermore, location-based social media data provide not only the space and time information of activities, but also the social connections among individuals. This is particularly advantageous to research of human activities, as the context of social connection is considered important to human activities. It has been argued that time, space and social differentiation should be coupled in the study of practices or phenomena (Schwanen and Kwan

2012). From the relationalism-idealism perspective, the assumed existence of social networks sets the scope to which space and time should be conceptualized and analyzed in human activity analysis (Yuan, Nara, and Bothwell 2014). In the age of big data, details of human activities can now be extracted from the social media to reveal when and where people interact with others, collections of such interactions reveal social network among people. Different types of social media allow for different types of connections. For example, Twitter fosters an asymmetric network structure that people prefer to broadcast individual activities, while LinkedIn and Facebook aim to capture pre-existing ties by focusing on social interactions among friends (Takhteyev, Gruzd, and Wellman 2012). Previous studies have investigated the content and friendship structure on Twitter (Takhteyev, Gruzd, and Wellman 2012; Naaman, Boase, and Lai 2010), Facebook (Lewis et al. 2008) and Weibo (Li et al. 2013). The spatial distribution of location-based social activities from different social media has also been explored in recent studies. For example, Cho et al (Cho, Myers, and Leskovec 2011) devised a periodic & social mobility model to study the interaction of geographic, temporal and social aspects of human mobility. Eagle et al (2009) compared the resulting behavioral social networks with self-reported relationships based on mobile phones data and accurately infer friendships based on the observational data alone. Cheng et al (2011) analyzed human mobility patterns from Twitter in terms of spatial, temporal, social and textual aspects.

There is a long tradition that human activities are visually represented and analyzed, particularly in GIS. Starting from the space-time prism (Hägerstrand 1970), trajectories of human activities are visually represented as a series of locations in space-time dimensions. Because human activities have innate spatial component, geographic information system (GIS) is naturally the most desirable environment for the visualization and analysis of it. Sui and

Goodchild (2011) suggest that GIS is a media in terms of communicating and sharing knowledge and supporting location-based social networking. With increasing popularity of location-based social networking technologies and data, scientific investigations have expanded to include data about activities in both physical and virtual spaces. Meanwhile, the convergence of geographic information systems (GIS) and social media has resulted in a data avalanche that creates new challenges in GIScience (Sui and Goodchild 2011). Although location-based social media activities have been examined in many studies as discussed previously, a structured GIS representation is still absent for all dimensions of space, time, and the context of social connections in which activities take place. GIS representation of space and time alone is already a critical research theme in the literature, adding more dimensions obviously is not a trivial issue. The goal of this paper is to fill the gap by developing a conceptual model and a computer prototype to organize location-based social media activity (LBSMA) data in GIS, so that the positions and interrelationship of LBSM activities in the spatial, temporal, and social dimensions can be represented and analyzed further. In this research, as highlighted in Figure 2-1, the LBSMA data refer to the subset of human activities of which locations can be georeferenced in the geographical space and contents are advertised in the networked social media such as Facebook, Twitter and others. The scope of the study is limited to human activities that are recorded explicitly or implicitly in the LBSMA data, the yellow-highlighted area in Figure 2-1.

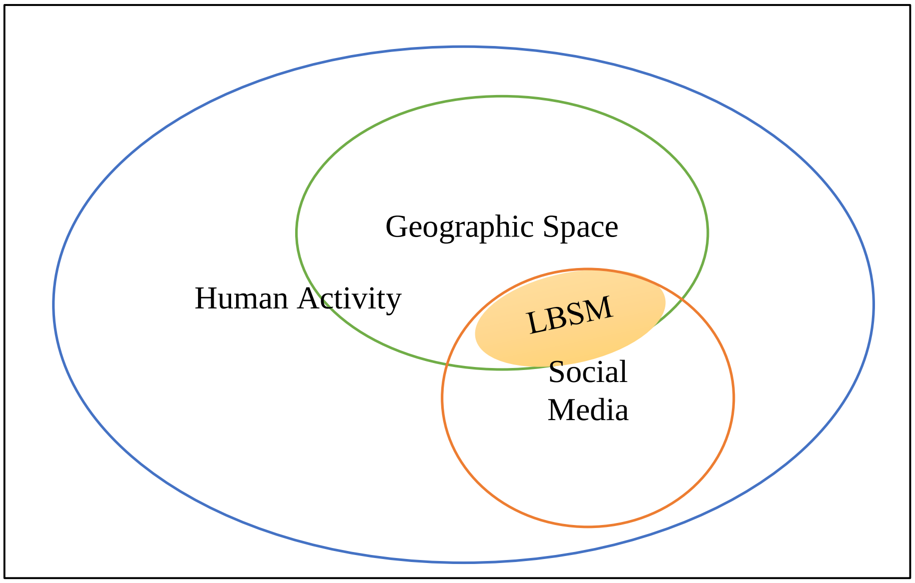


Figure 2-1 Definition of Location-based Social Media Activity

The paper is organized as follows. The next section reviews related prior work about space-time representation in GIS and those about visual models of human activities in GIS environments. Section 3 presents an ontological framework of the concepts and a conceptual model for the representation of LBSMA data in GIS. Section 4 introduces a corresponding logical model and the pilot implementation of a computer system that can import, organize and analyze LBSMA data. A case study is conducted in the prototype to illustrate the feasibility of the whole process and the capability of enabling new types of geographic inquiries in such a system. The paper is concluded with discussions of major findings and future research in Section 5.

Related Work

This research is related to two threads of research about data modeling in GIS. The first thread is about the conceptual and logical data models for space-time representation in GIS, while the other thread is the representation of human activities in GIS.

Space-time Data Modeling in GIS

Time consciousness has been discussed in many discourses of geographical inquiry (Thrift 1996). Depending on the philosophical considerations of space-time representation, space and time have been conceptualized in either object view in which space is composed of points while time consists of instants, or subject view in which space and time are positional qualities attached to objects (Peuquet 1994, 2001; Yuan, Nara, and Bothwell 2014). In the early stage of GIS, time was simply considered an abstract parameter that is devoid of any differentiating contents or sequences. Gradually, time became both experientially and spatially referenced. Excellent examples include those analyses extended from Hägerstrand’s space-time (ST) prism. More recently, many researchers believe that geographic data models must serve as an acceptable reflection of the real-world phenomena (Pickles 1995) and consequently many ST-explicit models have been proposed. In the literature, space-time data models in GIS can be classified into three categories: spatially focused data models, temporally focused data models, and integral data models. The framework of spatially focused data models treats time as an additional dimension to the traditional location-based spatial data. Examples include the snapshot model (Dragicevic and Marceau 2000), spatio-temporal raster data structure (Mennis 2000), or the traditional feature-based spatial data such as amendment vectors and spatial-temporal place model (Yao 2010). The framework of temporally focused data models organizes spatial objects according to a time line of spatial object that changes (Peuquet and Duan 1995), evolves (Hornsby and Egenhofer 2000), or involves more complex occurrences and relationships (Worboys 2005). The family of integral data models combine the location-based or the feature-based spatial data structure with the temporal coordination in attribute sets. Those integral data frameworks include the spatiotemporal triad framework (Peuquet 1994), the three domain model

(Yuan 1997), the geo-atom (Goodchild, Yuan, and Cova 2007; Pultar et al. 2010), and the unified spatial-temporal data model (Yuan et al. 2010).

Representation of Human Activities in GIS

The literature has a wide range of research topics about human activities, such as human travel behavior, way-finding, migration and residential mobility, decision making and choice behavior, as well as spatial cognition and environmental perception (Kwan 2010). This study focuses on the representation of human movement dynamics in GIS. Up to this point, the most widely used are the space-time representations.

Hägerstrand (1970) introduced the time dimension to the traditionally space-only approach to modeling human activities. An individual is located at a specific location in space and time, while the individual’s next location in space at in a given time duration is constrained by several factors. Hägerstrand addressed three types of constraints, i.e., capability constraints, coupling constraints, and authority constraints (Hägerstrand 1970). Almost simultaneously, Anderson proposed a similar idea using the concept of time-budget diary which incorporates timing, duration and location when modeling individual's activities (Anderson 1970). Hence Kuijpers et al. (2010) argued that an individual movement implies a trade-off between inseparability and scarce nature of space and time, and is conditioned by various constraints and opportunities. These ideas became known as time geography. Time geography offers a people-oriented extension to place-based tools in GIS (Miller 2003) that indicates individual travel possibilities (Neutens et al. 2008).

Rooted in time geography, various types of spatio-temporal constructs have been developed in GIS environments to represent human activities. Space-time path, space-time prism, and potential path are well-known examples. A space-time path traces an individual's movements in space with respect to time, as illustrated in Figure 2-2. A space-time prism delimits possible locations for the space-time path during a period of potential activity participation. Potential path areas are the projection of a space-time prism to the geographical plane (Neutens et al. 2008). Those three ST constructs can be used directly to measure individual accessibility to resources (Miller and Bridwell 2009).

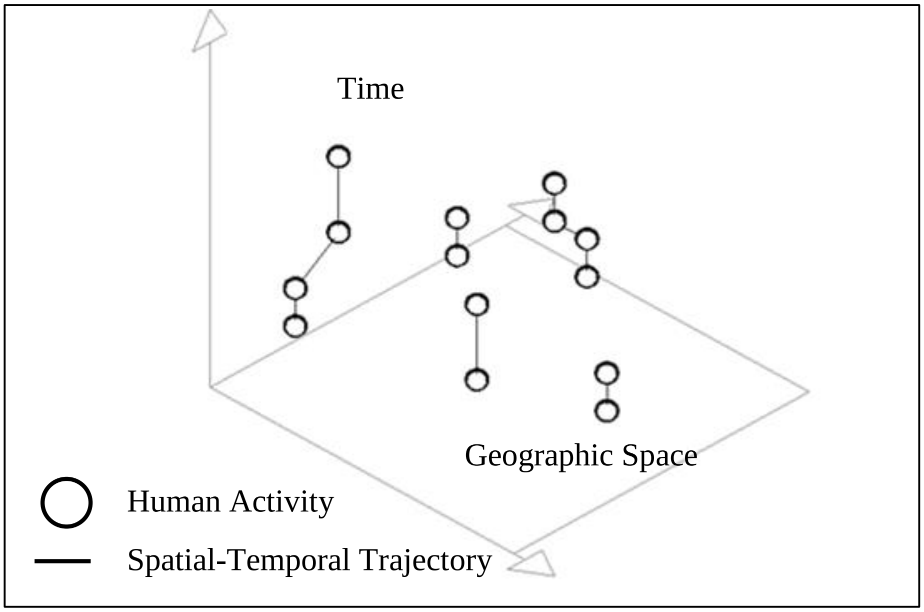


Figure 2-2 Space-Time Path Model

The space-time path model is the most widely adopted approach in the representation of human activities in GIS. For instance, it has been applied in the analysis of individual accessibility and geo-visualization of human activities (Miller 1999; Kwan 1998; Kim and Kwan 2003). More recently, to account for human activities by information communication technology (ICT), Yu proposed a revised design of spatial-temporal GIS for the analysis of human activities in both physical and virtual spaces by using linear referencing and dynamic segmentation (Yu 2006). Shaw, Yu and Bombom (2008) later presented a generalized space-time path for visual exploration of spatiotemporal changes of individual activities. Chen et al (2011) developed

ArcGIS Extension named Activity Pattern Analyst that is able to generate, filter, and query space-time path and to perform spatial-temporal density and path clustering analysis. Yin et al (2011) introduced an extended time-geographic analytical framework to illustrate how human interactions opportunities can be impacted under the use of phone communication. Yin and Shaw (2015) proposed a spatiotemporal exploratory analysis approach to examine the relationship between physical separation and social interactions at the individual level.

Some criticisms towards the time-space path model are discussed in the literature. It was argued that the model use arbitrary spatial and temporal resolutions (Tavakoli and Fakhraie 2011). In addition, innovative communication technologies, such as location-based social networking, relaxes some of the space-time constraints and creates new topologies of spatial interaction of human activities (Kwan 2007). A series of improvements of the time-space path model have been addressed to tackle various issues (Wang and Cheng 2001; Kwan 2004; Raubal, Miller, and Bridwell 2004; Neutens et al. 2007; Yu and Shaw 2008; Shaw and Yu 2009; Miller and Bridwell 2009; Kuijpers et al. 2010). However, most human activity models in GIS do not incorporate social connections, with few exceptions such as Kwan's hypertext metaphor that considers social network in modeling human spatial interaction (Kwan 2007),

LBSMA Representation in a GIS environment

Ontological Framework

Equipped with geo-referencing capabilities, social media platforms nowadays are not only where people interact with others in the virtual space, but also data resources from which human physical activities can be recorded and visualized in geographic space. Thus it provides an excellent opportunity to connect the social dimension with the space-time dimensions.

Traditional GIS conceptual models use either object-based or field-based representation. The former distinguishes each spatial object with delineated spatial boundaries, while the latter enumerates all spatial locations systematically and stores attribute values for each location. However, none of them is able to directly account for social network (or social associations) or human activities in the context of such a network. Aiming to have a conceptual underpinning for later technical deployment to fill the gap, this paper first develops an ontological framework that categorizes related concepts and their relationships in the scope of the study. The ontological framework is graphically illustrated in Figure 2-3.

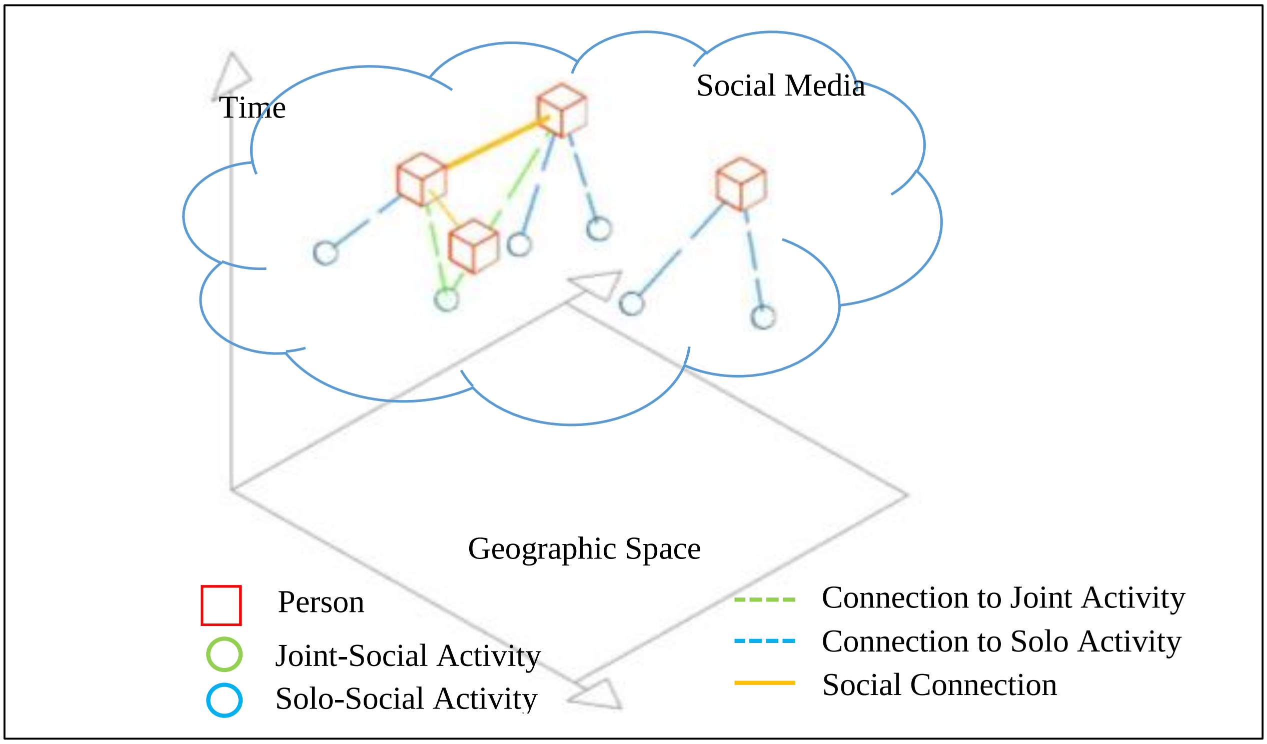


Figure 2-3 Human Activities in the Geographic Space and Social Media Space

The framework identifies four primary categories for the topic under study. They are Person, Activity, Location, and Social Connection.

- Person

Person is represented with red boxes in Figure 2-3. In the LBSMA ontological framework, Person refers to an individual who has one or more active accounts in social media platforms. A person may have different identifications in both the physical world and virtual world of social media. A person can participate in any number of activities and form social connections with other persons.

- Activity

Activity in this framework is defined as any action of a person in either the physical space or the virtual space. For example, visiting, commuting, reading, participating in a party are examples of activities in the physical world, while posting a message and following another account on Facebook or Twitter are examples of activities in the virtual space. Human activities can be further classified based on the number of participants. Because the scope of this study is the intersection area of human activities in the physical space and the social media space. It means no matter which space the activity took place, it would have been reflected in one or more social media platforms and thus is made visible to some or all people. Thus we call it social activity. Specifically, a joint-social activity (green circle in Figure 2-3) is an event participated by multiple persons (red box in Figure 2-3) while a solo-social activity (blue circle in Figure 2-3) is an action performed by a single person.

- Location

A person may conduct a series of activities that occur either in the physical geographic space or in the virtual social media space, or in both spaces. Therefore two types of locations are distinguished in the framework: the geographic location and the virtual location. For example, just like an address can refer to the geographic location of a person’s home, a uniform resource locator (URL) of a user’s profile page refers to the person’s virtual location on social media from which the user’s information, such as online activities, can be stored and retrieved. A restaurant can be represented as a point with a specific set of latitude and longitude in GIS, and its menu or reviews can be retrieved from its public pages on the website. The category of virtual locations are important in the framework because they not only facilitate the organization of human activities in the virtual world, but also provide the source of rich information about people, activities and the context environment.

- Social Connection

In the scope of this research, social connections are personal relationships expressed via social interactions. It is a mind-dependent construct that can be reflected by mind-independent human activities. Social connections can be explicitly expressed or identified through their self-reported relationships such as kinship, workplace connection, friendship, and so on, which can be explicitly indicated in the profiles or connections between profiles on some social media platforms. However, many additional social connections can be identified through spatial-temporal reasoning implicitly. Social connections can be assessed based on the characteristics of activities in spatial, virtual and temporal dimensions. For instance, two persons may have or develop social connections if they participate in joint-activities that occur in the same temporal-spatial space or at the same temporal-virtual locations. Frequent joint-activities at the same home address suggests close family ties (as shown with thick yellow line in Figure 2-3). Discussing an identical football team or related topics on social media suggests common interests in sports. Locations also cater various social connections among visitors depending on the characteristics of the locations. For example, home or work places can be considered spatial locations where people foster personal or professional relationships respectively, while public space or third space (generic designation of public places that host the regular gatherings of individuals beyond the realms of home and place) are neutral ground where patrons feel like they are on a level field with one another (Humphreys 2007). Implicit social connections can thus be identified via people-based or location-based approaches. People-based social connections construct an individual social network that reflects personal social capitals. Location-based social connections reveal how the location is experienced by different persons in the physical and the virtual worlds.

Conceptual Model

A conceptual data model is designed in the paradigm of object-oriented modeling. The model is illustrated in Figure 2-4. The conceptual model consists of five classes: Location, Person, Activity, Social Connection, and Common Interest.

The Location class contains uniquely identifiable referents in the geographic space or the social media space, each referent is an instance. Because two types of spaces are of concern, there are also two types of locations accordingly. Therefore two subclasses are inherited form the superclass of Location, namely Spatial Location and Virtual Location. Theoretically, locations in the geographic space can be referents in any traditional GIS georeferencing frame of reference. Examples include coordinates, street addresses, city names, etc. As current location-enabled digital devices are typically able to provide precise coordinates of locations, we use coordinates in the model. But it should be noted that other types of referents are by no means excluded. Locations in the social media virtual space can be either the uniform resource locator (URL) or identification number (ID) on respective social media platforms.

The Person class refer to all individuals who have locations in the social media virtual space. Attributes in the class include individual biographic characteristics and virtual locations instantiated from the virtual location subclass. Each person is an instance of the class. A person's activities can be retrieved via his/her virtual location on social media.

The Activity class is the collection of activity objects. An activity is an action in the social media space performed by one or more participants who are instances of the Person class. Based on the number of participants, two subclasses are defined, namely solo activity which involves a single participant and joint activity which includes multiple participants. Both subclasses are inherited from the Activity superclass. The attributes of Activity class include at least time, spatial location, virtual location, and the message of this activity. The spatial or virtual location attributes are instantiated from the spatial or virtual location subclass separately. The message of an activity, which is the related description posted on the social media platform, provides implicit literal information about people, location, and the activity itself. Semantic analysis of the messages will help understand human behaviors in the real word and the virtual world.

The Common Interest class is designed to represent shared interest that anchor points of interest. For instance, if many persons visit a restaurant and mention it in their posts, the place of the restaurant is a common interest. If the topic of health reform or a recent movie appear many social media activities, the discussed topic could be a common interest. Thus two types Common Interest are differentiated, each as a subclass of Common Interest. Each instance of the Place subclass indicates a location that is frequently visited in the real world. Place class contains a attribute information of the associated spatial location and/or virtual location. Topic class also has an attribute of virtual location because some topics have permanent locations on social media. Such attribute information will help researchers to develop methods of connecting the social network and locational information. Two global methods are developed to populate the Place and Topic subclass. One method extracts hot places from activity class. The other method extracts hot topics that are broadly discussed in the virtual social media space.

Social Connection class store the identified relationship between any pair of persons. Social connections between people can be revealed during different social activities. Therefore, joint activity subclass has a method to identify social connections among persons. In facebook, social connection identified between persons who are tagged in the same joint activity, appear in the same geographic place (tagged on photo) or mentioned in the virtual location (tagged in status). All instances of joint activity class can be used to identify and populate the Social Connection class.

The purpose of the proposed conceptual model is to organize the data in a reasonable and retrievable way, so as to maximize the possibility to study hidden relationships and patterns in such rich big data. The most important aspect is to allow information in the spatio-temporal-social dimensions, expressed either explicitly or implicitly, to be identified and represented in the system. The ultimate goal is to facilitate further modeling and analysis of such data. Social connections are normally modeled in a network structure in which people are represented as nodes and their acquaintance with others are represented as links. All identified social connections linked together result in the social network which could evolve over time with continuously updated social media data. Traditional network analysis measures, such as those for connectivity, centrality, and vulnerability, can be directly utilized for the analysis of the social network. However, unique properties of the data provide opportunities for researchers to mine additional hidden patterns. The underlying social media data for developing social connections are usually time-stamped and many of them are location-stamped. When organizing the data in spatial, temporal, and social dimensions by the proposed conceptual model, new methods can be further developed to analyze multi-dimensional patterns. For instance, by extracting the social network based on specific location(S), a location-based social network can be constructed and the spatial-social patterns can be explored. By checking the temporal signatures of social connections, one can explore the socio-temporal patterns and relationship in such data.

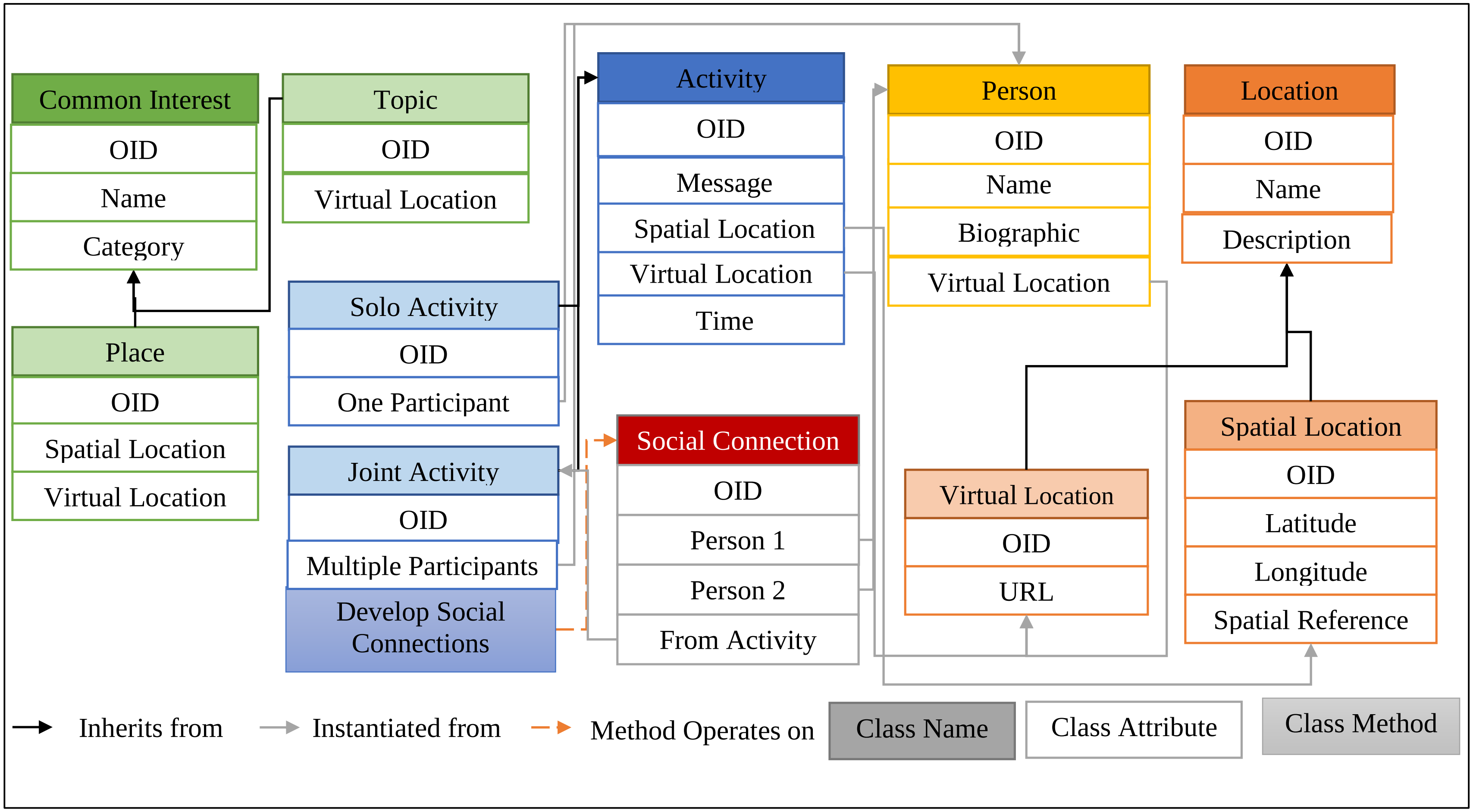


Figure 2-4 Conceptual Model of LBSMA

Pilot Prototype Implementation and Case Study

Based on the proposed conceptual model, a pilot prototype has been implemented and tested. A case study is performed to validate the prototype and most importantly to evaluate the usefulness of the proposed framework of LBSMA representation in GIS. Two most popular social media platforms, Facebook and Twitter, are used in the prototype and case study. But the model itself is equally applicable to other platforms. We started with collecting and tracking LBSMA data from Facebook and Twitter, then used the prototype to construct a LBSMA database from the collected social media data, and finally showed a few analysis examples that are otherwise not possible to be done in the traditional GIS environment.

LBSMA Data Collection

Although extracting twitter data is most commonly seen in current research projects due to Twitter's open API, The privacy and access controls make Facebook data much more closed (Russell 2013). This case study purposely chooses to extract Facebook data for two reasons. First, we want to study more about this more closed, and thus less explored social media platform. Second, Facebook is a social-networking type of platform while Twitter is a microblogging platform. Thus we will have better chance to observe and model social connections with Facebook data. The collection of users' location-based Facebook data requires a legitimate and secure procedure. A dedicated website for the collection of users' personal Facebook data has been established at www.lbsocial.net. This website obtained the Institutional Review Board (IRB) and Facebook App approval, and has gathered participants' information through the Facebook Application Programming Interface (API) with explicit authorizations of Facebook users. The website is running on the Google Application Engine. Any Facebook users can visit the designated website, read the IRB consent form, and log in with their Facebook account if they agree with the information provided in the consent form. Those who decide to log in become the volunteers for our data collection. The website will automatically record the volunteers’ and their friends’ posts that are embed with geo-location information. Those personal Facebook contents will be kept in Google App engine Database temperately, and then imported in a local GIS database.

Different from Twitter data which records latitudes and longitudes of users’ tweet ubiquitously, Facebook organizes user’s physical locations by using Open Graph protocol, a mechanism that enables any web page as an object in Facebook's Social Graph by injecting RDFa metadata into the page. The metadata uses a URL to represent any web page, i.e., a person, company, product, in a machine-readable way (Russell 2013). Therefore, several physical locations are grouped into a unique *place*, an Open Graph Object that has ID, name, coordination, visiting history and descriptions that are publicly available. Although the spatial resolution of the users’ location from Facebook is lower, the Open Graph protocol reduces the data uncertainty and equips the physical places with both physical and virtual locations that make the online resources of places available. Based on the visited places from the Facebook users, a places table is established and corresponding description are abstained from Facebook directly without users’ authorization. The data collection process is depicted in Figure 2-5.

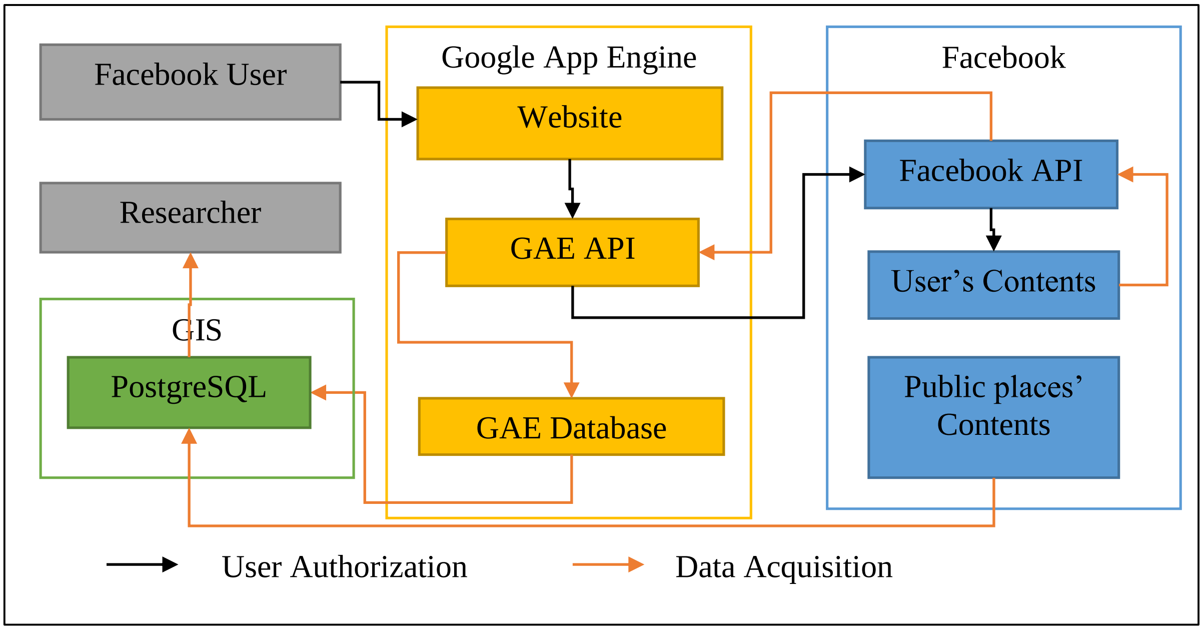


Figure 2-5 LBSMA Data Collection

Data Organization in GIS

The collected LBSMA data is then organized and maintained in a PostgreSQL database with the PostGIS plugin to provide GIS functions. Presented in Figure 6, this physical model has only implemented partial functions of the conceptual model. The collected posts from personal Facebook accounts are kept as records in the Activity table. A Facebook post can be published by the user of by the user's friend. The name and Facebook account ID for both the user and the user's friend are recorded in the Activity table. The other people that are tagged in this post are also kept in the Participants field. Since the number of the tagged people is not predictable, this data filed utilize the Json format to record all the participants. If the posts are embedded with geo-locations, the place name, place ID in Facebook Social Graph, and the latitude and longitude are kept in the Activity table. The time field indicates the time of publishing the post on Facebook, and the type field distinguishes whether the post is a check-in, a photo or a status.

The People table keeps unique users from the Activity table. Based on the participants of each activities, a people-based social network is constructed in which all the participants, (including user and from user if not tagged in post) in the same activity are assumed friends to each other, i.e., form connection in the social network. From the people-based social networks, a PeopleSocial table is created to keep some measurements of the social network for each people, including number of nodes, number of cliques, network density and etc., by using Networkx python library (Hagberg, Schult, and Swart 2008). The number of activities and total travel distance are also calculated for each people in the PeopleSocial table. Because people may post several photos or status for the same activity on Facebook, the activities with the identical participants on the same day at the same locations are treated as a single visit. The number of visits for each people is also computed in the PeopleSocial table.

The Place table contains all the unique places that are identified in the Activity table. Based on the Place ID in Facebook Social Graph, the additional information of these places are retrieved from Facebook, such as the category of the place, the total number of likes, the total number of talking about, and the total number of activities in this place. Some places also publish their website on Facebook. Similar to PeopleSocial table, place-based social networks are constructed from the participants in the same activity visiting the same place. Different to the people-based social network, the place-based social network reveal the social structure of the visitors to the spatial locations. Therefore, different sub-social groups are formed for place-based social networks. Some measurements of the place-based social network for each place are also calculated in the PlaceSocial table.

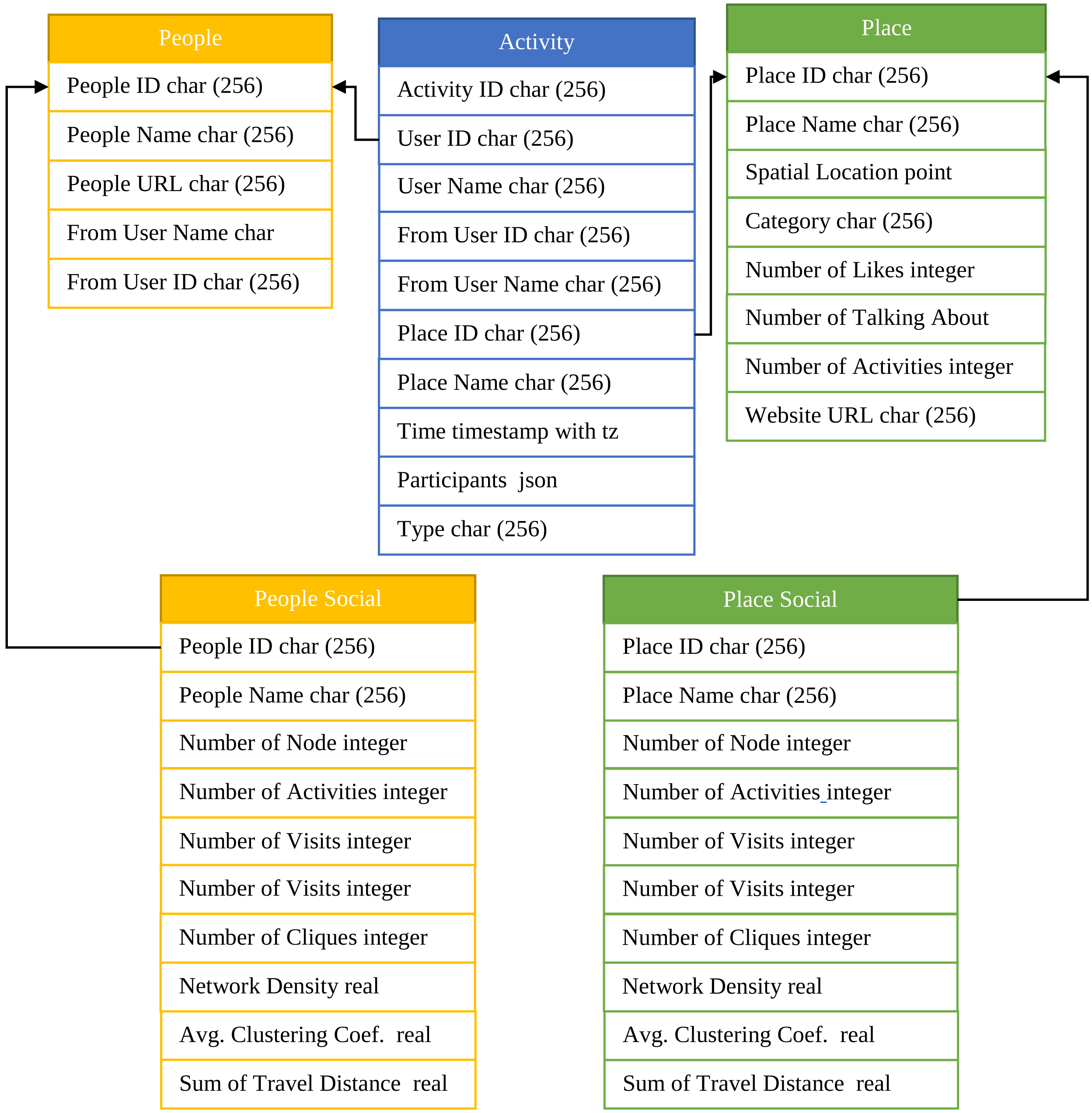


Figure 2-6 Physical Model of LBSMA

Visualization and Analysis Tools

A set of visualization and analysis tools for the LBSMA are developed in ArcGIS, including visualize activities and places, query people-based social network, create location-based social network and identify spatial-temporal interactions of activities.

- Visualize Activities and Places in GIS

This tool will read the activity table and place table from PostgreSQL in ArcGIS, and display the places and activities on a 2D map. In addition, since the activity records have the time coordination, the activities can also be visualized on a 3D map (Figure 2-7) in which the z coordination represents the time difference between the days of the publishing and a designated date.

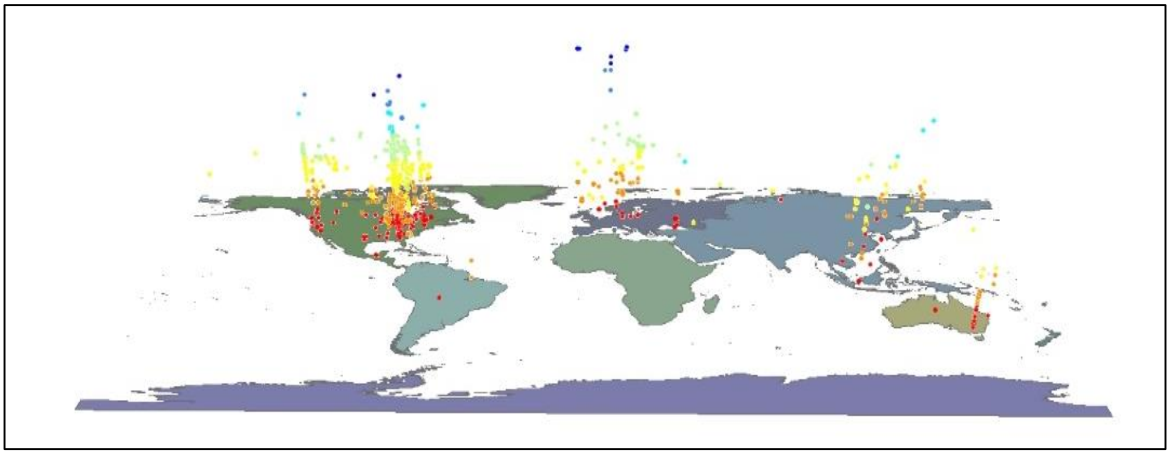
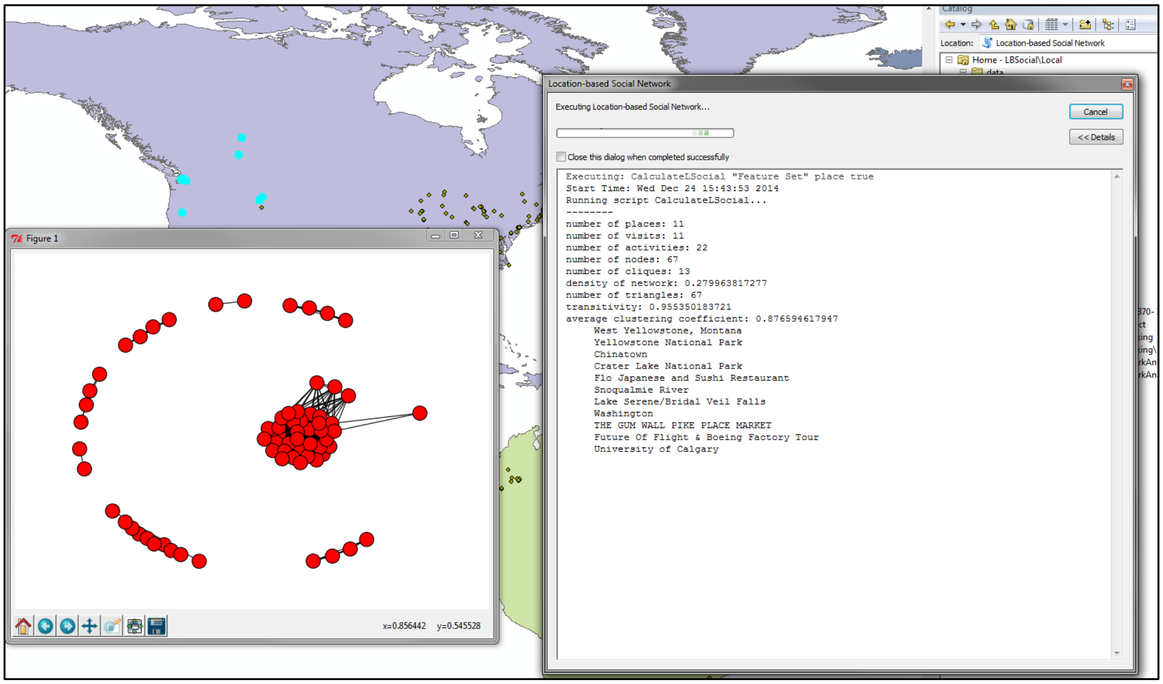


Figure 2-7 Visualization of Activities in 3 Dimension

- Create a Location-Based Social Network

The Create location-based social network tool allow users to interactively select the places in ArcGIS, and create a social network of the visitors from those places. The participants in the same activity are connected to each other in the social network. In addition to visualize the location-based social network, some network measures, e.g., number of nodes, number of cliques, average clustering coefficient and etc. are also reported in the result (Figure 2-8). If the user select only one single place, the reported result is identical to the record of this place in the PlaceSocial table.



Executing Location-based Social Network...

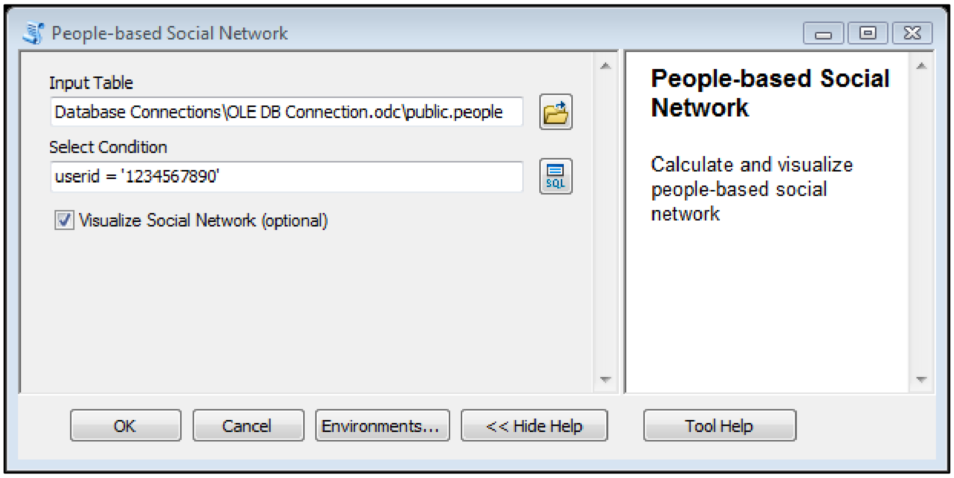
☐ Close this dialog when completed successfully

```
Executing: CalculateLSocial "Feature Set" place true
Start Time: Wed Dec 24 15:43:53 2014
Running script CalculateLSocial...
--------
number of places: 11
number of visits: 11
number of activities: 22
number of nodes: 67
number of cliques: 13
density of network: 0.279963817277
number of triangles: 67
transitivity: 0.955350183721
average clustering coefficient: 0.876594617947
     West Yellowstone, Montana
     Yellowstone National Park
     Chinatown
     Crater Lake National Park
     Flo Japanese and Sushi Restaurant
     Snoqualmie River
     Lake Serene/Bridal Veil Falls
     Washington
     THE GUM WALL PIKE PLACE MARKET
     Future Of Flight & Boeing Factory Tour
     University of Calgary
```

Figure 2-8 Result of location-Based Social Network Analysis

- Query People-Based Social Network

The third tool (Figure 2-9) allows users to query the people-based social network from the PeopleSocial table based on a user-defined SQL sentence. The user can also visualize and analyze the social networks for the selected people.



Input Table

Database Connections\OLE DB Connection.odc\public.people

Select Condition

userid = '1234567890'

☑ Visualize Social Network (optional)

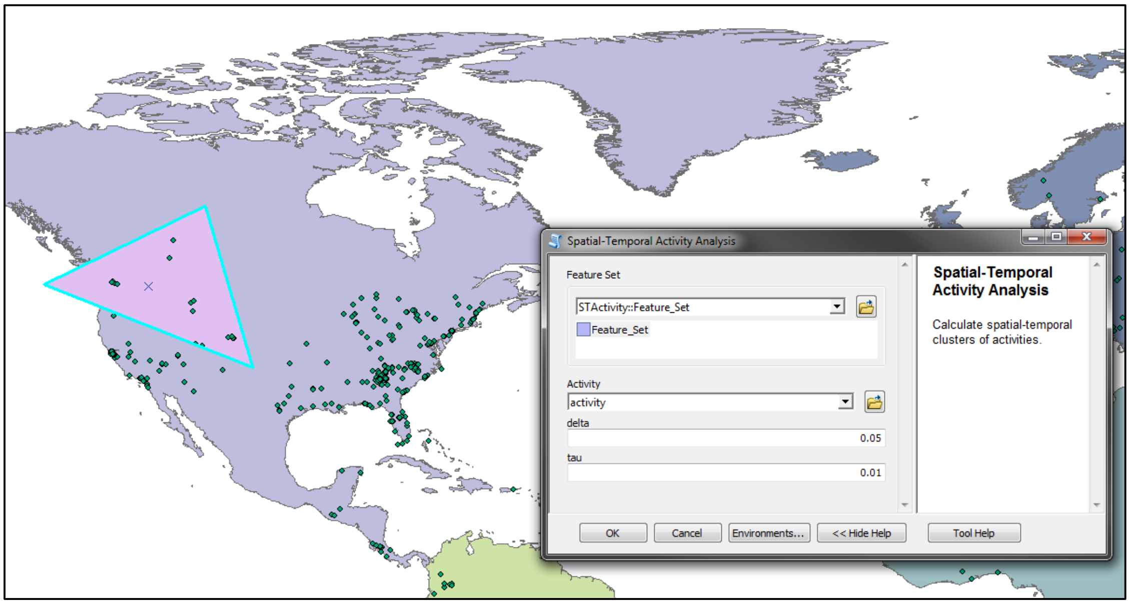
People-based Social Network

Calculate and visualize people-based social network

Figure 2-9 Interface of Query People-Based Social Network

- Identify Spatial-Temporal Interactions

The spatial-temporal interactions are identified with the Knox test (Knox and Bartlett 1964) by using the Pysal python library (Rey and Anselin 2007). This tool (Figure 2-10) will report the identified spatial-temporal interaction based on the user-defined spatial (delta) and temporal (tau) intervals.



Feature Set: STActivity::Feature_Set
Feature_Set
Activity: activity
delta: 0.05
tau: 0.01

Spatial-Temporal Activity Analysis

Calculate spatial-temporal clusters of activities.

Figure 2-10 Interface of Identification of Spatial-Temporal Interactions

Case study

A case study is conducted to test the validity of the proposed LBSMA model. This research has recruited several students in the University of Georgia to collect their Facebook data. The extracted Facebook data is organized in the implemented LBSMA data model. There are 500 unique Facebook accounts and nearly 2,500 posts collected in this case study (Table 2-1). Among those posts, nearly 900 places have been extracted and 48 place categories are identified. The LBSMA database is capable of supporting the analysis of student activities in the spatial-temporal-social dimension.

Table 2-1 Summary of Collected Data

Entity kind	No. of entities in GAE	No. of entities in PostgreSQL
Activity	2,964	2,467
People	568	500
Place		892
Place Category		48
People Social		500
Place Social		892

- Spatial-Temporal-Social Analysis of Human Activities

Table 2-2 summarizes the correlation of different measures of human activities in spatial, temporal and social aspects. The number of human activities and visits and the total travel distance of an individual person are positively correlated to the number of friends in people's social network. This finding confirm the hypothesis of previous study that users who travel have more chances to meet friends, and thus get involved in more social activities (Cheng et al. 2011). In addition, the number of activities and visits are both closely correlated to the number of cliques. This is because people who participant in more social activities are more likely to have a diverse social relationships, i.e., more closed sub-groups in their social networks.

Table 2-2 Spatial-Temporal-Social Analysis of Human Activities

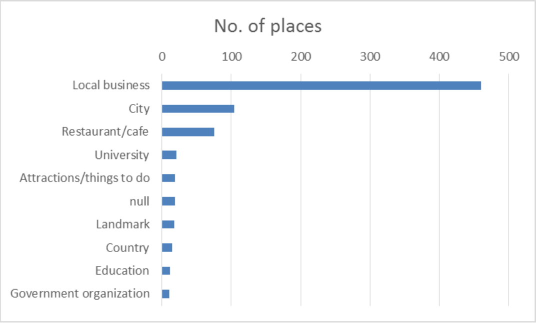
	No. of activities	No. of visits	Sum of travel distance
No. of nodes	0.6186*	0.6754*	0.6060*
No. of cliques	0.8150*	0.8426*	0.5677*
Network density	-0.5510*	-0.6506*	-0.5181*
Avg. Clustering Coef.	0.0757	0.0536	0.0899
* indicates significant at 5%			

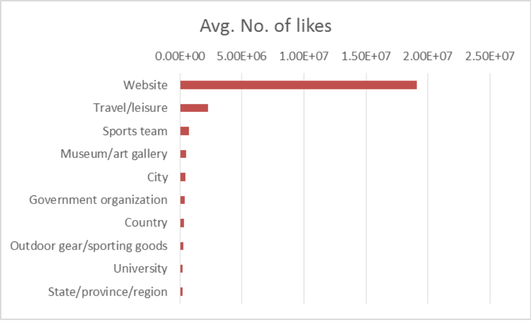
- Popularity Analysis of Places in Virtual and Physical World

The popularity of places are measured based on the number of likes, number of talking about and the number of activities. The total number of likes is how many times Facebook users push the *like* button on a place's Facebook page or website, thus can indicate of the popularity of places on Facebook. The sum of talking about is the total number of times that a place has been mentioned in Facebook user's posts. It is another popularity indicator of places on Facebook. The sum of activities counted the total number of Facebook posts that are embedded the spatial location of this place, hence stands for a popularity of places in the physical world. Table 2-3 demonstrates that the sum of talking about a place has higher correlation with the sum of activities than the sum of talking about. Therefore, the sum of talking about is a better indicator for the popularities of locations in both virtual and physical world.

Table 2-3 Comparison of Popularities of Spatial Locations in Virtual and Physical Space

	Sum of likes	Sum of talking about
Sum of talking about	0.6517*	1
Sum of activities	0.5102*	0.8029*
* indicates significant at 5%		





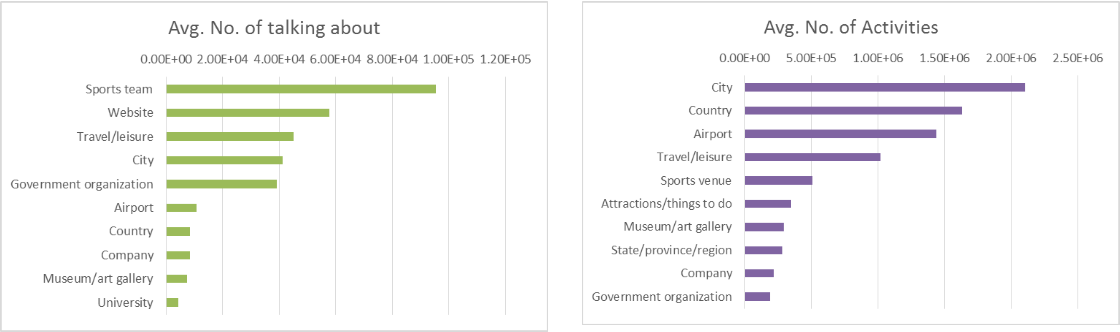


Figure 2-11 Popularity Analysis of Places in Different Categories

Figure 2-11 summaries the popularities of places in different categories. Local business, city and restaurant/caf éhave the most number of places extracted from the collected Facebook posts. The only place classified as website is Google which received the most likes among all the places and are also frequently talked about in Facebook. Places of the travel/leisure and sports team category are also hot topics on Facebook. The places that are classified as city, country and airport received the most physical visits. This demonstrates that people are more like to release their physical location on social media when they are traveling to different cities or countries. In addition to those three categories, travel/ leisure, sports venue, attractions/thing to do, and museum/art gallery are the most popular places in the physical world, because people are more likely to release their social activities in public spaces on Facebook.

- Location-Based Social Network Analysis

Based on the collected activities, Table 2-4 , Figure 2-12, Figure 2-13, Figure 2-14 and Figure 2-15 shows how the location-based social networks vary across space. Different to the people-based social network where the number of activities is closely related to the number of visits (Table 2-2), the number of activities (Figure 2-12) is moderately related to the number of visits (Figure 2-13) for places. This disparity is rooted in the behaviors of Facebook users. Some

Facebook users prefer to post a few status or photos for a single visit while others are eager to publish lots of Facebook posts for an individual visit. The number of visits to a place is also moderately related to the number of visitors (Figure 2-14) but closely related to the number of cliques (Figure 2-15), meaning that places foster a diverse society of customers rather than vast customers attract more visits.

Table 2-4 Location-Based Social Network Analysis

	No. of activities	No. of visits
No. of visits	0.5414*	1
No. of nodes	0.3926*	0.5417*
No. of cliques	0.6727*	0.8105*
Network density	-0.4283*	-0.5659*
Avg. Clustering Coef.	0.0711*	0.0369
* indicates significant at 5%		

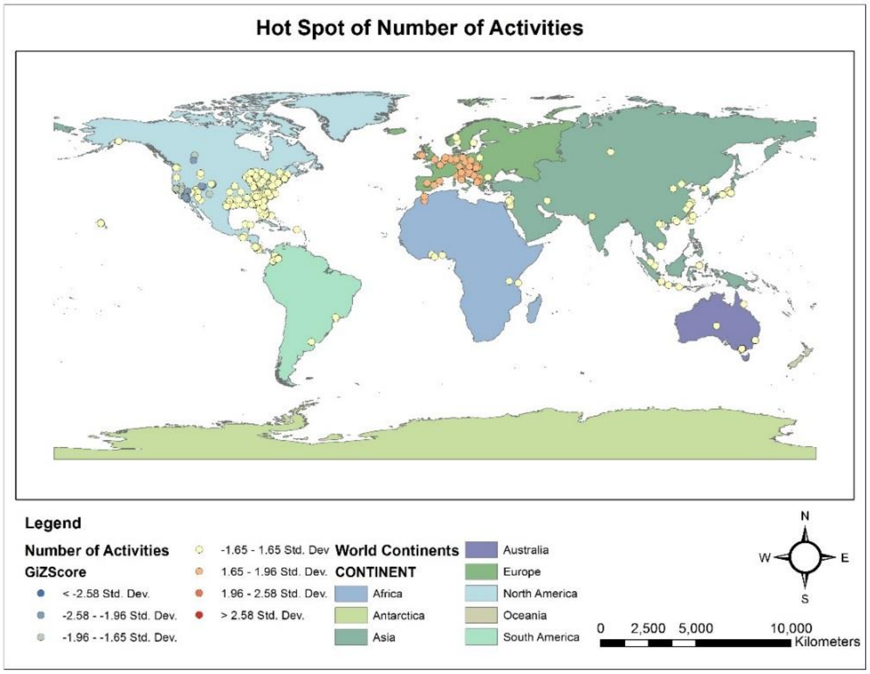


Figure 2-12 Hot Spot of Number of Activities

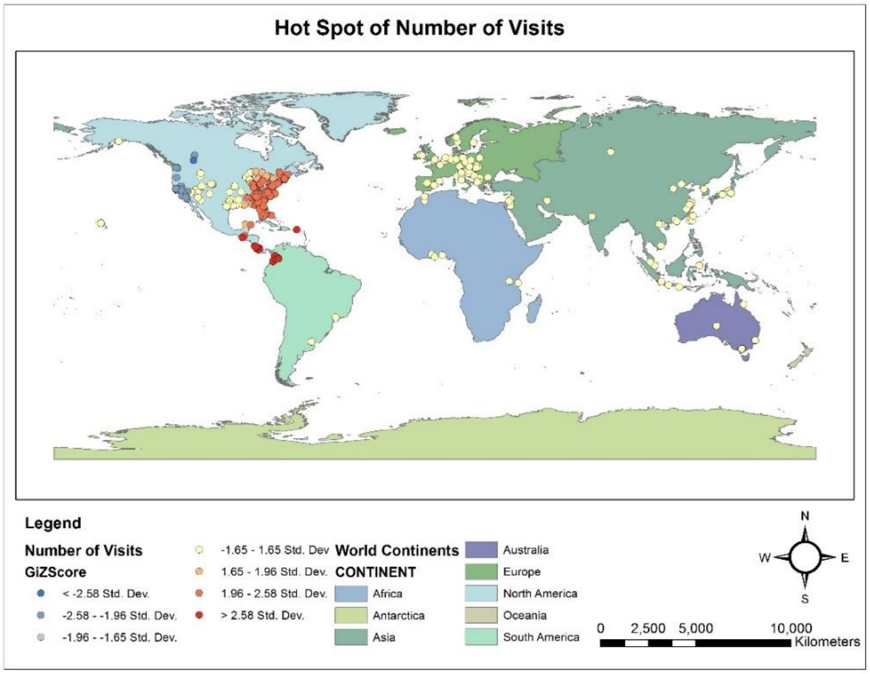


Figure 2-13 Hot Spot of Number of Visits

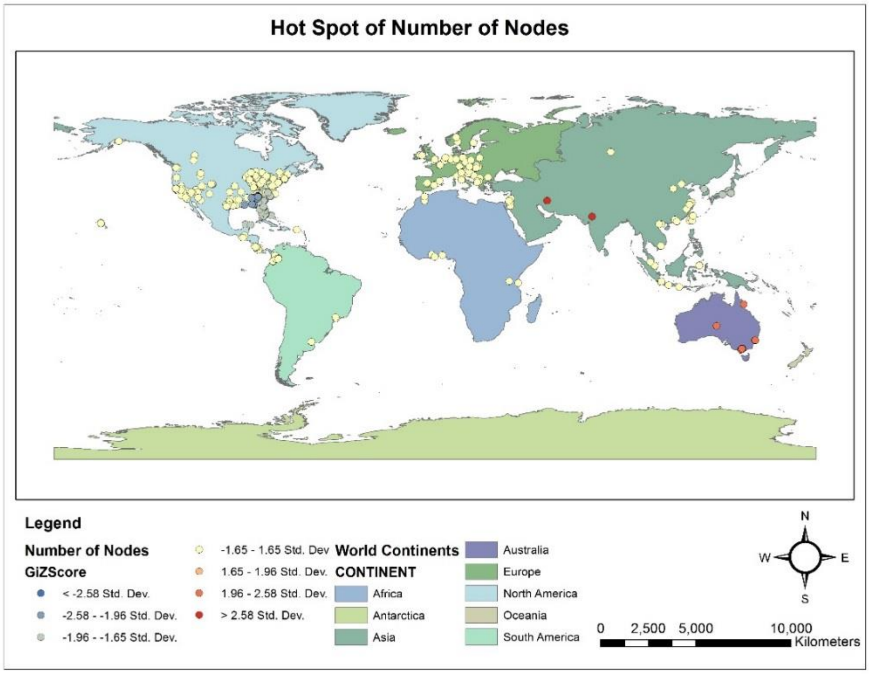


Figure 2-14 Hot Spot of Number of Nodes

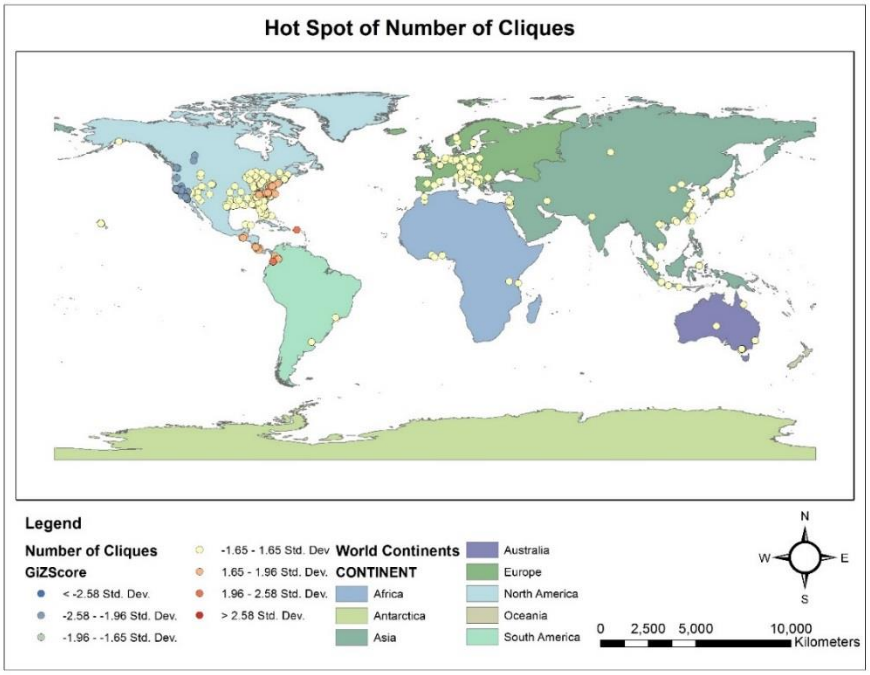


Figure 2-15 Hot Spot of Number of Cliques

Conclusion

Geo-tagged social media activities provide new channels to observe human activities at micro and macro scales. The social media data have naturally and sometimes implicitly embedded space, time, and social connections in the data itself. However, current GIS environment is not suitable for the representation and analysis of such rich data due to its lack of capability to represent the key components of the data. In response to the new opportunity and research challenge, this research aims to fill the blank by developing an ontological framework and a conceptual data model for the representation of social media data in the multiple-dimensional representational space of geography, time, and social connections. Furthermore, a prototype of the conceptual model is implemented in a pilot computer system. Several tools are also developed to query, calculate and visualize human activities in spatial-temporal-social dimension. People-based and location-based social networks can be created and analyzed based on the proposed LSBMA model to add our understanding of human interactions by providing innovative and applicable measures for places, social associations and human activities. The validity of LBSMA model is evidenced by a case study by collecting and analyzing Facebook data. A dedicated website (www.lbsocial.net) is established to conduct an authentic data collection of Facebook data, disseminate aggregated results and significant findings.

In the era of big data, people and environment interact in the physical space and the virtual space (social media) simultaneously. The findings of this research yields new insights regarding human activities in virtual and physical space, and will enhance technical capabilities for social media analysis in GIS. The developed methods can help identify place-based or people-based strategies, e.g., urban planning, traffic planning, commercial advertising or energy communicating. The proposed framework paves new avenues for future research, such as public health, transportation, urban geography and social science. Based on the proposed model and prototype, we believe there are many more potential ways to mine the organized datasets. This study has only provided a case study with a couple application examples, both of which asked questions that are only related to two dimensions of space, time, and social network. Starting from the LBSMA conceptual model, exciting future research avenues include developments of new analytical methods and explorations of new application studies, particularly those involve all three dimensions of the LBSMA data.

References

Anderson, J. 1970. Time-Budgets and Human Geography. *Area* 2 (1):50–51.

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

Chen, J., S.-L. Shaw, H. Yu, F. Lu, Y. Chai, and Q. Jia. 2011. Exploratory data analysis of activity diary data: a space–time GIS approach. *Journal of Transport Geography* 19 (3):394–404.

Cho, E., S. A. Myers, and J. Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1082–1090. ACM http://dl.acm.org/citation.cfm?id=2020579 (last accessed 6 April 2014).

Dodgshon, R. A. 2008. GEOGRAPHY'S PLACE IN TIME. *Geografiska Annaler Series B: Human Geography* 90 (1):1–15.

Dragicevic, S., and D. J. Marceau. 2000. A fuzzy set approach for modelling time in GIS. *International Journal of Geographical Information Science* 14 (3):225–245.

Eagle, N., A. Pentland, and D. Lazer. 2009. Inferring friendship network structure by using mobile phone data. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* 106 (36):15274 – 15278.

Goodchild, M. F., M. Yuan, and T. J. Cova. 2007. Towards a general theory of geographic representation in GIS. *International Journal of Geographical Information Science* 21 (3):239–260.

Hagberg, A. A., D. A. Schult, and P. J. Swart. 2008. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, 11–15. Pasadena, CA USA.

H ägerstrand, T. 1970. WHAT ABOUT PEOPLE IN REGIONAL SCIENCE? *Papers in Regional Science* 24 (1):7.

Hornsby, K., and M. J. Egenhofer. 2000. Identity-based change: a foundation for spatio-temporal knowledge representation. *International Journal of Geographical Information Science* 14 (3):207–224.

Humphreys, L. 2007. Mobile Social Networks and Social Practice: A Case Study of Dodgeball. *Journal of Computer-Mediated Communication* 13 (1):341–360.

Kim, H.-M., and M.-P. Kwan. 2003. Space-time accessibility measures: A geocomputational algorithm with a focus on the feasible opportunity set and possible activity duration. *Journal of Geographical Systems* 5 (1):71.

Knox, E. G., and M. S. Bartlett. 1964. The Detection of Space-Time Interactions. *Applied Statistics* 13 (1):25.

Kuijpers, B., H. J. Miller, T. Neutens, and W. Othman. 2010. Anchor uncertainty and space-time prisms on road networks. *International Journal of Geographical Information Science* 24 (8):1223.

Kwan, M.-P. 2010. A Century of Method-Oriented Scholarship in the Annals. *Annals of the Association of American Geographers* 100 (5):1060–1075.

———. 2004. GIS Methods in Time-Geographic Research: Geocomputation and Geovisualization of Human Activity Patterns. *Geografiska Annaler Series B: Human Geography* 86 (4):267.

———. 2007. Mobile Communications, Social Networks, and Urban Travel: Hypertext as a New Metaphor for Conceptualizing Spatial Interaction. *Professional Geographer* 59 (4):434–446.

———. 1998. Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-based Framework. *Geographical Analysis* 30 (3):191–216.

Lewis, K., J. Kaufman, G. A. Marco, W. B. Andreas, and C. A. Nicholas. 2008. Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks* 30:330–342.

Li, Y., H. Gao, M. Yang, W. Guan, H. Ma, W. Qian, Z. Cao, and X. Yang. 2013. What are Chinese Talking about in Hot Weibos? *arXiv preprint arXiv:1304.4682*. http://arxiv.org/abs/1304.4682 (last accessed 26 December 2013).

Mennis, J. L. P., Donna J.Qian, Liujian. 2000. A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science* 14 (6):501.

Miller, H. J. 1999. Measuring space-time accessibility benefits within transportation networks: basic theory and computational procedures. *Geographical analysis* 31 (1):1–26.

———. 2003. What about people in geographic information science? *Computers, Environment and Urban Systems* 27:447.

Miller, H. J., and S. A. Bridwell. 2009. A Field-Based Theory for Time Geography. *Annals of the Association of American Geographers* 99 (1):49–75.

Naaman, M., J. Boase, and C.-H. Lai. 2010. Is it really about me?: message content in social awareness streams. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 189–192. ACM http://dl.acm.org/citation.cfm?id=1718953 (last accessed 11 December 2014).

Neutens, T., N. Weghe, F. Witlox, and P. Maeyer. 2008. A three-dimensional network-based space–time prism. *Journal of Geographical Systems* 10 (1):89–107.

Neutens, T., F. Witlox, N. Van De Weghe, and P. H. De Maeyer. 2007. Space-time opportunities for multiple agents: a constraint-based approach. *International Journal of Geographical Information Science* 21 (10):1061–1076.

Peuquet, D., and N. Duan. 1995. An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographical Information Systems* 9 (1):7–24.

Peuquet, D. J. 1994. It’s about Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems. *Annals of the Association of American Geographers* (3):441.

———. 2001. Making Space for Time: Issues in Space-Time Data Representation. *GeoInformatica* 5 (1):11.

Pickles, J. C. 1995. Representations in an Electronic Age: Geography, GIS, and Democracy. In *Critical Geographies: A Collection of Readings*, eds. H. Bauder and S. E.-D. Mauro, 637–663. British Columbia, Canada.: Praxis (e)Press.

Pultar, E., T. J. Cova, Y. May, and M. F. Goodchild. 2010. EDGIS: a dynamic GIS based on space time points. *International Journal of Geographical Information Science* 24 (3):329–346.

Raubal, M., H. J. Miller, and S. Bridwell. 2004. User-Centred Time Geography for Location-Based Services. *Geografiska Annaler Series B: Human Geography* 86 (4):245–265.

Rey, S. J., and L. Anselin. 2007. PySAL: A Python Library of Spatial Analytical Methods. *The Review of Regional Studies* 37 (1):5–27.

Ronald, N., V. Dignum, C. Jonker, T. Arentze, and H. Timmermans. 2012. On the engineering of agent-based simulations of social activities with social networks. *Information and Software Technology* 54 (6):625–638.

Russell, M. A. 2013. *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. O'Reilly Media, Inc.

Schwanen, T., and M.-P. Kwan. 2012. Critical Space-Time Geographies: Guest Editorial. *Environment and Planning A* 44 (9):2043–2048.

Shaw, S.-L., and H. Yu. 2009. A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. *Journal of Transport Geography* 17 (2):141–149.

Shaw, S.-L., H. Yu, and L. S. Bombom. 2008. A Space-Time GIS Approach to Exploring Large Individual-based Spatiotemporal Datasets. *Transactions in GIS* 12 (4):425–441.

Sui, D., and M. Goodchild. 2011. The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science* 25 (11):1737.

Takhteyev, Y., A. Gruzd, and B. Wellman. 2012. Geography of Twitter networks. *Social Networks* 34 (1):73–81.

Tavakoli, M., and S. Fakhraie. 2011. THEORY OF TIME GEOGRAPHY (EXPLANATION OF ITS APPLICABLE QUANTITIES IN PLANNING). *International Journal of Academic Research* 3 (3):655–662.

Thrift, N. 1996. *Spatial Formations*. SAGE Publications. http://books.google.com/books?id=B2hw8X-yigIC.

Tsou, M.-H., and M. Leitner. 2013. Visualization of social media: seeing a mirage or a message? *Cartography and Geographic Information Science* 40 (2):55.

Wang, D., and T. Cheng. 2001. A spatio-temporal data model for activity-based transport demand modelling. *International Journal of Geographical Information Science* 15 (6):561–585.

Worboys, M. 2005. Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science* 19 (1):1–28.

Yao, X. 2010. Modeling Cities as Spatio-Temporal Places. In *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, 311–328. Netherlands: Springer http://proxy-remote.galib.uga.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=76899530&site=eds-live.

Yin, L., S.-L. Shaw, and H. Yu. 2011. Potential effects of ICT on face-to-face meeting opportunities: a GIS-based time-geographic approach. Journal of Transport Geography 19 (3):422–433.

Yin, L., and S.-L. Shaw. 2015. Exploring space–time paths in physical and social closeness spaces: a space–time GIS approach. International Journal of Geographical Information Science 29 (5):742.

Yuan, L., S. Chen, Y. Wang, Z. Yu, W. Luo, and G. Lü. 2010. CAUSTA: Clifford Algebra-based Unified Spatio-Temporal Analysis. *Transactions in GIS* 14:59–83.

Yuan, M. 1997. Use of knowledge acquisition to build wildfire representation in Geographical Information Systems. *International Journal of Geographical Information Science* 11 (8):723–746.

Yuan, M., A. Nara, and J. Bothwell. 2014. Space–time representation and analytics. *Annals of GIS* 20 (1):1–9.

Yu, H. 2006. Spatio-temporal GIS Design for Exploring Interactions of Human Activities. *Cartography and Geographic Information Science* 33 (1):3–19.

Yu, H., and S.-L. Shaw. 2008. Exploring potential human activities in physical and virtual spaces: a spatio-temporal GIS approach. *International Journal of Geographical Information Science* 22 (4):409–430.

CHAPTER 3 THE RANDOM WALK VALUE FOR RANKING SPATIAL CHARACTERISTICS IN ROAD NETWORKS [2]

[2] Wei, X., and X. A. Yao. 2014. Geographical Analysis 46 (4):411–434. Reprinted here with permission of publisher.

Abstract

This study proposes a new network index at both nodal and link levels to rank the spatial characteristics of individual network components. The objective is to create a network metric that captures socioeconomic characteristics in urban environments. Because this index is based on the random walk simulation modeling strategy, it is coined the Random Walk Value (RWV). An algorithm and an associated software tool were developed to calculate the RWVs of network components. Compared with other popular network indices, the unique advantage of the RWV is that this index considers not only spatial structural or topological characteristics, but also physical characteristics of network components. Two case study cities, the Chinese city of Wuhan and the United States city of Atlanta, were chosen to test the utility of the RWV. These two case studies yield several findings. First, the RWV is highly consistent with some of the most widely used network measures, such as closeness and connectivity measures, which was evidenced by strong correlations between RWV and other network. Second, the RWV has been proven to be a good indicator of spatial importance, and a better predictor of socioeconomic variables in urban environments. The RWV outperforms all other network indices in terms of its correlations with important socioeconomic variables, and its ability to predict some of them. Third, both case studies confirm that the RWV can be a good substitute for some important socioeconomic variables, such as population density and job density, in spatial modeling. This finding is significant for studies when population and job data are not available, or for studies that attempt to predict future scenarios.

Introduction

Road networks are essential components of urban systems that facilitate various types of activities in urban life. An urban road structures determine patterns of human movements (Jiang and Jia 2011), and the spatial layout of urban streets has a direct impact on human social activities (Jiang and Claramunt 2004). The structure of an urban transportation system and the spatial pattern of land use in a city are mutually dependent. The relationships have been well established and supported by many studies (Taaffe, Gauthier, and O’Kelly 1996). Dynamic relationships between the two subsystems of a city can be partly explained by the spatial distribution of network characteristics, which often have been measured by network indices such as connectivity and centrality.

A considerable body of literature supports that network characteristics of urban transportation can significantly influence the spatial distributions of socioeconomic characteristics. Influences of network structure are seen in a variety of urban characteristics, including traffic distribution (Wewal et al. 2010), population density (Chan, Donner, and Lämmer 2011), employment growth (Ozbay, Ozmen, and Berechman 2006), distribution of public facilities (Aderamo and Magaji 2010), and even economic prospects (Blanchard and Volchenkov 2008). Road networks also have been used as a spatial framework to estimate probable locations of heavily trafficked nodes (Horner, Zook, and Downs 2012) or critical facilities (Lei 2013). Therefore, a better understanding of urban road network characteristics is essential for improved outcomes of research about the spatial distributions of urban socioeconomic characteristics.

Many scholars posited that people's perceptions of space determine local movements (Hillier 1999; Blanchard and Volchenkov 2010). This study further argues that perception of spatial characteristics in a network also determines spatial decisions that ultimately lead to spatial patterns of socioeconomic activities. Prior studies already prove that characteristics of individual network components can be highly correlated with socioeconomic characteristics of each component's proximal area. A measure of node importance, for instance, provides a basis for the identification of urban agglomeration (Song et al. 2011). The Closeness centrality measure of nodes has been found to be correlated with both population and employment density (Wang, Antipova, and Porta 2011) and the location of economic activities (Porta et al. 2012). The Integration network metric, which is based on the theories and techniques of space syntax, can help predict the spatial variations of human movements in an urban environment (Jiang 2009). The same study also reports that, compared with space syntax metrics, Google's PageRank scores as well as *betweenness* and degree centrality indices are even better indictors of urban traffic volume(Jiang 2009).

However, existing network measures primarily are based only on the topological structure of a network. Physical characteristics of road network components (e.g., intersections and road segments) have been almost ignored in these measures. Yet their impacts on an urban economy have been intensively investigated in the field of transportation geography. The goal of transportation is believed to provide transportability; i.e., ease of movement of passengers, freight or information. These physical characteristics determine accessibility and travel costs on road networks, and further impact the location, scale, scope, agglomeration, and density of urban economic sectors (Rodrigue, Comtois, and Slack 2013). Based on empirical evidences, Wang et al. (2012) argue that road characteristics cannot be defined solely by topological properties, but by topological properties as well as other physical properties related to the usage of road segments. Therefore, previous network metrics may not be sufficient to maximally capture spatial characteristics of network components so as to use such metrics to predict the socioeconomic characteristics of corresponding proximal areas. The absence of a suitable metric in the literature may explain why even though correlations were observed between network measures and urban socioeconomic variables, few studies, if any, attempted to formally investigate the potential of network metrics for the prediction of socioeconomic characteristics. This study attempts to fill the blank by developing a new network metric that can be a good predictor of socioeconomic characteristics in urban environments. By incorporating both topological properties and physical characteristics of network components, we argue that the newly developed index can be more theoretically plausible than traditional ones.

Theoretical Framework

A network can be represented as a graph in which nodes are connected by links. Many concepts and associated quantifiable measures are developed with this node-link representation. Each node and link is a component of a network. Therefore, a network component refers to a node or a link in this paper. Network components are nodes-and-links will be used interchangeably hereafter. Based on such a network representation, various methods have been proposed to analyze characteristics of network components. For the convenience of discussion, this study groups the previous methods of characterizing network components into two major types of approaches, namely index assessment and simulation.

- Index Assessment

In a traditional transportation network, the index assessment approach either measures network components characteristics, or summarizes specific characteristics in respective proximal areas of network components. The former includes connectivity or centrality measures as well as other qualitative attributes of road networks, such as road class and traffic conditions (Bono, Gutiérrez, and Poljansek 2010). The latter type of indices can be socioeconomic variables, such as population, jobs, income (Song et al. 2011), or urban form variables such as mean NDVI values and land use categories (Lindsey et al. 2007; Vance and Hedel 2008).

Connectivity and centrality are the most commonly used network indices. Connectivity measures the strength of connections between nodes (or links) with respect to other nodes (or links) (Brandes and Erlebach 2005). At the nodal (or link) level, the degree of a node (or link) in a network can be defined as the number of nodes (or links) that are directly connected to it. The concept of centrality concerns the relative importance of a node or a link in a network. As such, a link may be said to have higher centrality if it is used more often for traverses. A number of indices have been proposed to measure centrality. The four most popular ones are degree, closeness, betweenness, and eigenvector centralities. Degree centrality is based on the number of direct connections to each component, it essentially is a measure of connectivity as well. Closeness centrality of a given node is defined by an inverse function of the sum of distances between this node and all other nodes in a network. Betweenness centrality measures the number of times a node serves on the shortest paths among all pairs of nodes in a network. Eigenvector centrality considers reciprocal relationships between connected neighbors by finding a principal eigenvector of the adjacency matrix of a network representation. It assigns relative centrality scores to each node with more contributions from higher-scoring neighbors. Although centrality measures are generally applied in network science, many studies use these indices to characterize the structural properties of urban street networks (Jiang and Claramunt 2004; Bono, Gutiérrez, and Poljansek 2010; Wang et al. 2011; Zhang and Li 2011).

More recently, the PageRank algorithm was developed to rank the importance of webpages and has been used by the Google web search engine (Brin and Page 2012). The algorithm has also been applied in transportation network analysis. It is considered a variant of the eigenvector centrality measure because it considers not only the number of neighbors, but also the quality of each neighbor. The weighted PageRank algorithm, a modified version of the original, takes into account the importance of linked pages, and distributes rank scores based on the popularity of the pages (Xing and Ghorbani 2004). Both algorithms have been utilized in urban network analysis, and have been found to outperform other metrics in estimating human movements or predicting traffic (Jiang 2009; Jiang and Jia 2011).

Space syntax metrics (Hillier 1999) have been widely adopted to measure network characteristics. Space syntax includes a set of new concepts and techniques that can analyze spatial configurations. Space syntax metrics, such as connectivity, depth, and integration are popular in research (Li and Guo 2003). Connectivity in space syntax refers to the number of spatial units connected to a given spatial unit. Depth indicates the minimum number of spatial units that are connected to a given spatial unit in a certain step. Integration measures the number of turns one has to make from one link to reach other links by way of the shortest paths in a network. In street network analysis, many studies show that the space syntax measures are able to quantify traffic characteristics such as vehicle traffic volume (Croxford, Penn, and Hillier 1996) and pedestrian volume (Porras et al. 2002; Baran, Rodríguez, and Khattak 2008). In a study that predicted human movements in an urban environment, Jiang (Jiang 2009) found that

60 percent of the spatial variation in human movements can be explained by a space syntax measure from a topological point of view. Interestingly, the same study also showed that physical characteristics of roads such as road width, and building height as well as the land use type of the vicinity area, may account for the other 40 percent of the spatial variations.

- Simulation models

Various simulation models have been developed to imitate human activities and responses to increase understanding of complex systems, such as cities (Benenson and Torrens 2004). The literature contains many traffic simulation modeling strategies, including discrete agent-based methods (Benenson and Torrens 2004), the grid model method (Miyagawa 2009), and continuous traffic flow methods that explore traffic jams or evaluate network configurations (Wewal et al. 2010).

In comparison with the index assessment approach, most simulation models for networks predict or estimate traffic dynamics on a network by imitating agent behavior within a network configuration. Only a few previous studies along this line of research specifically assess the characteristics of network components. In the small number of studies, the random walk method has been particularly popular.

Random Walk is a special type of simulation modeling. The concept initially was proposed by Pearson in 1905 to estimate the probability that after n steps of random walks a person is at a distance between r and r + d from his/her starting point (Pearson 1905). This method has attracted much interest through a variety of research efforts, ranging from electric communication (Leng et al. 2007) and social problems (Short et al. 2008), to complex network (Yang 2005). In time geography, random walk simulation has been used to predict the probability distribution of an agent's location over time (Winter and Yin 2011). In urban studies, the random walk algorithm has been applied to detect community structures (Pons and Latapy 2005), and to identify isolated urban areas (Volchenkov and Blanchard 2007). Li, Zhao, and Yuki (2009) apply the random walk algorithm to rank node importance of each road intersection with regard to its likelihood to be visited.

Although the random walk algorithm shows promise in characterizing network components, few studies can be found in the literature and two problems have been uncovered in the few available prior studies. First, previous studies only considered un-weighted networks (Chen and Chen 2007). The probability of path selection is assumed to be uniform and all of the walking distances are set to a fixed value. However, some important characteristics of network components, such as capacity, speed limit, and the turning angle at an intersection, may have direct or indirect, and significant or minor, impacts on route choices (Behrens and Kane 2004); (Watling, Milne, and Clark 2012). Second, previous studies of the random walk algorithm never explored its potential to develop characteristic measures of network components.

Research Design

The research integrates both topological characteristics and other innate attributes of nodes and links in a network to rank the spatial importance of each network component. The objective is to establish a new metric for nodes and links that can serve as a significant indictor of some socioeconomic characteristics in the proximal area around network components. Because the new index is based on the random walk simulation approach, it is dubbed the random walk value (RWV). As described in Figure 3-1, the RWV attempts to capture both physical and topological characteristics of network components. Therefore the authors hypothesize that the RWV might more representative and more strongly associated with socioeconomic characteristics of the proximal areas of network components.

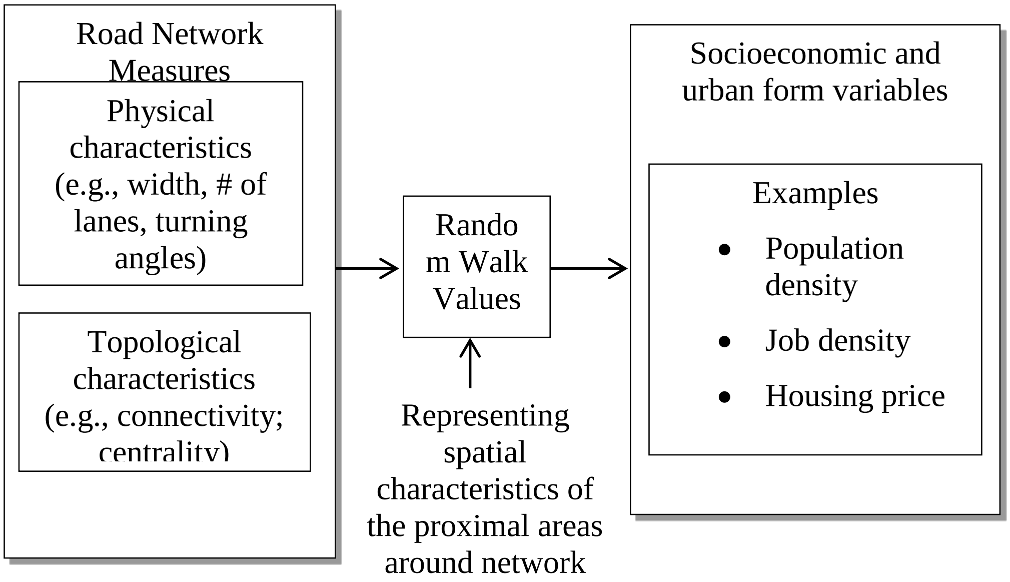
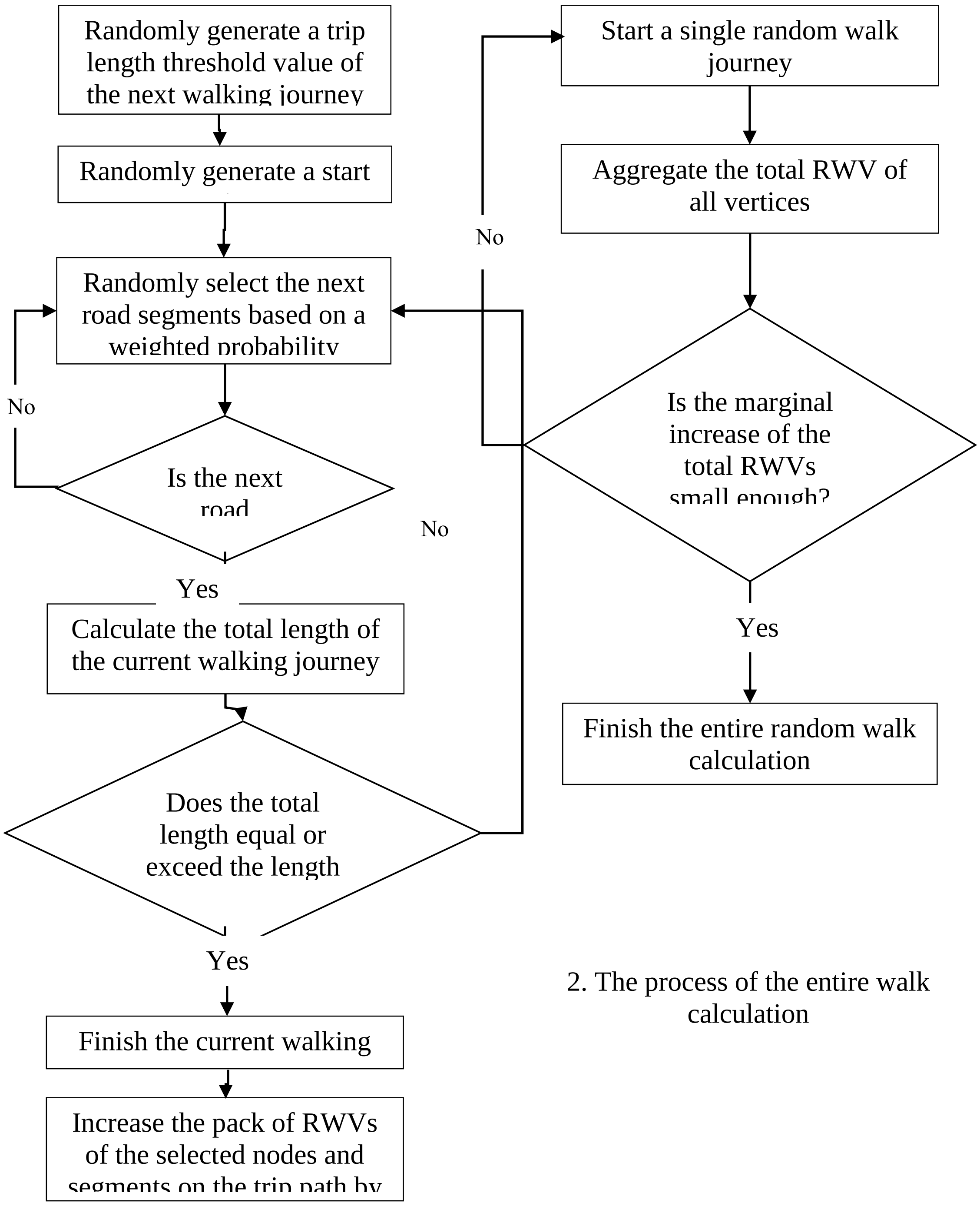

Road Network Measures
Physical characteristics (e.g., width, # of lanes, turning angles)
Topological characteristics (e.g., connectivity; centrality)
Random Walk Values
Representing spatial characteristics of the proximal areas around network
Socioeconomic and urban form variables
Examples
• Population density
• Job density
• Housing price

Figure 3-1. A Conceptual Framework

The RWV algorithm

Based on the original idea of random walk simulation, this algorithm makes improvements in the following ways. First, it not only considers the topological structure of road networks, but it also incorporates the physical characteristics of network components. Secondly, it considers the statistical distribution of trip lengths when randomly generating the trip lengths. Thirdly, it simulates a large number of walks on a network until a convergence criterion is met. The general process is illustrated in Figure 3-2. To start a trip, a node of trip origin is randomly selected, and a trip length is randomly generated. Then the actual path of each trip is incrementally formed by choosing a link segment each time a road intersection (or node) is visited. A path is complete when the trip length is reached. The RWV of every traversed network component on the trip path increases by 1 when a trip path is generated.


Randomly generate a trip length threshold value of the next walking journey
Randomly generate a start
Randomly select the next road segments based on a weighted probability
No
Is the next road
Yes
Calculate the total length of the current walking journey
Does the total length equal or exceed the length
No
Yes
Finish the current walking
Increase the pack of RWVs of the selected nodes and segments on the trip path by
Start a single random walk journey
Aggregate the total RWV of all vertices
No
Is the marginal increase of the total RWVs small enough?
Yes
Finish the entire random walk calculation

2. The process of the entire walk calculation

1.The process of single random walk journey

Figure 3-2. A Flowchart of the RWV Algorithm

Figure 3-3 shows a simulated trip path, which starts with a random point *A*. A decision of the next road segment is to be made at nodes *A*, *B*, and *C*, respectively, until a randomly generated trip length *l* is reached.

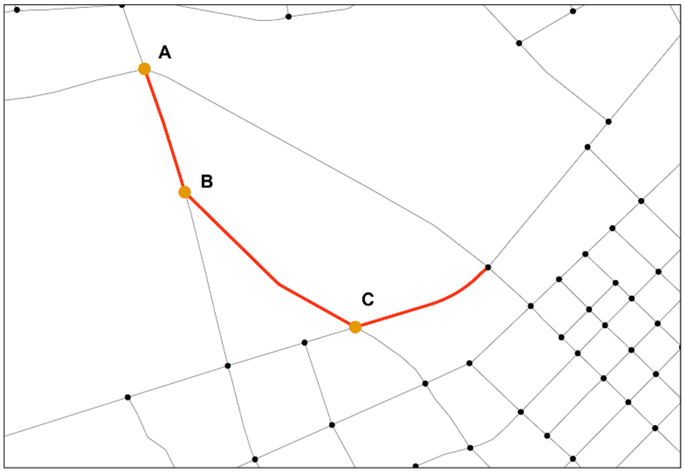


Figure 3-3. A Simulated Trip Path

Apparently, the choice of the next link in a sequence is a critical decision. This also is the step in which characteristics of network components are the primary consideration. This process basically is about solving a discrete choice problem of choosing one of the several connected roads at an intersection. A number of discrete choice models are available, including the most popular random utility model (refer to (Train 2009) for more details). Most models establish a probability distribution expressed as a function of characteristics of the available choices. The rationale here is that the probability for a link selection is proportional to the chosen key characteristic of the link. As expressed in Equation (1) and Figure 3-4, this study applies a weighted proportional function to estimate the probability for each link to be taken as the next step. In the Figure 3-4 example, a random traveler coming from node N_1 and arriving at node N_2

faces a choice between links $l_{2,3}$ (link from N_2 to N_3), $l_{2,4}$, and $l_{2,5}$ to continue the trip. The probability that this walker selects $l_{2,3}$ is modeled by Equation (1):

$$p_{l_{2,3}} = \frac{w_{l_{2,3}}}{w_{l_{2,3}}+w_{l_{2,4}}+w_{l_{2,5}}} \qquad (1)$$

where p stands for the probability of selecting a link, and w is the value of the weight variable of the associated link, such as the width or speed limit of a road segments. Based on p values, a Monte Carlo simulation can be performed to make a choice according to the preceding probability distribution. The algorithm excludes the incoming link ($l_{1,2}$ in this case) from the set of candidates in order to avoid the occurrence of a meaningless circular path.

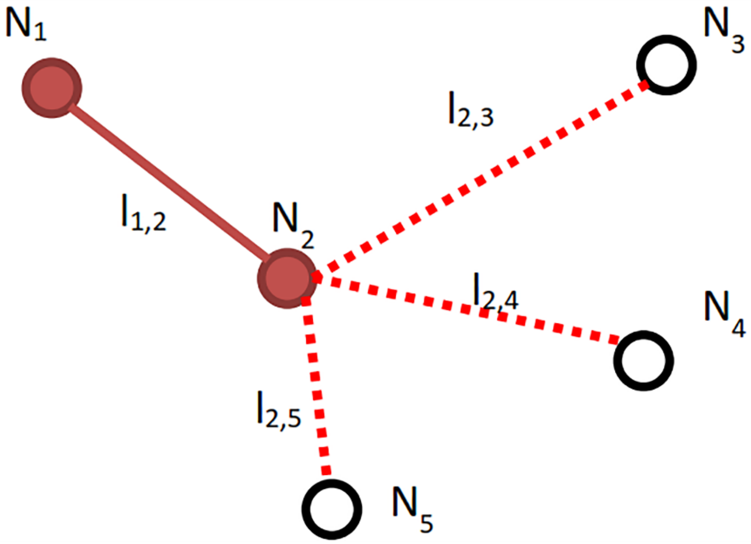


Figure 3-4 Link Selection

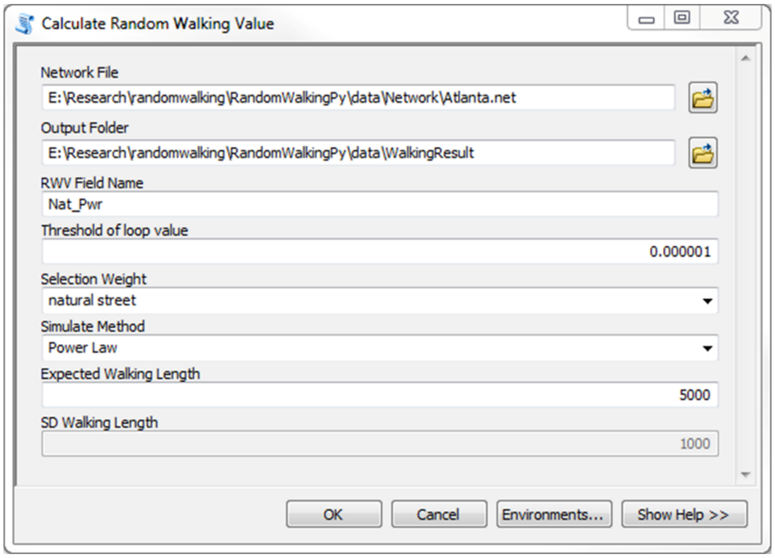
After choosing a link and adding it to the simulated travel path, the total length of the current journey is updated. On the condition that the total length equals or exceeds the predefined threshold trip length, the current random walk trip path is complete, and the RWVs of all nodes and links (road segments) on the path are increased by 1. Then the sum of the RWVs for all nodes in the network is compared with the sum before this travel path was added. If the relative contribution of the new path is smaller than a threshold value (e.g. 0.00001), the RWVs of the network components are considered to be in a stable situation, which signals the conclusion of the algorithm. The stopping condition is specified by Equation (2):

$$\frac{\sum R_{i,t} - \sum R_{i,t-1}}{\sum R_{i,t}} \leq \text{threshold} \qquad (2)$$

where *R* stands for the RWV scores, subscript *i* indicates node *i*, and *t* is the iteration sequence number of the simulation.

Critical Parameters and a Software Tool

A software tool was developed to implement the RWV algorithm. The tool is programmed in Python in the form of an ArcGIS python toolbox. Figure 3-5 portrays the interface of the tool. The python tool creates a net file from an existing road network shape file, calculates RWVs and exports the results to a point shapefile for nodes, and to a polyline shapefile for links. In the interface, users can define the stopping condition in Equation (2), and the two critical parameters, as discussed next.



Calculate Random Walking Value

Network File
E:\Research\randomwalking\RandomWalkingPy\data\Network\Atlanta.net

Output Folder
E:\Research\randomwalking\RandomWalkingPy\data\WalkingResult

RWV Field Name
Nat_Pwr

Threshold of loop value
0.000001

Selection Weight
natural street

Simulate Method
Power Law

Expected Walking Length
5000

SD Walking Length
1000

Figure 3-5. The Interface of the RWV Calculation Tool

Two critical factors require special attention. The first is about trip distribution in the simulation process, and the second is the weight variable. When randomly generating a set of trip lengths, the distribution of the simulated trip lengths needs to conform to observed distributions in the real world. Power-law (Clauset, Shalizi, and Newman 2009) and normal distributions are the two most popular forms of length distribution for urban trips. A power distribution of trip lengths suggests that the corresponding frequency (or probability) y is the power function of the trip length x, as shown in Equation (3).

$$y = x^{a} \qquad (3)$$

In this paper, the exponent a is set to 2.5, which has been widely adopted for urban trips. This study applies Newman's (Newman 2005) method for generating trip lengths following a power-law distribution. In comparison, the normal distribution of trip lengths means that the frequency or probability follows a Gaussian distribution.

Many possible road characteristics are viable choices for the weight variable for link selection. This study tests three representative choices: *natural-street* which captures angular relationships in the network structure, *connectivity* which measures the centrality of a network component from a topological perspective, and road width (or speed limit) which describes a physical characteristic of network components. Natural streets are joined road segments of good continuity and often are identified from adjacent street segments that have the smallest deflection angles (Liu and Jiang 2012). Subscribing to the Gestalt principle of good continuity, Jiang and colleagues (Jiang et al. 2008; Liu and Jiang 2010; Jiang and Jia 2011) argue that drivers often prefer to choose natural streets whenever possible. This position seems to be supported by not only their empirical studies, but also by the pattern of human travel routes that were identified by massive GPS data (Turner 2009). When a natural street was chosen as the weight variable, the algorithm prioritizes a connected natural street as the most likely road segment for the next step in a simulation.

Validation, Sensitivity Analysis, and Case studies

To test the validity of the developed algorithm and the software tool, a preliminary study was performed using a road network dataset for Wuhan, China. The preliminary study also evaluated the sensitivity of results to the aforementioned critical parameters of the algorithm. After the validation and sensitivity analysis, two case studies were conducted to examine the usefulness of the RWV. Wuhan in China and Atlanta in the United State are the two case study cities. They are one of the largest metropolitan areas in their respective countries. Both cities are characterized by high population density, and both serve an important regional role as economic and transportation hubs. In the validation tests and in the case studies, the RWV algorithm was run in six different parameter settings, as enumerated in below Table 3-1.

Table 3-1 The RWV Index Obtained with Different Parametric Settings

	Weight variable	Trip Length Distribution
RWV1	width (in Wuhan study)	power-law distribution
	speed (in Atlanta study)	
RWV2	width (in Wuhan study)	normal distribution
	speed (in Atlanta study)	
RWV3	natural street	power-law distribution
RWV4	natural street	normal distribution
RWV5	connectivity	power-law distribution
RWV6	connectivity	normal distribution

For comparison, several widely used network indices also were calculated based on the same data. Then a correlation analysis was conducted to examine relationships between the RWV and these indices, which include centrality measures, space syntax metrics, and PageRank scores. Space syntax metrics were computed by the program XWoman (Jiang and Claramunt

2002). Centrality measures and PageRank scores were calculated in Python with the NetworkX library. All of these measures were estimated for each network component, whenever possible. Some indices, such as the PageRank scores, are inherently applicable at the node level only. Similarly, the space syntax measures were calculated only for links. Although only the correlation analysis is reported for these case studies, significant correlations between the RWV and these widely used network metrics also evidence the validity of the algorithm and the associated tool.

Validity and Stability Tests

Because the RWV algorithm is simulation based, the stability of resulting RWV scores is important. To test this stability, the tool of the RWV algorithm is performed six consecutive times on the same data with the same parameter setting. Statistics for the six sets of RWVs are reported in Table 3-2. They suggest that results from multiple runs are consistent, given the inherent randomness of simulation.

Table 3-2 Descriptive Statistics for Generated RWVs of Nodes for Six Tests

	Test1	Test2	Test3	Test4	Test5	Test6
Average	830.59	828.53	829.73	828.80	828.53	829.56
Standard Error	8.90	8.92	9.05	8.92	8.79	8.85
Median	779.00	783.00	781.00	780.00	780.00	786.00
Mode	692.00	630.00	600.00	739.00	853.00	804.00
S.D.	309.20	309.80	314.27	309.79	305.28	307.46
Variance	95602.48	95973.40	98766.13	95970.81	93193.89	94529.82
Range	1802.00	1859.00	1782.00	1839.00	1836.00	1851.00
Min	140.00	122.00	123.00	117.00	108.00	99.00
Max	1942.00	1981.00	1905.00	1956.00	1944.00	1950.00

Sensitivity Analysis

An analysis is conducted to test the sensitivity of the results to the two critical parameters. Table 3-1shows six parameter settings, each of which is a unique combination of the realized choices of the two critical parameters. The corresponding six sets of RWVs are calculated using the developed tool. Figure 3-6 is the scatterplot matrix showing relationships between each pair of the RWV sets. Results are not sensitive to the parameter of trip length distribution. RWV1 and RWV2 are two sets of results based on different forms of trip length distribution, but with an identical weight matrix. The scatterplot of RWV1 versus RWV2 follows the diagonal line closely, which suggests that the two sets of values are highly similar. The same is true for the other two pairs of RWVs that have different types of trip length distribution when other parameters are kept the same. However, the simulated RWV results are relatively more sensitive to the weight variable parameter. When other parameters are held constant and only the weight variable changes, the set of RWVs (e.g., RWV1, RWV3, and RWV5) are more obviously dispersed away from the diagonal lines in respective scatterplots. Among them, the choice of "Natural Street" yields results that are slightly more different from that of other choices (i.e., road width and connectivity).

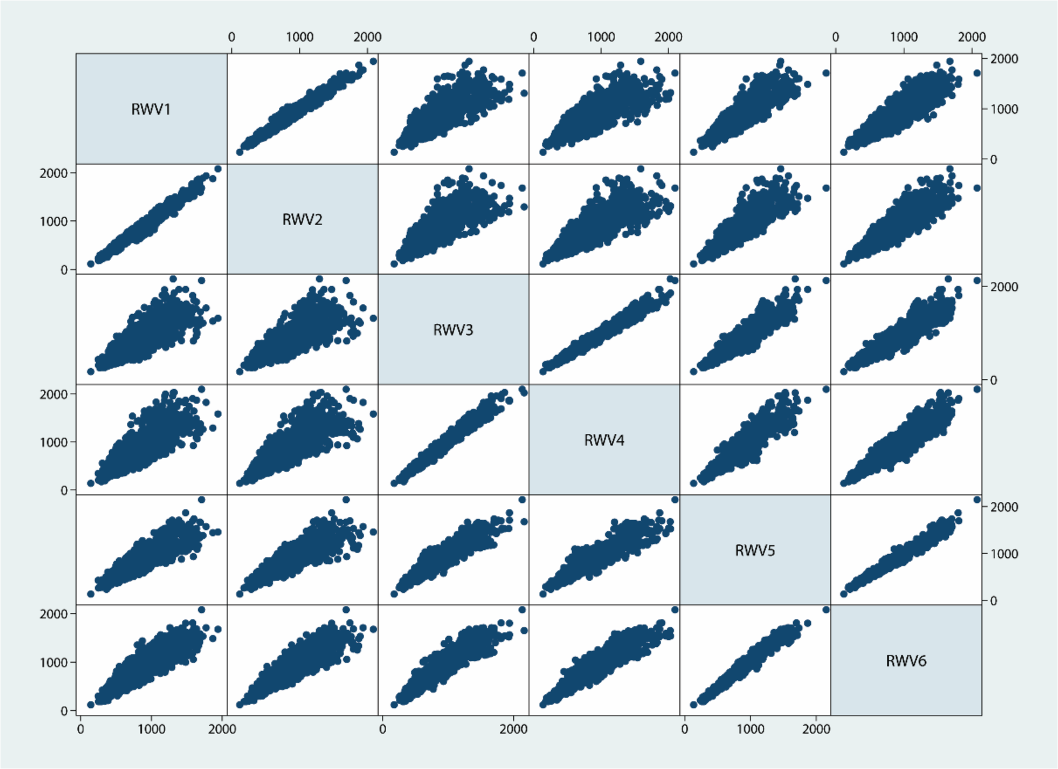


Figure 3-6. Scatterplot Matrix for the Parameter Sensitivity Analysis

Case Study 1: Wuhan, China

Wuhan is the capital city of Hubei Province, China and has been considered a transportation hub and a center of industry, commerce, and education in Central China. The primary data used in the study are from its road network for the year 2000. Population and employment census data for the same time period also were obtained. Additionally, the sale prices of more than one hundred randomly located houses were collected between 2001 and 2005.

RWVs show various degrees of correlation with the other network measures (Table 3-3). Particularly, they have the strongest correlations with degree centrality, followed by PageRank indices. The choice of the weight variable exerts significant influences on both PageRank scores and RWVs. The RWVs and PageRank scores have a higher correlation if the same weight variable is applied.

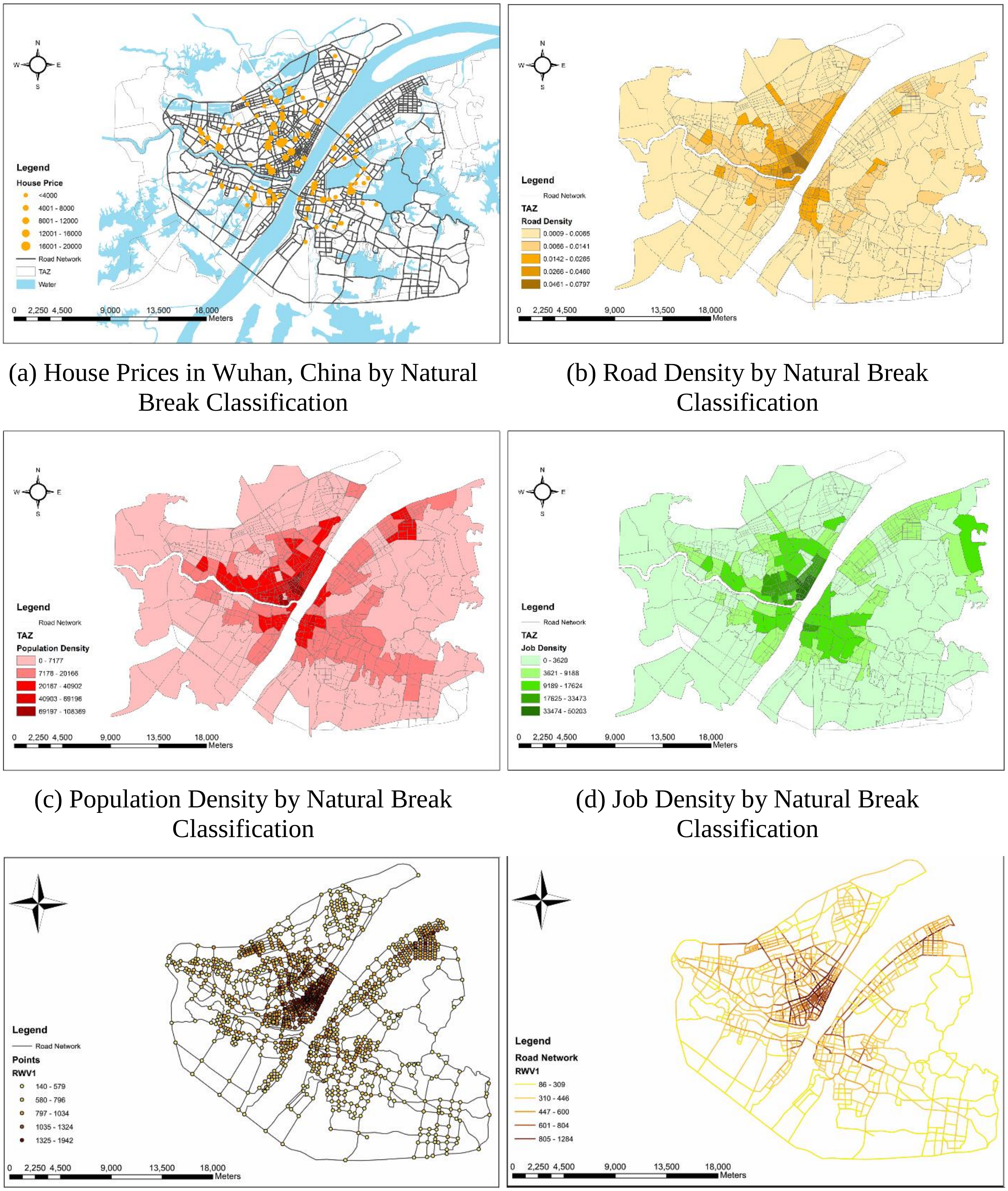
Table 3-3 Correlation of the RWV with Other Indices for Wuhan

Node level	**RWV1**	**RWV2**	**RWV3**	**RWV4**	**RWV5**	**RWV6**
PageRank	0.30*	0.29*	0.26*	0.25*	0.40*	0.38*
WPageRank1	0.48*	0.46*	0.09*	0.11*	0.17*	0.17*
WPageRank2	0.30*	0.28*	0.26*	0.25*	0.41*	0.38*
Degree of Centrality	0.46*	0.45*	0.44*	0.43*	0.58*	0.56*
Closeness	0.00	0.03	-0.02	0.02	-0.05	-0.05
Betweenness	0.01	0.03	-0.06*	-0.04	-0.07*	-0.05
Eigenvector	-0.13*	-0.14*	-0.16*	-0.16*	-0.16*	-0.17*
Link level	**RWV1**	**RWV2**	**RWV3**	**RWV4**	**RWV5**	**RWV6**
Connectivity	0.10*	0.12*	0.11*	0.12*	0.32*	0.31*
Mean Depth	-0.05*	-0.07*	-0.03	-0.07*	0.00	0.01
Global Integration	0.05*	0.07*	0.02	0.06*	-0.01	-0.01
Local Integration	0.15*	0.17*	0.16*	0.18*	0.37*	0.37*
Total Depth	-0.05*	-0.07*	-0.03	-0.07*	0.00	0.01
Local Depth	0.29*	0.31*	0.29*	0.31*	0.48*	0.49*

* denotes that the correlation is statistically significant at the 95% confidence level

For PageRank,WPageRank1 is weighted by width, and WPageRank2 is weighted by connectivity

Figure 3-7 (a) shows the geographic distribution of house sale prices from the survey data. It reveals that house prices in Wuhan are highest in central urban areas, and decrease gradually in the direction toward the outskirts. However, despite the general trend, housing price varies across space in a way that cannot simply be explained by distance from the urban center. Panels (b) through (f) in Figure 3-7 display the spatial distributions of related attributes.



(a) House Prices in Wuhan, China by Natural Break Classification

(b) Road Density by Natural Break Classification

(c) Population Density by Natural Break Classification

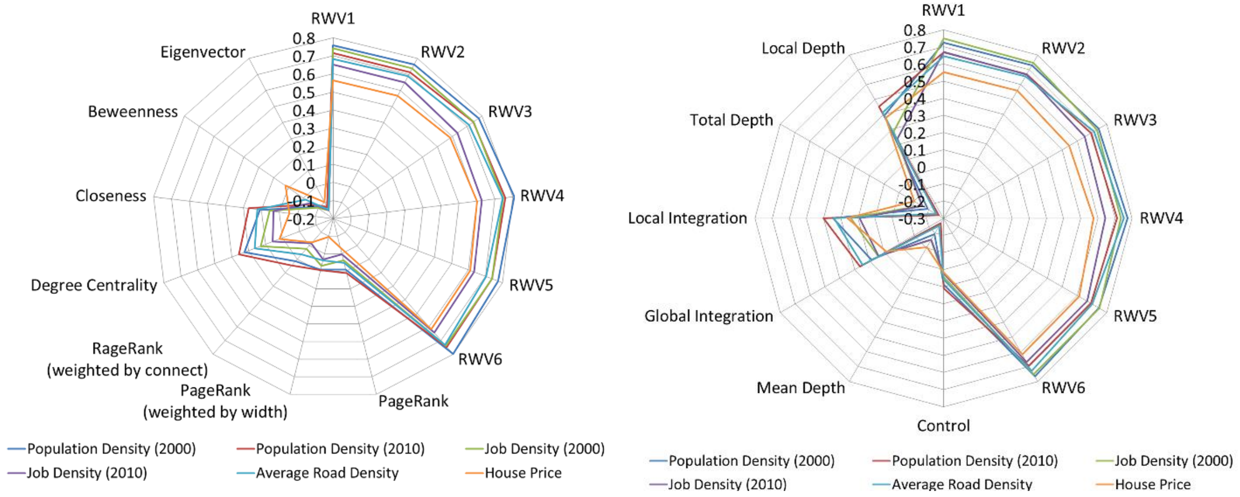
(d) Job Density by Natural Break Classification

(e) RWV1 of Nodes

(f) RWV1 of Links

Figure 3-7. Spatial Distributions of Selected Socioeconomic and Network Attributes in Wuhan, China

Figure 3-8 depicts correlations among all the network indices and the social-economic variables at the same aggregation level. In the radar charts, each color represents one socioeconomic variable. A dot of that color on each radius indicates the correlation between the variable and the corresponding network measure that is labeled for the radius. Because population and job densities are measured at the TAZ level, all of the network measures were also aggregated as averages in the TAZs for consistency. The Figure shows that the RWVs are highly correlated with population density, job density, and road densities, with much stronger coefficients than all other network measures. Among the six RWVs, all of those with a normal distribution of trip lengths (RWV2,4,6) perform slightly, but consistently, better than their counterparts with a power law distribution.



(a) Correlations Based on Nodal Indices (b) Correlations Based on Link Indices

Figure 3-8. Correlations Between Network Indices and Socioeconomic Variables at TAZ Level in Wuhan

To test the predictive power of RWVs as a surrogate for urban form characteristics, housing price is estimated using multivariate regression. The reason for this choice is that housing price has been a uniquely important socioeconomic indicator in Chinese cities, and is closely associated with not only land values, but also with other urban form characteristics.

Previous studies indicate that house prices in China are significantly influenced by both the land supply policy (Yu 2010) and locations of houses. The regression analysis was conducted separately for node RWV and link RWV measures. The general trends for these two measures are very consistent, although the node RWV yields slightly better results.

Twelve regression models are specified to estimate the house sale price in Wuhan, and results are compared. All variables used in the regression, including population density, job density, road density, RWV, and centrality indices, were recalibrated as averages in buffers of 500 meters around all the sample house locations. The reasons are twofold. First, it is to ensure consistent spatial units among all variables in the analysis. More importantly, the second reason is to test the penetration of influence of network components into its proximal areas. The spatial characteristics of the nearby nodes and links of the network determine the accessibility and other characteristics of the proximal areas where the houses are located.

Table 3-4 summarizes the specifications and results of all twelve models. The first two models are the base model which predict house sale price without RWVs. These two models are selected best models with a stepwise regression process that considers candidate predictor variables including population density, job density, road density, and the year of transaction. Because population density and job density are very highly correlated, the inclusion of both in a model does not produce better results. So they did not both appear in the two selected best model. As suggested by the statistics in Table 4, the problem of multicollinearity in Model-1 is of concern because the variance inflation factor (VIF) is larger than 5. This problem was solved in Model-2 by dropping road density at the cost of reduced explanation power of the model.

The next six models (Model 3-8) substitute RWV for job density. All of them exhibit better explanatory power and better quality, evidenced by the higher R-square scores and lower AIC values. Moreover, all of the six models are free of the multicollinearity problem. This outcome probably is because the RWVs are embedded with a richer set of information, including topological and physical characteristics of the local nodes in the road network, which are particularly important to house prices in the Chinese housing market. Among the six models, Model-8 is the best because it gives the highest adjusted R-square and the lowest AIC values. The corresponding RWV (RWV6) uses connectivity as weight, and adopts normal distribution for the simulation of trip lengths. However, the models using RWV5 and RWV2 also give highly comparable results. Those RWVs have various parameter choices, including a power law for trip length distribution, and the road width for weight. This suggests that the results are not very sensitive to the two critical parameters.

Four additional models (Models 9-12) were calibrated to test the performance of other traditional measures including betweenness, degree centrality, and the two weighted PageRank (WPR) measures. Betweenness and degree are chosen because they have higher correlations with housing price than any of the other traditional network measures (see Figure 3-8). The two WPRs are chosen because they have the highest correlations with the RWV (see Table 3-3). Unfortunately, all of them, except betweenness, are statically insignificant in the respective models (Table 3-4). Therefore, the explanatory power of Models 10, 11, and 12 are merely the same as the base model because they cannot receive any contribution from the network measures. Betweenness is found to be a significant predictor in Model-9, which helps to improve the model. However, with a lower R-square and a higher AIC score, Model-9 still is inferior to all of the RWV models in comparison. These consistent results suggest that RWV is a better network index for the substitution and estimation of selected socioeconomic variables in spatial analysis. In addition, to examine the effect of spatial autocorrelation in the study, a spatial lag model and a spatial error model were constructed. No significant spatial autocorrelation was found, and these spatial autoregressive models did not improve results.

Table 3-4 Regression Results for House price (101 Observations)

		Std. Coef.	P	VIF	R-square	Adj. R Square	AIC
Model 1	Average Job Density	0.11	0.51	5.26	0.48	0.47	1851.85
	Average Road Density	0.44	0.01	5.17			
	Transaction Year	0.36	0.00	1.04			
Model 2	Average Job Density	0.50	0.00	1.04	0.45	0.43	1856.76
	Transaction Year	0.35	0.00	0.00			
Model 3	Average Road Density	0.37	0.00	1.47	0.54	0.53	1838.91
	Transaction Year	0.35	0.00	1.46			
	Average RWV1	0.31	0.00	1.03			
Model 4	Average Road Density	0.35	0.00	1.50	0.55	0.54	1837.23
	Transaction Year	0.36	0.00	1.49			
	Average RWV2	0.33	0.00	1.02			
Model 5	Average Road Density	0.34	0.00	1.99	0.52	0.51	1844.01
	Transaction Year	0.34	0.00	1.97			

	Average RWV3	0.29	0.01	1.04			
Model 6	Average Road Density	0.31	0.00	2.01	0.53	0.52	1841.34
	Transaction Year	0.35	0.00	1.99			
	Average RWV4	0.33	0.00	1.03			
Model 7	Average Road Density	0.30	0.00	1.73	0.56	0.54	1836.50
	Transaction Year	0.35	0.00	1.73			
	Average RWV5	0.36	0.00	1.03			
Model 8	Average Road Density	0.28	0.00	1.78	0.56	0.55	1834.32
	Transaction Year	0.36	0.00	1.77			
	Average RWV6	0.39	0.00	1.02			
Model 9	Average Road Density	0.54	0.00	1.03	0.51	0.49	1846.17
	Transaction Year	0.38	0.00	1.03			
	Average Betweenness	0.18	0.02	1.01			
Model 10	Average Road Density	0.51	0.00	1.08	0.49	0.48	1849.75
	Transaction Year	0.39	0.00	1.08			
	Average Degree Centrality	0.12	0.12	1.06			
Model 11	Average Road Density	0.53	0.00	1.04	0.48	0.46	1852.30
	Year of Transaction	0.37	0.00	1.03			

	Average WPG (width)	0.00	1.00	1.03			
Model 12	Average Road Density	0.53	0.00	1.07	0.48	0.47	1851.46
	Year of Transaction	0.38	0.00	1.05			
	Average WPG (connect)	0.07	0.37	1.02			

Note: highlighted variables are NOT statistically significant at the 0.05 level

Case study 2: Atlanta, GA, the United States (U.S.)

The Atlanta metropolitan area is one of the top ten most populous areas in the U.S., according to the 2010 census data. It also is a world city with important roles in the global economic system. The road network data in the study are based on 2010 TIGER road GIS data of Atlanta from the U.S. census. In consideration of data volume and computational load, the dataset included all highways, primary roads, as well as secondary roads, while other local and neighborhood roads were excluded. For socioeconomic data, this study obtained the American Community Survey (ACS) 5-Year estimates for time period 2007-2011 from the U.S. census. All of the RWVs and other network indices are aggregated at the census tract level in order to be consistent with the spatial unit for the available socioeconomic data. For the simulation of trip lengths, the average trip length of 10 miles was used according to an online report (Atlanta Regional Commission 2004).

The correlations between the RWVs and other network measures are reported in Table 9 for measures at the nodal level and at the link level. The Table 3-5 shows that the RWVs have the highest correlations with degree centrality, closeness, and weighted PageRank for nodal measures, and with mean depth, total depth, and global integration for link measures.

Correlations between the RWV and centrality measures as well as space syntax indices are higher in Atlanta than those in Wuhan. Such differences may result from the distinct spatial structures of road networks in the two cities.

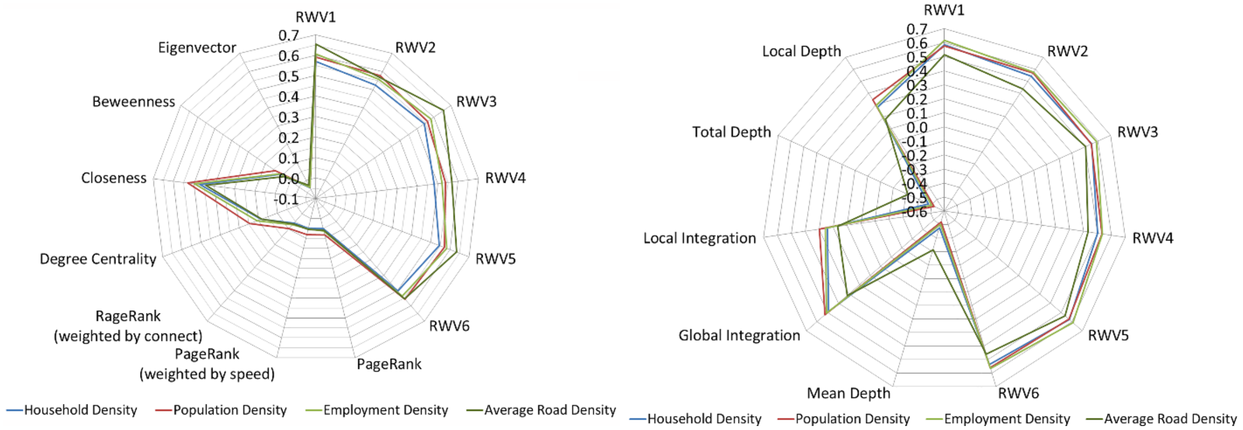
Table 3-5 Correlation of the RWV with Other Indices for Atlanta

Node level	**RWV1**	**RWV2**	**RWV3**	**RWV4**	**RWV5**	**RWV6**
PageRank	0.41*	0.44*	0.38*	0.41*	0.47*	0.51*
WPageRank (weighted by speed)	0.46*	0.49*	0.41*	0.45*	0.49*	0.52*
WPageRank (weighted by connectivity)	0.45*	0.48*	0.41*	0.45*	0.51*	0.55*
Load	0.14*	0.24*	0.13*	0.22*	0.16*	0.26*
Degree of Centrality	0.61*	0.65*	0.56*	0.61*	0.66*	0.70*
Closeness	0.51*	0.60*	0.49*	0.55*	0.50*	0.57*
Betweenness	0.14*	0.25*	0.14*	0.22*	0.17*	0.26*
Eigenvector	0.12*	0.14*	0.09*	0.14*	0.10*	0.13*
Link level	**RWV1**	**RWV2**	**RWV3**	**RWV4**	**RWV5**	**RWV6**
Connectivity	0.13*	0.15*	0.13*	0.14*	0.28*	0.29*
Mean Depth	-0.50*	-0.60*	-0.46*	-0.52*	-0.52*	-0.60*
Global Integration	0.48*	0.59*	0.44*	0.51*	0.51*	0.60*
Local Integration	0.15*	0.17*	0.15*	0.16*	0.29*	0.31*
Total Depth	-0.50*	-0.59*	-0.46*	-0.52*	-0.52*	-0.60*
Local Depth	0.29*	0.31*	0.27*	0.28*	0.39*	0.40*

Note: * indicates that a correlation is statistically significant at the 0.05 level

Figure 3-9 depicts correlations between the network measures and socioeconomic variables. Overall, all six types of RWVs have much higher correlations with the four socioeconomic variables than the traditional network measures do. The immediate second is closeness, which also has relatively high correlations. The result echoes findings from previous studies (Wang, Antipova, and Porta 2011) that the closeness centrality is positively correlated with population and employment densities. However, the newly developed RWV index clearly has a competitive edge over closeness.

Comparing the six RWVs among themselves, all three types with a power law distribution of trip lengths (RWV1, RWV3, and RWV5) have slightly but consistently higher correlations with the socioeconomic attributes than their counterparts based on a normal distribution. All four socioeconomic variables, namely population density, employment density, road density, and household density, consistently have significant correlations with all RWVs.



(a) Correlations Based on Node Indices (b) Correlations Based on Link Indices

Figure 3-9 Correlations between Network Indices and Socioeconomic Variables at the TAZ Level in Atlanta

Although the findings about strong correlations between RWVs and socioeconomic attributes are consistent between the two case studies, housing price in Atlanta is not found to be significantly associated with RWVs. It is also found that even population and job densities are not good predictors of housing price in Atlanta. This difference may result from differences in the two housing markets, in which property laws, people’s travel behavior, transportation systems, as well as economic and political contexts, are all different.

Discussion

The cases study results support the contention that the RWV is superior to the traditional network measures for ranking spatial characteristics of network components. First, in both studies, all RWVs consistently have higher correlations with all tested socioeconomic variables than other network measures do. In addition, the regression model for Wuhan's housing prices indicates that the RWV can be an excellent predictive factor, because adding it to the base model significantly improves the quality of the regression model. The RWV can be an excellent substitute for population and job densities in spatial analysis and modeling. In fact, by substituting the RWV for job density, the new models are significantly better than the original ones, evidenced by higher R-square, lower AIC and resolving the multicollinearity problem in the case study of Wuhan. In comparison, most other traditional network measures cannot make any contribution to the base model. The only exception is the betweenness network measure. However, all RWVs outperform betweenness with regard to the performances of respective models. The superiority of the RWV over other traditional network measures has its theoretical underpinning. As argued in the first section, a representative network measure should consider both topological properties and physical characteristics of network components. Traditional measures almost exclusively rely on topological properties only, while the RWV algorithm incorporates both topological structure and physical characteristics. Other less significant factors also may contribute to the better performance of the RWV. Compared with other network measures, the RWV algorithm implicitly incorporates road density for a road network measures by limiting the walking length in its calculation. In an urban area, road density is closely related to urban development intensity and economic prosperity.

The second finding is that the RWV algorithm is not extremely sensitive to the two critical parameters, as long as the parameters are chosen from the set of reasonable options. All RWVs under various parameter settings consistently outperform other network measures in the case studies, suggesting that none of the available parameter settings would dramatically affect the general pattern of RWVs. That said, moderate variations are observed among RWVs from different parameter settings. For instance, normal distribution of trip lengths was found to be slightly but consistently better than power law distribution in the case study of Wuhan. However, a power law distribution was found to be a better option for the case study of Atlanta. Such a difference in trip length distributions may result from the differences in prevailing travel modes and urban forms in the two cities. Therefore, careful choice of parameters is highly recommended, and should take into account the social, political, and economic context of a study area.

Third, the RWV is not always a good predictor for all socioeconomic variables. The result really depends on the underpinning spatial processes behind the variables to be estimated. For instance, housing price was found to be predictable by the said variables in Wuhan but not in Atlanta. This is probably because different mechanisms governing the two housing markets. In the Chinese housing market, house values are largely determined by its accessibility to socioeconomic resources. However, in the U.S. housing market, where accessibility is generally good in cities and the prevailing travel mode is private car, the network related attributes become less important, and local neighborhood characteristics may play more important roles. The RWV is expected to be a good predictor only when a socioeconomic variable to be predicted results from spatial processes that are closely related to its network characteristics.

Conclusion

This study proposes a new network measure, called the random walk value (RWV), for ranking spatial characteristics of network components and associated proximal areas. In general, the RWV is consistent with some of the most widely used network measures. Two case studies reveal high correlations between the RWV and several popular network indices, such as PageRank, closeness, and connectivity. However, these case studies also reveal that RWV outperforms all traditional network indices in terms of its correlation with important socioeconomic variables, and in its ability to predict other variables. Specifically, both cases studies confirm that the RWV tends to be highly correlated with population and job densities, and may serve as a powerful substitute for them in spatial analysis models. This finding is significant for many studies when population and job distribution data are not available, or for studies that predict future scenarios.

The RWV is more broadly grounded in both topological network structure and physical road characteristics. The algorithm considers not only spatial structural characteristics, but also physical characteristics of network components. Implicit spatial characteristics, such as natural streets, also can be incorporated in the algorithm. The more comprehensive consideration of network properties provides a competitive advantage for it, compared with traditional network indices that are primarily based on topological properties only.

Although the newly developed network index is innovative and powerful, the RWV algorithm is still young, and can be further refined. Many potential research avenues exist for further development and applications. One is discussed here as a starting point. The current algorithm for calculating the RWV assumes absolute independence for each simulated trip. In other words, the algorithm does not consider the previously generated trips that have already been assigned to road segments. However, in reality, a dynamic relationship exists between trip volume and travel speed on a road segment. Traffic congestion may occur when traffic volume exceeds road capacity. Research is needed to revise the algorithm in consideration of such a dynamic relationship.

Acknowledgement

The research is supported by a special fund of the Key Laboratory of Eco Planning & Green Building, Ministry of Education (Tsinghua University), China. The authors are grateful to Dr. Bin Jiang for his illuminating discussions, to anonymous reviewers for their constructive comments and suggestions, and to Dr. Daniel Griffith for his valuable and meticulous edits, all of which have greatly helped to improve an earlier version of this manuscript. The views and insufficiencies of the paper remain the authors' responsibility.

References

Adams, P. 1998. Network topologies and virtual place. *Annals of the Association of American Geographers* 88 (1):88.

Aderamo, A. J., and S. A. Magaji. 2010. Rural Transportation and the Distribution of Public Facilities in Nigeria: A Case of Edu Local Government Area of Kwara State. *Journal of Human Ecology* 29 (3):171–179.

Adnan, M., A. D. Singleton, P. A. Longley, and C. Brunsdon. 2010. Towards Real-Time Geodemographics: Clustering Algorithm Performance for Large Multidimensional Spatial Databases. *Transactions in GIS* 14 (3):283–297.

Anderson, J. 1970. Time-Budgets and Human Geography. *Area* 2 (1):50–51.

Atlanta Regional Commission. 2004. Commute Options. *www.atlantaregional.com*. http://www.atlantaregional.com/transportation/commute-options (last accessed 7 January 2014).

Backstrom, L., E. Sun, and C. Marlow. 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. New York, NY, USA: ACM http://doi.acm.org/10.1145/1772690.1772698 (last accessed 10 December 2014).

Bahir, E., and A. Peled. 2013. Identifying and Tracking Major Events Using Geo-Social Networks. *SOCIAL SCIENCE COMPUTER REVIEW* 31 (4):458–470.

Barabasi, A.-L., and R. Albert. 1999. Emergence of Scaling in Random Networks. *Science* 286 (5439):509.

Baran, P. K., D. A. Rodr ǵuez, and A. J. Khattak. 2008. Space Syntax and Walking in a New Urbanist and Suburban Neighbourhoods. *Journal of Urban Design* 13 (1):5–28.

Batty, M. 1997. Virtual geography. Time and Space Geographic Perspectives on the Future., eds. M. Batty and S. Cole, 337–352. Great Britain, BUTTERWORTH-HEINEMANN.

Behrens, R. B., and L. A. Kane. 2004. Road capacity change and its impact on traffic in congested networks: evidence and implications. *Development Southern Africa* 21 (4):587–602.

Benenson, I., and P. M. Torrens. 2004. *Geosimulation: automata-based modelling of urban phenomena*. Hoboken, NJ: John Wiley & Sons.

Blanchard, P., and D. Volchenkov. 2008. Exploring Urban Environments By Random Walks. *arXiv:0801.3216v1 [physics.soc-ph]* 1021:183–203.

———. 2010. Random Walks Estimate Land Value. *arXiv:1003.0384 [physics]* :1–15.

Bono, F., E. Guti érrez, and K. Poljansek. 2010. Road traffic: A case study of flow and path-dependency in weighted directed networks. *Physica A: Statistical Mechanics and its Applications* 389 (22):5287–5297.

Brandes, U., and T. Erlebach. 2005. *Network analysis methodological foundations*. NewYork: Springer Berlin Heidelberg.

Brin, S., and L. Page. 2012. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer Networks* 56 (18):3825 – 3833.

Butts, C. T., R. M. Acton, J. R. Hipp, and N. N. Nagle. 2012. Geographical variability and network structure. *Social Networks* 34 (1):82–100.

Chan, S. H. Y., R. V. Donner, and S. L ämmer. 2011. Urban road networks — spatial networks with universal geometric features? *The European Physical Journal B - Condensed Matter and Complex Systems* 84 (4):563–577.

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

Chen, J., S.-L. Shaw, H. Yu, F. Lu, Y. Chai, and Q. Jia. 2011. Exploratory data analysis of activity diary data: a space–time GIS approach. *Journal of Transport Geography* 19 (3):394–404.

Chen, Q., and S. Chen. 2007. A highly clustered scale-free network evolved by random walking. *Physica A: Statistical Mechanics and its Applications* 383 (2):773–781.

Cho, E., S. A. Myers, and J. Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1082–1090. ACM.

Clauset, A., C. R. Shalizi, and M. E. Newman. 2009. Power-law distributions in empirical data. *SIAM review* 51 (4):661–703.

Crandall, D. J., L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. 2010. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences* 107 (52):22436–22441.

Cranshaw, J., R. Schwartz, J. I. Hong, and N. M. Sadeh. 2012. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In *ICWSM*. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/download/4682/4967 (last accessed 6 April 2014).

Croitoru, A., A. Crooks, J. Radzikowski, and A. Stefanidis. 2013. Geosocial gauge: a system prototype for knowledge discovery from social media. *International Journal of Geographical Information Science* 27 (12):2483 – 2508.

Croxford, B., A. Penn, and B. Hillier. 1996. Spatial distribution of urban pollution: civilizing urban traffic. *Science of The Total Environment* 189:3–9.

Dodgshon, R. A. 2008. GEOGRAPHY’S PLACE IN TIME. *Geografiska Annaler Series B: Human Geography* 90 (1):1–15.

Dragicevic, S., and D. J. Marceau. 2000. A fuzzy set approach for modelling time in GIS. *International Journal of Geographical Information Science* 14 (3):225–245.

Eagle, N., A. Pentland, and D. Lazer. 2009. Inferring friendship network structure by using mobile phone data. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* 106 (36):15274 – 15278.

Eagle, N., and A. Sandy Pentland. 2006. Reality mining: sensing complex social systems. *Personal & Ubiquitous Computing* 10 (4):255–268.

Easley, D., and J. Kleinberg. 2010. *Networks, Crowds, and Markets : Reasoning About a Highly Connected World*. Cambridge University Press.

Emch, M., E. D. Root, S. Giebultowicz, M. Ali, C. Perez-Heydrich, and M. Yunus. 2012. Integration of Spatial and Social Network Analysis in Disease Transmission Studies. *Annals of the Association of American Geographers* 102 (5):1004–1015.

ESRI. 2013. GeoEvent Processor Extension. *ArcGIS for Server*. http://www.esri.com/software/arcgis/arcgisserver/extensions/geoevent-extension (last accessed 18 January 2014).

Facebook. 2013. Facebook Developers. *Facebook Developers*. https://developers.facebook.com/ (last accessed 25 December 2013).

Geofeedia. 2014. Search & Monitor Social Media by Location. *Geofeedia*. http://corp.geofeedia.com/ (last accessed 18 January 2014).

Gidofalvi, G., and T. B. Pedersen. 2005. Spatio-temporal Rule Mining: Issues and Techniques. In *Data Warehousing and Knowledge Discovery*, Data warehousing and knowledge discovery; DaWaK 2005., eds. J. M. Morvan and J. M. Morvan, 275–284. Berlin, [Great Britain], Springer.

Goodchild, M. F. 2010. Twenty years of progress: GIScience in 2010. *Journal of Spatial Information Science* (1):3.

Goodchild, M. F., M. Yuan, and T. J. Cova. 2007. Towards a general theory of geographic representation in GIS. *International Journal of Geographical Information Science* 21 (3):239–260.

Google. 2013. Google App Engine — Google Developers. *Google App Engine: Platform as a Service*. https://developers.google.com/appengine/?csw=1 (last accessed 24 January 2014).

Hagberg, A. A., D. A. Schult, and P. J. Swart. 2008. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, 11–15. Pasadena, CA USA.

H ägerstrand, T. 1970. What About People in Regional Science? *Papers in Regional Science* 24 (1):7.

Hanneman, R. A., and M. Riddle. 2005. *Introduction to social network methods*. University of California Riverside.

Hillier, B. 1999. The common language of space A way of looking at the social, economic and environmental functioning of cities on a common basis. *Journal of Environmental Sciences (IOS Press)* 11 (3):344–349.

Hipp, J. R., R. W. Faris, and A. Boessen. 2012. Measuring “neighborhood”: Constructing network neighborhoods. *Social Networks* 34 (1):128–140.

Horner, M., B. Zook, and J. Downs. 2012. Where were you? Development of a time-geographic approach for activity destination re-construction. *COMPUTERS ENVIRONMENT AND URBAN SYSTEMS* 36 (6):488–499.

Hornsby, K., and M. J. Egenhofer. 2000. Identity-based change: a foundation for spatio-temporal knowledge representation. *International Journal of Geographical Information Science* 14 (3):207–224.

Humphreys, L. 2007. Mobile Social Networks and Social Practice: A Case Study of Dodgeball. *Journal of Computer-Mediated Communication* 13 (1):341–360.

Jiang, B. 2009. Ranking Spaces for Predicting Human Movement in an Urban Environment. *International Journal of Geographical Information Science* 23 (7):823–837.

Jiang, B., and C. Claramunt. 2002. Integration of Space Syntax into GIS: New Perspectives for Urban Morphology. *Transactions in GIS* 6 (3):295–309.

Jiang, B., and C. Claramunt. 2004. A Structural Approach to the Model Generalization of an Urban Street Network*. *GeoInformatica* 8 (2):157–171.

Jiang, B., and T. Jia. 2011. Agent-based simulation of human movement shaped by the underlying street structure. *International Journal of Geographical Information Science* 25 (1):51–64.

Jiang, B., and Y. Miao. 2014. The Evolution of Natural Cities from the Perspective of Location-Based Social Media. *arXiv:1401.6756 [nlin, physics:physics]* :1–14.

Ji, S. Y., E. Niklas, and S. Lee. 2010. TimeMatrix: Analyzing Temporal Social Networks Using Interactive Matrix-Based Visualizations. *International Journal of Human-Computer Interaction* 26 (11/12):1031–1051.

Kim, H.-M., and M.-P. Kwan. 2003. Space-time accessibility measures: A geocomputational algorithm with a focus on the feasible opportunity set and possible activity duration. *Journal of Geographical Systems* 5 (1):71.

Kleinfeld, J. S. 2002. The Small World Problem. *Society* 39 (2):61–66.

Knox, E. G., and M. S. Bartlett. 1964. The Detection of Space-Time Interactions. *Applied Statistics* 13 (1):25.

Kuijpers, B., H. J. Miller, T. Neutens, and W. Othman. 2010. Anchor uncertainty and space-time prisms on road networks. *International Journal of Geographical Information Science* 24 (8):1223.

Kwan, M.-P. 1998. Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-based Framework. *Geographical Analysis* 30 (3):191–216.

———. 2002. Feminist Visualization: Re-Envisioning GIS as a Method in Feminist Geographic Research. *Annals of the Association of American Geographers* (4):645.

———. 2004. GIS Methods in Time-Geographic Research: Geocomputation and Geovisualization of Human Activity Patterns. *Geografiska Annaler Series B: Human Geography* 86 (4):267.

———. 2007. Mobile Communications, Social Networks, and Urban Travel: Hypertext as a New Metaphor for Conceptualizing Spatial Interaction. *Professional Geographer* 59 (4):434–446.

———. 2010. A Century of Method-Oriented Scholarship in the Annals. *Annals of the Association of American Geographers* 100 (5):1060–1075.

Lampos, V., and N. Cristianini. 2010. Tracking the flu pandemic by monitoring the social web. *2010 2nd International Workshop on Cognitive Information Processing (CIP)* :411.

Leng, S., L. Zhang, H. Fu, and J. Yang. 2007. Mobility analysis of mobile hosts with random walking in ad hoc networks. *Computer Networks* 51 (10):2514–2528.

Lewis, K., J. Kaufman, G. A. Marco, W. B. Andreas, and C. A. Nicholas. 2008. Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks* 30:330–342.

Liben-Nowell, D., and J. Kleinberg. 2007. The link-prediction problem for social networks. *Journal of the American Society for Information Science & Technology* 58 (7):1019–1031.

Licoppe, C., and Y. Inada. 2008. Geolocalized Technologies, Location-Aware Communities, and Personal Territories: The Mogi Case. *Journal of Urban Technology* 15 (3):5–24.

Li, J., and Q. Guo. 2003. Quantitative research of urban spatial morphology based on syntactic analysis. *Enginering Journal of Wuhan University* 36 (2):69–73.

Lindsey, G., J. Wilson, E. Rubchinskaya, J. Yang, and Y. Han. 2007. Estimating urban trail traffic: Methods for existing and proposed trails. *Landscape and Urban Planning* 81:299–315.

Liu, X., and B. Jiang. 2012. Defining and Generating Axial Lines from Street Center Lines for better Understanding of Urban Morphologies. *International Journal of Geographical Information Science* 26 (8):1521–1532.

Liu, Y., Z. Sui, C. Kang, and Y. Gao. 2014. Uncovering Patterns of Inter-Urban Trip and Spatial Interaction from Social Media Check-In Data. *PLoS ONE* 9 (1):1–11.

Li, Y., H. Gao, M. Yang, W. Guan, H. Ma, W. Qian, Z. Cao, and X. Yang. 2013. What are Chinese Talking about in Hot Weibos? *arXiv preprint arXiv:1304.4682*. http://arxiv.org/abs/1304.4682 (last accessed 26 December 2013).

Mennis, J. 2010. Multidimensional Map Algebra: Design and Implementation of a Spatio-Temporal GIS Processing Language. *Transactions in GIS* 14 (1):1–21.

Mennis, J. L. P., Donna J.Qian, Liujian. 2000. A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science* 14 (6):501.

Merriman, P. 2012. Human geography without time-space. *Transactions of the Institute of British Geographers* 37 (1):13–27.

Mika, P. 2005. Flink: Semantic Web technology for the extraction and analysis of social networks. *Web Semantics: Science, Services and Agents on the World Wide Web* 3 (2–3):211–223.

Milgram, S., L. Mann, S. Harter, and B. Kass. 1965. THE LOST-LETTER TECHNIQUE: A TOOL OF SOCIAL RESEARCH. *Public Opinion Quarterly* 29 (3):437.

Miller, H. J. 1999. Measuring space-time accessibility benefits within transportation networks: basic theory and computational procedures. *Geographical analysis* 31 (1):1–26.

———. 2003. What about people in geographic information science? *Computers, Environment and Urban Systems* 27:447.

Miller, H. J., and S. A. Bridwell. 2009. A Field-Based Theory for Time Geography. *Annals of the Association of American Geographers* 99 (1):49–75.

Miyagawa, M. 2009. Optimal hierarchical system of a grid road network. *Annals of Operations Research* 172 (1):349–361.

Naaman, M., J. Boase, and C.-H. Lai. 2010. Is it really about me?: message content in social awareness streams. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 189–192. ACM http://dl.acm.org/citation.cfm?id=1718953 (last accessed 11 December 2014).

Neutens, T., N. Weghe, F. Witlox, and P. Maeyer. 2008. A three-dimensional network-based space–time prism. *Journal of Geographical Systems* 10 (1):89–107.

Neutens, T., F. Witlox, N. Van De Weghe, and P. H. De Maeyer. 2007. Space-time opportunities for multiple agents: a constraint-based approach. *International Journal of Geographical Information Science* 21 (10):1061–1076.

Newman, M. E. 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary physics* 46 (5):323–351.

Newman, M. E. J. 2003. The Structure and Function of Complex Networks. *SIAM Review* (2):167.

———. 2006. Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences of the United States of America* (23):8577.

Ozbay, K., D. Ozmen, and J. Berechman. 2006. Modeling and Analysis of the Link between Accessibility and Employment Growth. *Journal of Transportation Engineering* 132 (5):385–393.

Padmanabhan, A., S. Wang, G. Cao, M. Hwang, Z. Zhang, Y. Gao, K. Soltani, and Y. Liu. 2014. FluMapper: A cyberGIS application for interactive analysis of massive location-based social media. *Concurrency and Computation: Practice and Experience* 26 (13):2253–2265.

Pearson, K. 1905. The Problem of the Random Walk. *Nature* 72 (1865):294.

Peuquet, D., and N. Duan. 1995. An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographical Information Systems* 9 (1):7–24.

Peuquet, D. J. 1994. It's about Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems. *Annals of the Association of American Geographers* (3):441.

———. 2001. Making Space for Time: Issues in Space-Time Data Representation. *GeoInformatica* 5 (1):11.

Pickles, J. C. 1995. Representations in an Electronic Age: Geography, GIS, and Democracy. In *Critical Geographies: A Collection of Readings*, eds. H. Bauder and S. E.-D. Mauro, 637–663. British Columbia, Canada.: Praxis (e)Press.

Pons, P., and M. Latapy. 2005. Computing communities in large networks using random walks. In *20th International Symposium on Computer and Information Sciences,* 284–293. Istanbul, Turkey: Springer.

Porras, R., T. Takeshita, M. Ikezoe, and R. Araya. 2002. A Study on the Pedestrian Space applying Space Syntax and the Segment Unit. *Journal of Asian Architecture and Building Engineering* 1 (1):197–203.

Porta, S., V. Latora, F. Wang, S. Rueda, E. Strano, S. Scellato, A. Cardillo, E. Belli, F. Cardenas, B. Cormenzana, and L. Laura. 2012. Street Centrality and the Location of Economic Activities in Barcelona. *Urban Studies* 49 (7):1471–1488.

Pultar, E., T. J. Cova, Y. May, and M. F. Goodchild. 2010. EDGIS: a dynamic GIS based on space time points. *International Journal of Geographical Information Science* 24 (3):329–346.

Raubal, M., H. J. Miller, and S. Bridwell. 2004. User-Centred Time Geography for Location-Based Services. *Geografiska Annaler Series B: Human Geography* 86 (4):245–265.

Rey, S. J., and L. Anselin. 2007. PySAL: A Python Library of Spatial Analytical Methods. *The Review of Regional Studies* 37 (1):5–27.

Rodrigue, J.-P., C. Comtois, and B. Slack. 2013. *The Geography of Transport Systems* 3 edition. NewYork: Routledge.

Ronald, N., V. Dignum, C. Jonker, T. Arentze, and H. Timmermans. 2012. On the engineering of agent-based simulations of social activities with social networks. *Information and Software Technology* 54 (6):625–638.

Russell, M. A. 2013. *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. O'Reilly Media, Inc.

Scellato, S., A. Noulas, R. Lambiotte, and C. Mascolo. 2011. Socio-Spatial Properties of Online Location-Based Social Networks. *ICWSM* 11:329–336.

Schwanen, T., and M.-P. Kwan. 2012. Critical Space-Time Geographies: Guest Editorial. *Environment and Planning A* 44 (9):2043–2048.

Shaw, S.-L., and H. Yu. 2009. A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. *Journal of Transport Geography* 17 (2):141–149.

Shaw, S.-L., H. Yu, and L. S. Bombom. 2008. A Space-Time GIS Approach to Exploring Large Individual-based Spatiotemporal Datasets. *Transactions in GIS* 12 (4):425–441.

Short, M. B., M. R. D'Orsogna, V. B. Pasour, G. E. Tita, P. J. Brantingham, A. L. Bertozzi, and L. B. Chayes. 2008. A statistical model of criminal behavior. *Mathematical Models & Methods in Applied Sciences* 18:1249–1267.

De Smith, M. 2010. *Statistical Analysis Handbook: Concepts, Techniques, Tools*. www.statsref.com/HTML/.

Song, X., X. Wang, A. Li, and L. Zhang. 2011. Node Importance Evaluation Method for Highway Network of Urban Agglomeration. *Journal of Transportation Systems Engineering and Information Technology* 11 (2):84–90.

Sui, D., and M. Goodchild. 2011. The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science* 25 (11):1737.

Taaffe, E. J., H. L. Gauthier, and M. E. O'Kelly. 1996. *Geography of transportation*. Upper Saddle River, N.J.: Prentice Hall.

Takhteyev, Y., A. Gruzd, and B. Wellman. 2012. Geography of Twitter networks. *Social Networks* 34 (1):73–81.

Tavakoli, M., and S. Fakhraie. 2011. THEORY OF TIME GEOGRAPHY (EXPLANATION OF ITS APPLICABLE QUANTITIES IN PLANNING). *International Journal of Academic Research* 3 (3):655–662.

Thrift, N. 1996. *Spatial Formations*. SAGE Publications. http://books.google.com/books?id=B2hw8X-yigIC.

Thrift, N. J. 1996. *Spatial formations*. Sage.

Torrens, P., X. Li, and W. Griffin. 2011. Building Agent-Based Walking Models by Machine-Learning on Diverse Databases of Space-Time Trajectory Samples. *Transactions in GIS* 15:67.

Train, K. 2009. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Tsou, M.-H., and M. Leitner. 2013. Visualization of social media: seeing a mirage or a message? *Cartography and Geographic Information Science* 40 (2):55.

Turner, A. 2009. The Role of Angularity in Route Choice. In *Spatial Information Theory*, Lecture Notes in Computer Science., eds. K. S. Hornsby, C. Claramunt, M. Denis, and G. Ligozat, 489–504. Springer Berlin Heidelberg.

Twitter. 2013. Exploring the Twitter API. *Twitter Developers*. https://dev.twitter.com/console (last accessed 25 December 2013).

Vance, C., and R. Hedel. 2008. On the Link Between Urban Form and Automobile Use: Evidence from German Survey Data. *Land Economics* 84 (1):51–65.

Vasardani, M., S. Winter, and K.-F. Richter. 2013. Locating place names from place descriptions. *International Journal of Geographical Information Science* 27 (12):2509–2532.

Volchenkov, D., and P. Blanchard. 2007. Random walks along the streets and canals in compact cities: Spectral analysis, dynamical modularity, information, and statistical mechanics. *Physical Review E* 75 (2):026104.

Wang, D., and T. Cheng. 2001. A spatio-temporal data model for activity-based transport demand modelling. *International Journal of Geographical Information Science* 15 (6):561–585.

Wang, F., A. Antipova, and S. Porta. 2011. Street centrality and land use intensity in Baton Rouge, Louisiana. *Journal of Transport Geography* 19 (2):285–293.

Wang, J., X. Wu, Y. Bo, and J. Guo. 2011. Improved Method of Node Importance Evaluation Based on Node Contraction in Complex Networks. *Procedia Engineering* 15:1600–1604.

Wang, P., T. Hunter, A. M. Bayen, K. Schechtner, and M. C. Gonz ález. 2012. Understanding road usage patterns in urban areas. *Scientific Reports* 2 (1001):1–6.

Watling, D., D. Milne, and S. Clark. 2012. Network impacts of a road capacity reduction: Empirical analysis and model predictions. *Transportation Research Part A: Policy and Practice* 46 (1):167–189.

Wewal, J., D. Wilkie, P. Merrell, and M. C. Lin. 2010. Continuum Traffic Simulation. *Computer Graphics Forum* 29 (2):439–448.

Winter, S., and Z.-C. Yin. 2011. The elements of probabilistic time geography. *GeoInformatica* 15 (3):417–434.

Worboys, M. 2005. Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science* 19 (1):1–28.

Wright, D. J., M. F. Goodchild, and J. D. Proctor. 1997. GIS: Tool or Science? Demystifying the Persistent Ambiguity of GIS as "Tool" Versus "Science." *Annals of the Association of American Geographers* (2):346.

Xing, W., and A. Ghorbani. 2004. Weighted PageRank algorithm. In *Communication Networks and Services Research*, 305–314. IEEE.

Yang, S.-J. 2005. Exploring complex networks by walking on them. *Physical Review E* 71 (1):016107.

Yao, X. 2010. Modeling Cities as Spatio-Temporal Places. In *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, 311–328. Netherlands: Springer http://proxy-remote.galib.uga.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=76899530&site=eds-live.

Yuan, L., S. Chen, Y. Wang, Z. Yu, W. Luo, and G. L ü. 2010. CAUSTA: Clifford Algebra-based Unified Spatio-Temporal Analysis. *Transactions in GIS* 14:59–83.

Yuan, M. 1997. Use of knowledge acquisition to build wildfire representation in Geographical Information Systems. *International Journal of Geographical Information Science* 11 (8):723–746.

Yuan, M., A. Nara, and J. Bothwell. 2014. Space–time representation and analytics. *Annals of GIS* 20 (1):1–9.

Yu, H. 2006. Spatio-temporal GIS Design for Exploring Interactions of Human Activities. *Cartography and Geographic Information Science* 33 (1):3–19.

Yu, H. 2010. China's House Price: Affected by Economic Fundamentals or Real Estate Policy? *Frontiers of Economics in China* 5 (1):25–51.

Yu, H., and S.-L. Shaw. 2008. Exploring potential human activities in physical and virtual spaces: a spatio-temporal GIS approach. *International Journal of Geographical Information Science* 22 (4):409–430.

Zhang, H., and Z. Li. 2011. Weighted ego network for forming hierarchical structure of road networks. *International Journal of Geographical Information Science* 25 (2):255–272.

Zhang, S., and X. Yao. 2011. *Social-spatial structure of Beijing : a spatial-temporal analysis*. 2011.

CHAPTER 4 ANALYZING LOCATION-BASED SOCIAL MEDIA ACTIVITY IN SPATIAL-SOCIAL DIMENSION[3]

[3] Xuebin Wei and Xiaobai Yao. To be submitted to Professional Geographer

Abstract

Human activities include spatial, temporal and social components that should be comprehensively analyzed. However, the dynamics of social connections between human activities are over simplified, and the spatial characteristics of social connections are seldom investigated. Meanwhile, collecting social relationship data is time consuming and always ended up with incomplete population. With the availability of location-based social media that enables people to publish their social events with geographical positions, participants and time of events, this research has introduced an innovative methodology of extracting location-based social media activities and organizing those activities in a way that social connections and spatial locations of human activities can be integrated. The proposed method is able to visualize and measure the dynamics of human connections in spatial-social dimension, and identify spatial-social clusters of human activities.

Introduction

Social relationships are molded in a network structure where points represent individuals and edges represent social connection between those individuals (Hanneman and Riddle 2005). Such representations of human relationships are solely based on whether or not those individuals recognize each other, and generate topological measures of how individuals are embed in social networks. However, human activities involve in interacting with other people or visiting geographical places at different time periods. Those activities form social connections and occupy spatial-temporal positions. Social connections are contingent to surround environment which includes human neighborhoods and geographical spaces.

The current studies of human activities focus on only one or two aspects of those aspects, such as time-space prisms (Kwan 1998) that analyze trajectories of human activities in spatial-temporal dimension, or TimeMatrix (Ji, Niklas, and Lee 2010) that emphasizes changes of social connections over time. How the dynamics of human connections vary over space has been seldom investigated. Particularly, the characteristics of human connections are simplified as static features in traditional social network analysis (Hanneman and Riddle 2005), and their spatial characteristics are neglected.

In addition, collecting social relationship data between human beings has traditionally been a challenging endeavor that requires long hours of observation and interviews (Cranshaw et al. 2012). Enabled with GPS functions, many social media platforms allow people to constantly publish posts about human activities. Those social media posts include explicit participants and precise geographic locations of social events. Therefore, location-based social media has become new channels for observing human activities in spatial and social dimensions.

This research has innovatively extracted human activities from location-based social media data where social acquisitions and locations are constantly published by social media users (Russell 2013). By constructing social connections from the collected social media data, this research proposed a methodology that can integrate social connections among people with geographic locations of their activities. The proposed method is able to measure the dynamic of social connection in spatial-temporal dimension, and produce useful visualization and analysis measurements, such as the identification of spatial-social clusters.

Related Works

Constructing Social Networks from Human Activities

Social-spatial structure cannot be explained by examining geographic factor or social mechanisms only (Scellato et al. 2011). On the one hand, social relations can be inferred from human activities in spatial and temporal dimension. On the other hand, social relations constrain human activities geographically and temporally (Cheng et al. 2011). Therefore, many studies have examined social relations incorporating with geographical locations of human activities. The inference about social ties between people is based on the factor those people were in the same geographical locations at roughly the same time (Crandall et al. 2010). Eagle et al. collected communication information and location data form mobile phones, and have successfully inferred 95% of friendships based on the observational data alone (Eagle, Pentland, and Lazer 2009). It has also been found that the social relationship can explain 10%-30% of human movements (Cho, Myers, and Leskovec 2011). Zhang and Yao applied a spatial-temporal analysis to identify social-spatial structural changes in Beijing (Zhang and Yao 2011). Butts et al. performed the exploratory simulation to explore the potential implications of geographical variability for the structure of social networks (Butts et al. 2012). Hipp et al. measured neighborhood boundaries based on the density of social ties among adolescents (Hipp, Faris, and Boessen 2012). Emch et al. also found that simultaneous spatial and social network analysis can add to the understanding of disease transmissions (Emch et al. 2012).

Analyzing Social Dynamics from Social Media Data

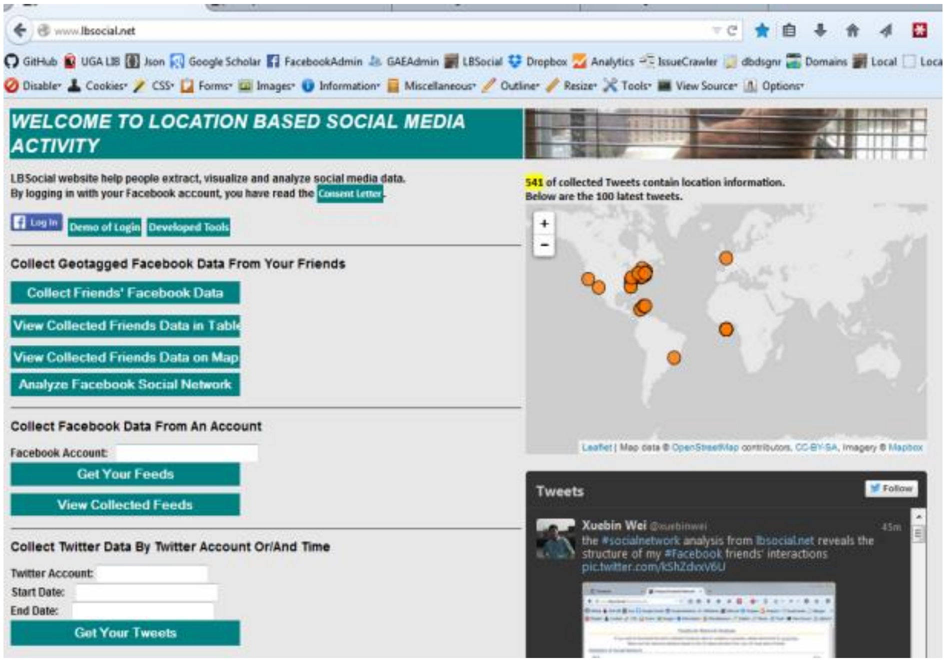
With the popularization of location-based social media data, social media feeds are becoming increasingly “geosocial” (Croitoru et al. 2013), meaning that social structure and its impact on human can be directly observed form social media (Cheng et al. 2011). Some scholars have studied human activities and interactions from different social media, such as Facebook (Backstrom, Sun, and Marlow 2010), Twitter (Cheng et al. 2011; Croitoru et al. 2013; Padmanabhan et al. 2014), Dodgeball (Humphreys 2007), Foursquare (Scellato et al. 2011), and Flickr and YouTube (Croitoru et al. 2013). Several interesting findings are reported, for example: the likelihood of friendship with a person is found to be decreasing with distance among persons (Backstrom, Sun, and Marlow 2010); people that travels has more chances to meet friends and thus gets involved in more social activities (Cheng et al. 2011); persons with more friends tend to create triangles with individuals further apart (Scellato et al. 2011). Exchange of social and locational information is accelerated in the era of social media that allow persons to make a decision about physical movement based on social and spatial information (Humphreys 2007). Croitoru et al. identified and mapped connected communities and their structure based on social media feeds (Croitoru et al. 2013). Cranshaw et al. introduced a clustering model and research methodology for studying the structure and composition of a city on a large scale based on social media data (Cranshaw et al. 2012).

Constructing Social Connections from Facebook

Data Collection on Facebook

Facebook is one of the most popular social media that foster close personal relationships that are projected from real life (Russell 2013). The research for this dissertation has conducted an authentic, transparent, repeatable and accessible data collection mechanism for extracting location-based social media activities from Facebook. Because the posts on Facebook are not entirely public, this research has applied and obtained the IRB approval from the University of Georgia. A dedicated website (www.lbsocial.net) is established for the Facebook data collection

(Figure 4-1). In addition to get IRB approval, the website itself is an authentic Facebook Application (Facebook 2013). All the collected data items have been reviewed and approved by the Facebook Company. Once people log in this data collection website, Facebook will send out a confirmation window explicitly explaining the items that will be collected by this website (Figure 4-2).


www.lbsocial.net
GitHub UGA LIB Json Google Scholar FacebookAdmin GAEAdmin LBSocial Dropbox Analytics IssueCrawler dbdsgnr Domains Local
Disable Cookies CSS Forms Images Information Miscellaneous Outline Resize Tools View Source Options

WELCOME TO LOCATION BASED SOCIAL MEDIA ACTIVITY

LBSocial website help people extract, visualize and analyze social media data.
By logging in with your Facebook account, you have read the Consent Letter.

Log In
Demo of Login
Developed Tools

Collect Geotagged Facebook Data From Your Friends

Collect Friends' Facebook Data
View Collected Friends Data in Table
View Collected Friends Data on Map
Analyze Facebook Social Network

Collect Facebook Data From An Account

Facebook Account:

Get Your Feeds
View Collected Feeds

Collect Twitter Data By Twitter Account Or/And Time

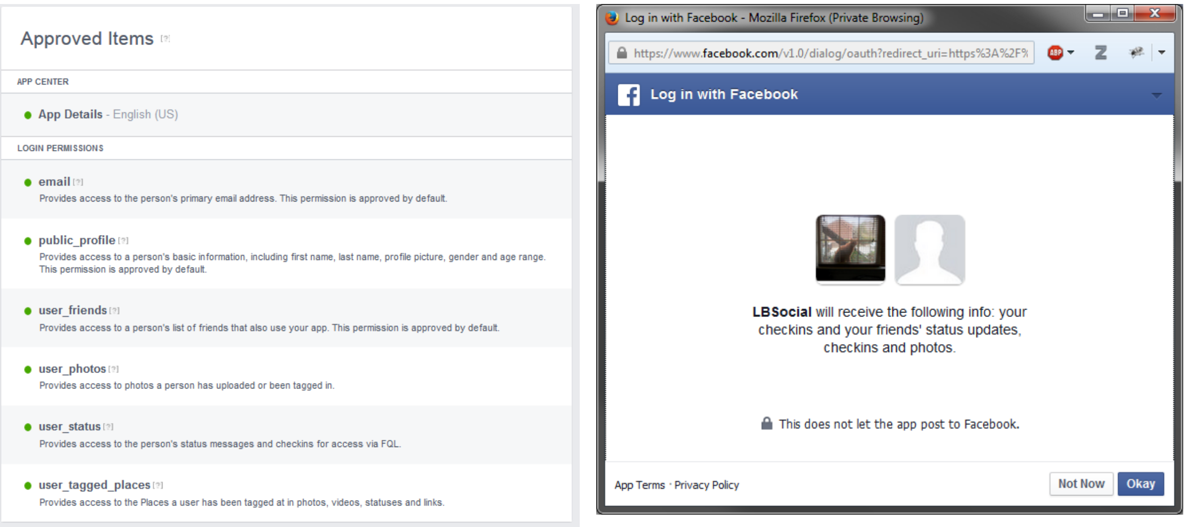
Twitter Account:
Start Date:
End Date:

541 of collected Tweets contain location information.
Below are the 100 latest tweets.

+
−
Leaflet | Map data © OpenStreetMap contributors, CC-BY-SA, Imagery © Mapbox
Tweets
Follow

Xuebin Wei @xuebinwei 45m
the #socialnetwork analysis from lbsocial.net reveals the structure of my #Facebook friends' interactions pic.twitter.com/kShZdxV6U

Figure 4-1 Interface of Data Collection Website



Approved Items [?]

APP CENTER

- App Details - English (US)

LOGIN PERMISSIONS

- email [?]
  Provides access to the person's primary email address. This permission is approved by default.
- public_profile [?]
  Provides access to a person's basic information, including first name, last name, profile picture, gender and age range. This permission is approved by default.
- user_friends [?]
  Provides access to a person's list of friends that also use your app. This permission is approved by default.
- user_photos [?]
  Provides access to photos a person has uploaded or been tagged in.
- user_status [?]
  Provides access to the person's status messages and checkins for access via FQL.
- user_tagged_places [?]
  Provides access to the Places a user has been tagged at in photos, videos, statuses and links.

Log in with Facebook - Mozilla Firefox (Private Browsing)
https://www.facebook.com/v1.0/dialog/oauth?redirect_uri=https%3A%2F%
Log in with Facebook

LBSocial will receive the following info: your checkins and your friends' status updates, checkins and photos.

This does not let the app post to Facebook.

Figure 4-2 Approved Items by Facebook Company

The website is built in a Python environment on the Google App Engine (Google 2013), and is accessible to all users from the entire world to log in and collect data for their own purposes. Meanwhile, the website is customizable that can support collecting data from other social media platform, or transform data into a different format. The current version of the website fully supports Facebook API (Facebook 2013) and Twitter API (Twitter 2013) that can extract location-based Facebook posts or Tweets.

This research project implemented a Snow-ball data collection mechanism to extract social relationships on Facebook (Figure 4-3). The researcher invited several volunteers who have valid Facebook accounts to log into this website. Those volunteers become the seed Facebook users in this snow-ball data collection. Once the volunteers agreed with the IRB approved consent and logged in the website, the website will retrieve all the Facebook friends of the seed users, and collect the Facebook posts, such as status and photos that are embedded with geographic locations, for every Facebook friend of those seed users. Those Facebook friends of the seed users are egos (Hanneman and Riddle 2005) in the collected social relations because all the information is directly related to them. For each Facebook post, the website has recorded the time, the textual description, the tagged Facebook users on photos or status, and geographic locations where the post is published. Those tagged Facebook users are actors in the social relations since they all have direct connections to the ego Facebook users. The collected data is maintained in local GIS database for further analysis.

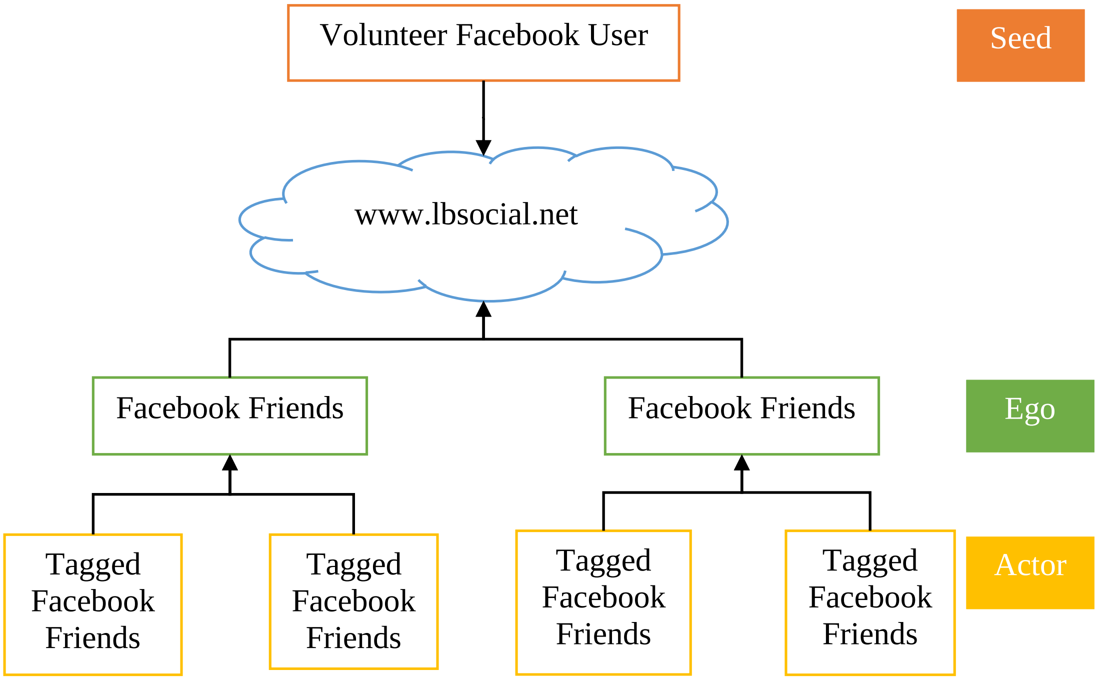


Figure 4-3 Snow Ball Data Collection of Social Relationships

Constructing Social Connection

With 50 seed Facebook users logged in the data collection website, 1,213 ego Facebook users and their location-based posts are recorded into the database (Table 4-1). For each single post, all the tagged participants are the Facebook friends of the collected Facebook user, and have physically (photo post) presented or literately (status post) mentioned in the Facebook post. Therefore, each participant mentioned in this post is assumed as a friend in the real world to all the other participants, and form social connections with all the other participants.

There are 16,612 actors identified from those 1,213 egos. In addition, 103,232 social connections are formed from those 19,756 location-based Facebook posts. Figure 4-4 depicts the locations and traveling paths of the collected Facebook posts all over the world. Those Facebook posts are published at 6,027 different places that have been classified into 520 categories by the Facebook Company.

Table 4-1 Summary of Collected Facebook Data

Type		Number
Seed Facebook User		50
Ego Facebook User		1,213
Actor Facebook User		16,612
Connection		103,323
Post		19,756
	Photo	15,167
	Status	4,589
Place		6,027
Place Type		520

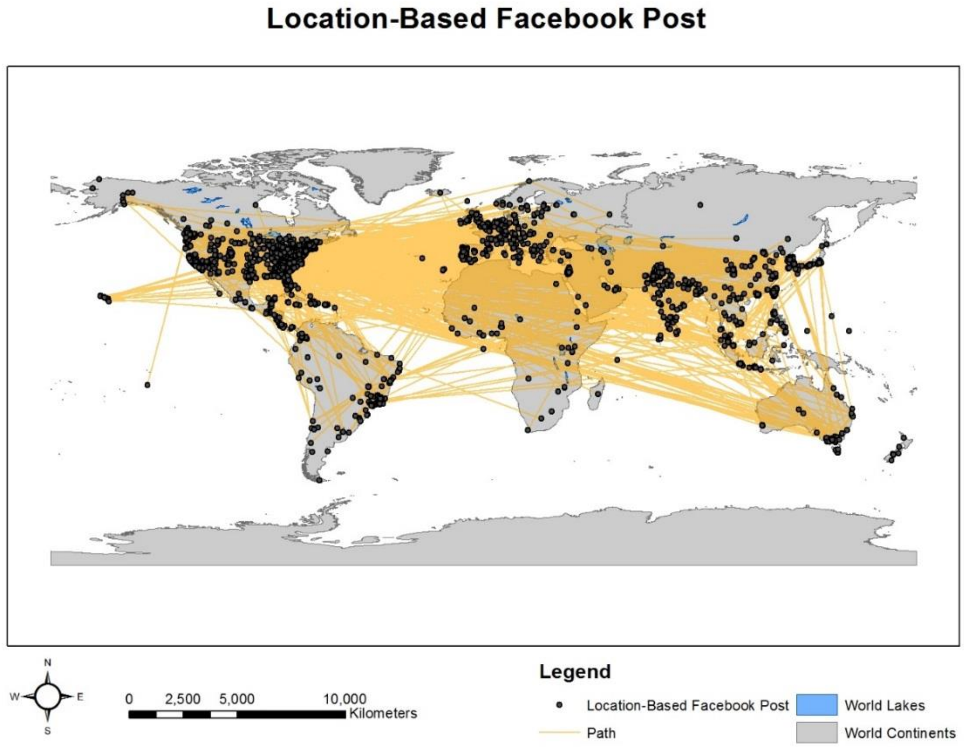


Figure 4-4 Location-based Facebook Posts

All the collected actors have visited 6027 different places (Table 3-1). Those places have been classified into 520 sub-categories by Facebook. Figure 4-5 illustrates the top 10 most popular place types. The most visited places are cities, following by the college and university, and local business. The type of places provides additional information about the activities and the social connection that happening at those places.

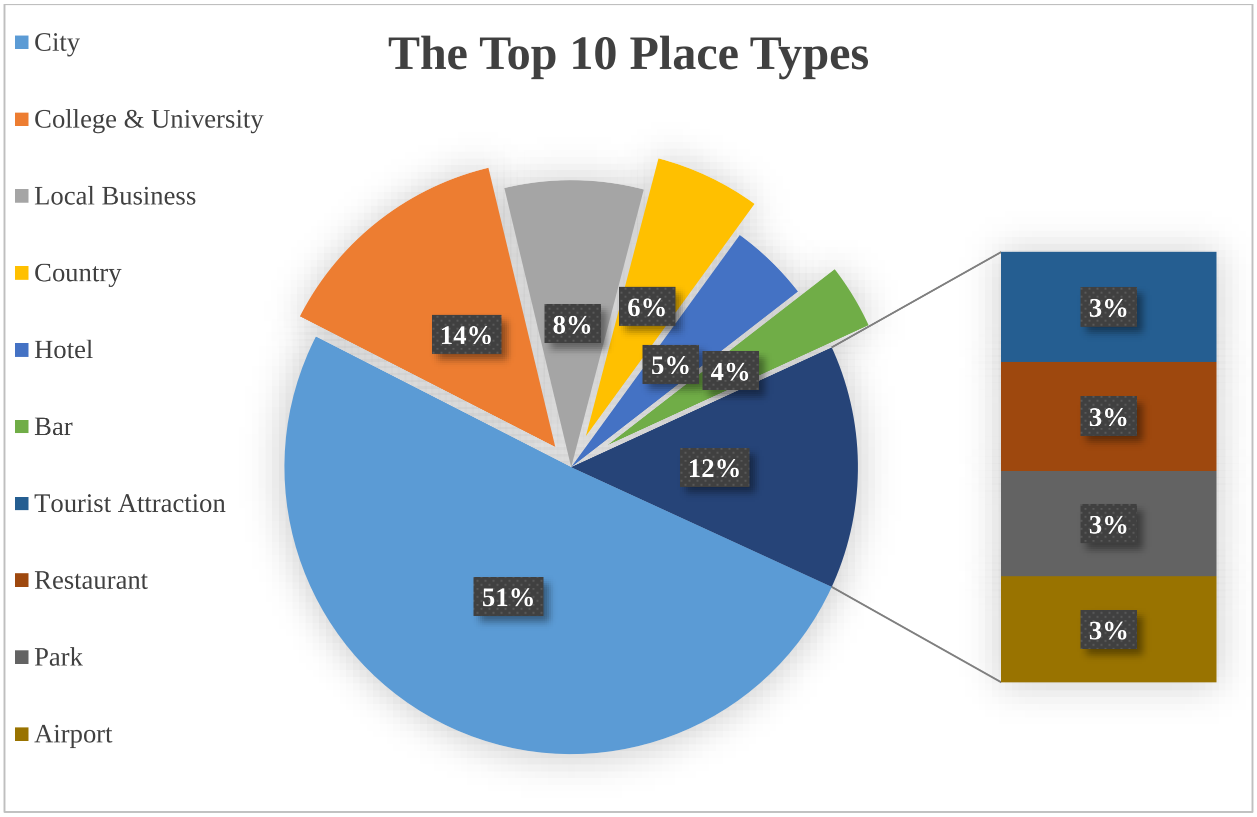


Figure 4-5 The Top 10 Place Types

For each place, Facebook provided the total number of likes that this place has received on line; the total number of times this place has been mentioned in Facebook posts; and the actual number of check-ins at this place. Those measurements can serve as popularity indices for places in the virtual world and the physical world. For example, Figure 4-6 shows how those measures vary across geographical space in Atlanta. The places in downtown are more popular in terms of receiving likes, talking about or check-ins.

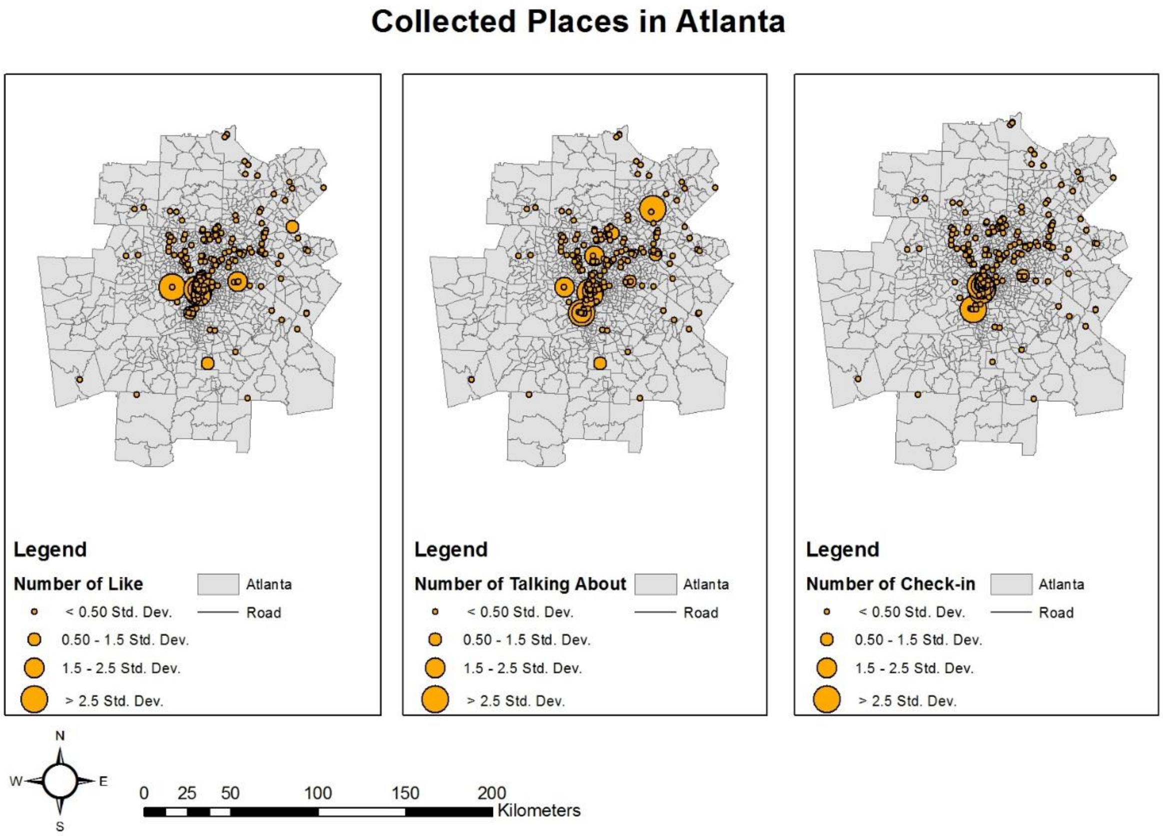


Figure 4-6 Places in Atlanta

The textual descriptions of the Facebook posts provide contextual information of the location-based human activities in both the physical world and the virtual world. From the word cloud of the Facebook posts (Figure 4-7), it is easy to find that most posts on Facebook are discussing positive activities, evidenced by the high frequency of the joyful vocabularies.

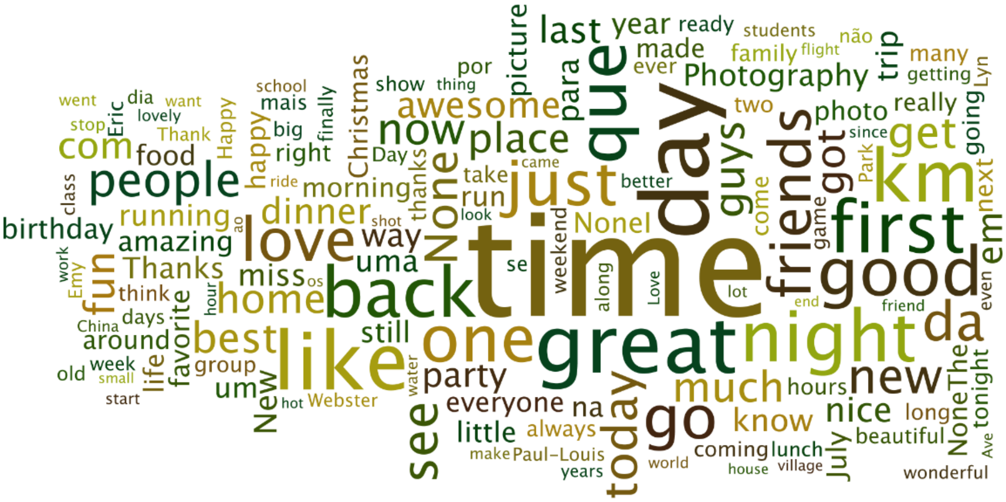


Figure 4-7 Word Cloud from Facebook Posts

Social Network Analysis from the Constructed Social Connections

The way of defining social connection in this paper provides a flexible method to construct social networks for different purposes or at different scales. For example, the entire social network has been constructed that comprises all the egos and their connections to other actors. Figure 4-8 depicts the structure of the entire social network. The different colors distinguish the persons into different communities.

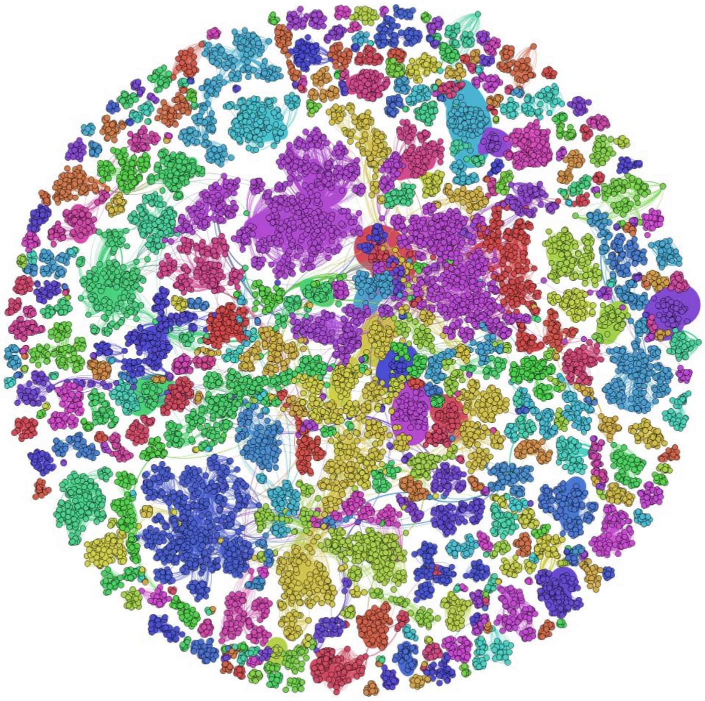


Figure 4-8 The Entire Social Network

Table 4-2 reports the statistics of the entire social network. The degree of a node is the number of the other directly connected nodes to this node. On average, each person who is mentioned the collected Facebook posts interacted with 12 other actors, and most persons have less than 40 direct connections (Figure 4-9). The density of a network is the proportion of all possible links that actually presents. Since the seed Facebook users and the ego Facebook users are from all over the world, their collected Facebook friends don not overlap with each other, resulting a low density of 0.001. The modularity and the connected components identify groups of nodes that are closely connected to each other within the group. Those persons consist of 376 separated sub-groups, i.e., connected components within which members of each group do not interact with members outside of their sub-groups. Among those components, 417 communities are identified. The modularity and the average clustering coefficient measure the average densities of neighborhoods or communities of all the nodes. Therefore, within each communities, persons almost know every other persons that is evidenced by the high modularity and the high average clustering coefficient. The sizes of the most communities are less than 50 (Figure 4-10), indicating that there less than 50 individuals in those identified groups. Those findings prove that the human interactions reflected on Facebook are constrained in small but close groups.

Table 4-2 Measurement of the Entire Social Network

Measurement	Value
Average Degree	12.440
Density	0.001
Modularity	0.983
Number of Communities	417
Connected Components	376
Average Clustering Coefficient	0.915

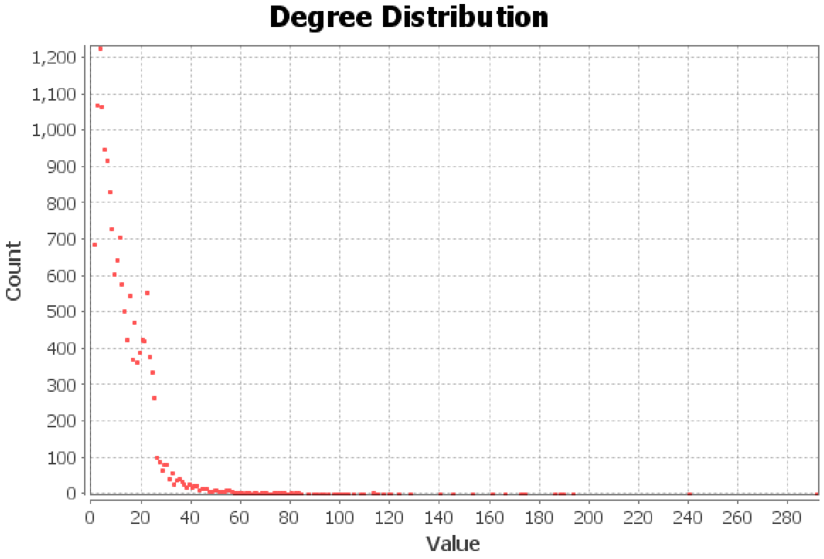


Figure 4-9 Degree Distribution of the Entire Social Network

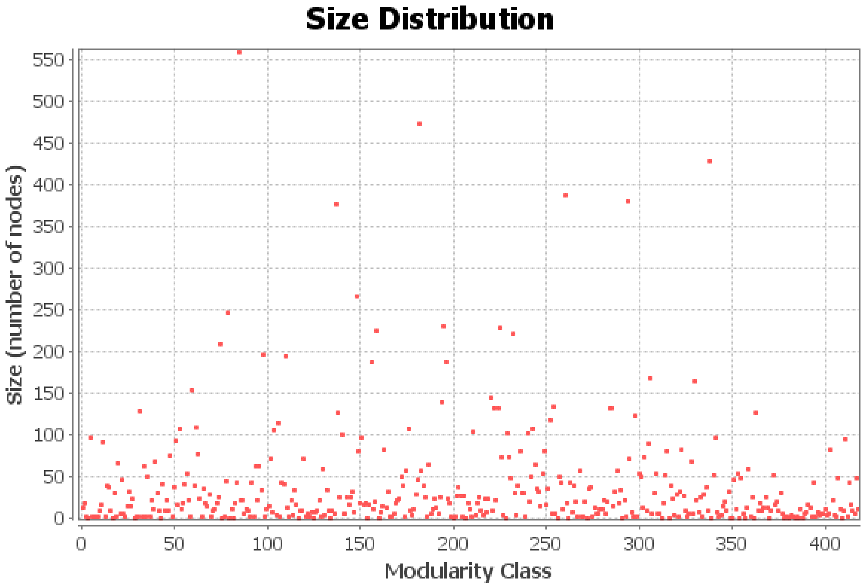


Figure 4-10 Community Size Distribution of the Entire Social Network

For each ego Facebook user, the ego-actors network is constructed to investigate the social structure at an individual level. In addition to calculate the basic network measures of each ego network, the total travel distance of each ego is also computed. Meanwhile, if an ego Facebook user publishes several posts at a same geographic place with same actors on a same day, those posts are counted as a single visit to this place. Therefore, the number of visits for each ego Facebook user is calculated as well. Table 4-3 summarizes how those measures correlated with each other. The high correlations among the number of nodes, the number of posts, the number of visits and the total travel distance confirm the hypothesis of previous study (Cheng et al. 2011) that people travel a lot will have more friends and get involved in more social activities.

Table 4-3 Summary of the Ego Networks

	Number of Node	Number of Post	Number of Visit
Number of Post	0.6788*	1	
Number of Visit	0.7258*	0.8741*	1
Number of Clique	0.7495*	0.6068*	0.6270*
Density	-0.7024*	-0.5997*	-0.6368*
Average Clustering Coefficient	-0.0879*	-0.1567*	-0.1600*
Sum of Travel Distance	0.5071*	0.5969*	0.6907*
* indicates statistical significant at 5%			

Location-Based Social Connection

Visualization of Location-Based Social Connection

The social connections can be quantified and spatialized based on the connection formation from the collected Facebook posts. For example, if user *A* published more posts tagging act user or B than posts tagging user *C*, it is fair to conclude that user *A* has stronger social connections with user *B* than with user *C*. Therefore, the strength of ono-to-one social connection can be defined as the times of the formation of such one-to-one social connections in different social activities. In addition, because in each location-based Facebook post tagged participants and geographical location are well known, the social connections between user *A* and user *B* can be further distinguished based on the geographical locations. This research has modified the social network in the way that each social connection includes (and thus is distinguished by) the latitude and longitude where the post is published. Those social connections embed with geographical locations are called location-based social connection.

Such location-based social connections can be visualized in GIS. Figure 4-11 displays the location-based social connections in Atlanta where each point represents a single social connection between two people, and the size of the point indicate the times of this location-based social connection occurs at this place, i.e., strength of the location-based social connection.

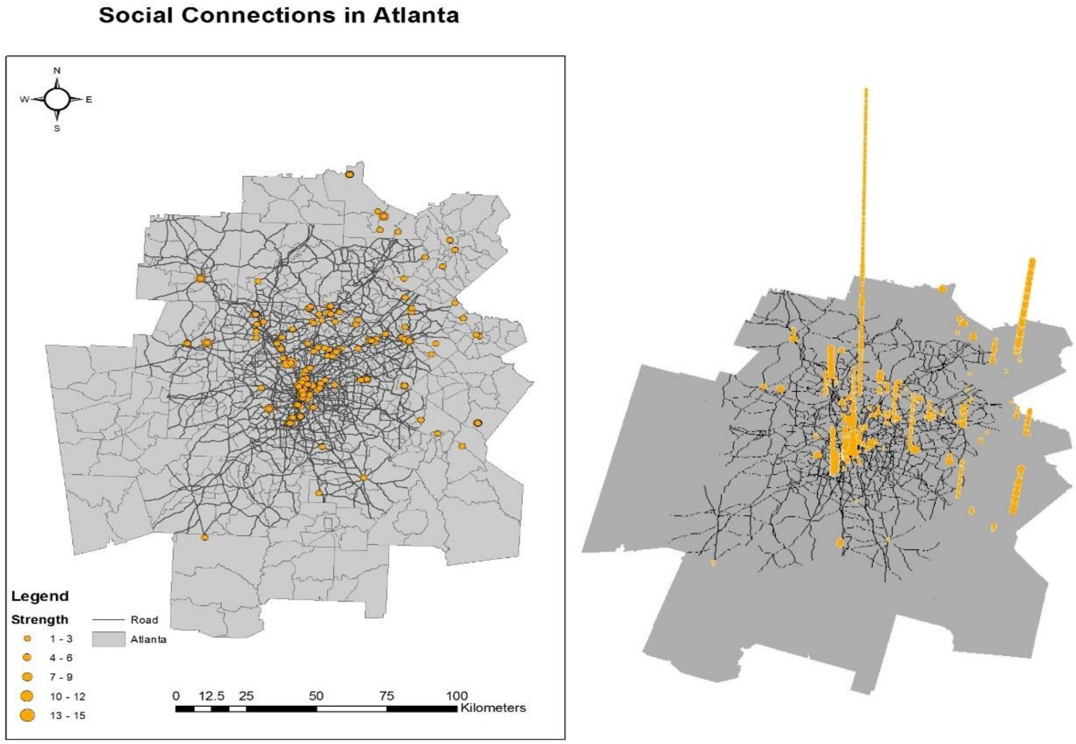


Figure 4-11 Visualization of Social Connections in Atlanta

Spatial-Social Clusters

Based on the location-based social connections, social connections can be geographically analyzed. For example, the local Moran's I of the location-based social connections based on the connection strength can identify clusters of social interactions. Since local Moran's I can determine the cluster points where high values are surrounded by other high values, the positions where strong social connections locating nearby the other strong social connections can be identified, i.e., the social-spatial clusters of activities indicating the place where those users involved in those activities are socially and spatially close.

In the example of Figure 4-12, several sets of location-based social connections are distributed on two places. Most of the connections are relative strong evidenced by the great size of the connection points. The local Moran's I identified at least 4 of those connections as H-H clutters, because they are strong social connections located within shorter distance, namely 260 – 5422, 260 – 1281, 492 – 4928, and 492 – 1120. Figure 4-13 illustrates how the social network looks like from the Figure 4-12 example. User 260 knows both user 1281 and User 5422, and forms very strong social connections with the both users at the same location. It is also highly possible that user 1281 will form social connection with user 5422 given that their social interactions are spatially clustered, and both have strong social connection to user 260.

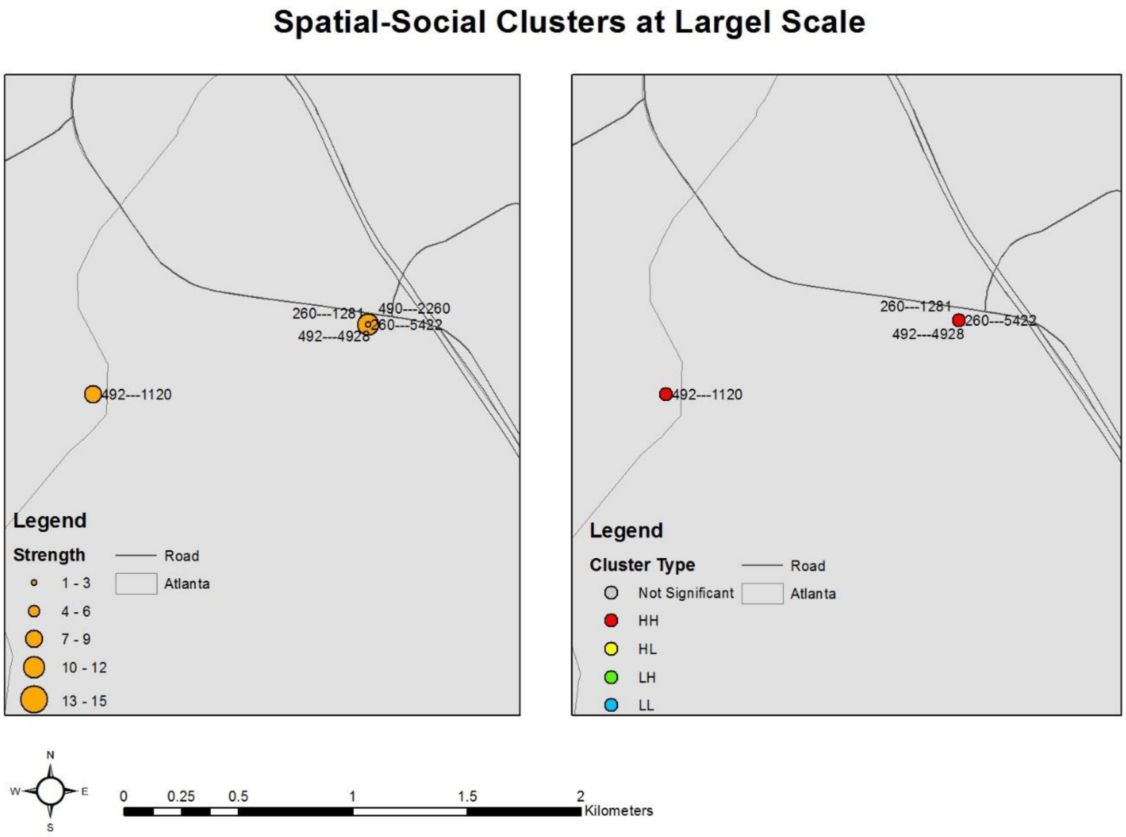


Figure 4-12 Example of Spatial-Social Clusters at Fine Scale

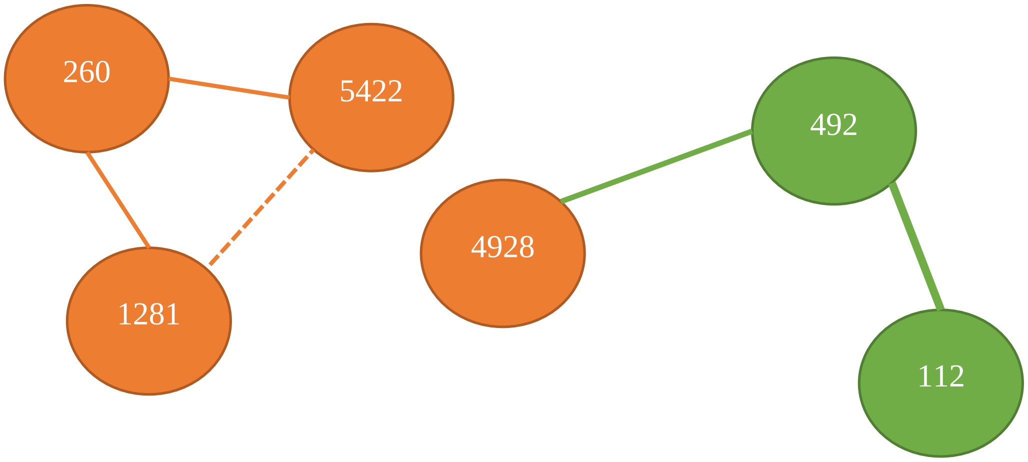


Figure 4-13 Social Network of Figure 4-12 Example

Those spatial-social clusters can be visualized spatially (Figure 4-14) and socially (Figure 4-15). The red points in Figure 4-14 indicate the geographical locations of those spatial-social clusters and the red links in Figure 4-15 illustrate how those spatial-social clusters are embedded in the social network. Figure 4-16 combines the information from Figure 4-14 and Figure 4-15 demonstrating how people (node) are interacted with others with the spatial-social connections (edge). This visualization integrates persons and their social connections in a spatial-social dimension that can generate meaningful information regarding human activities. For example, by retrieving several persons, the spatial-temporal paths or interactions can be visualized and analyzed. In addition, by selecting a specific spatial-social cluster, the involved users can be identified and visualized.

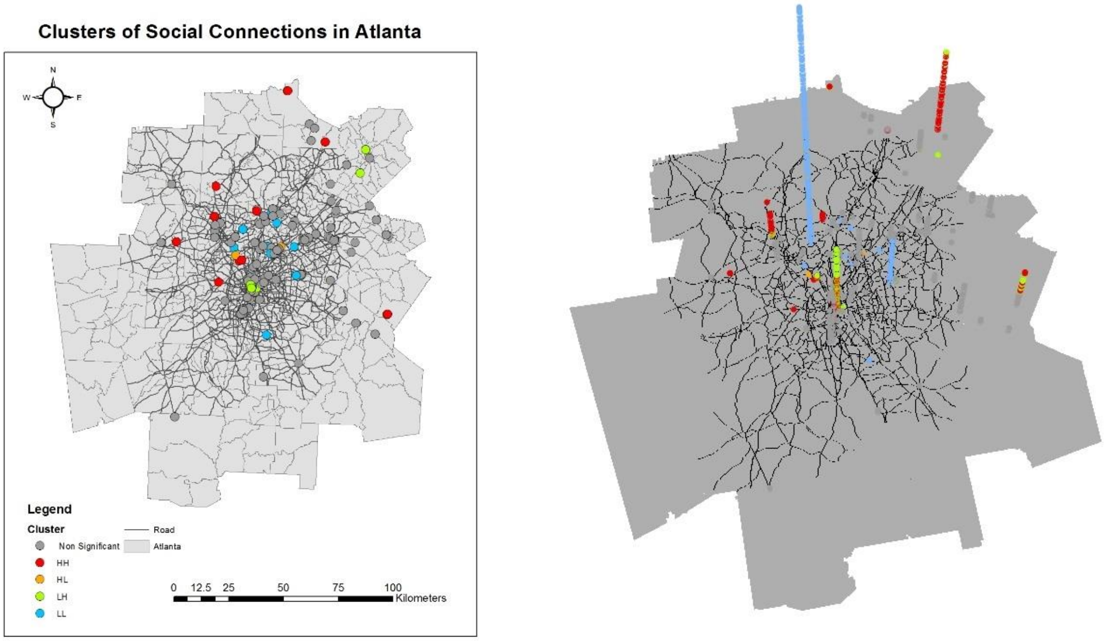


Figure 4-14 Clusters of Social Connections in Atlanta

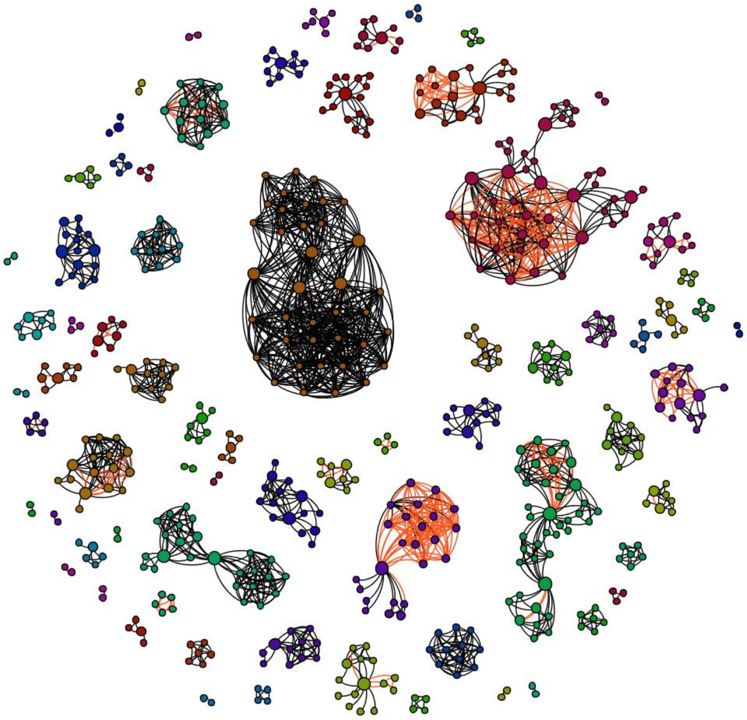


Figure 4-15 Social Connections in Atlanta

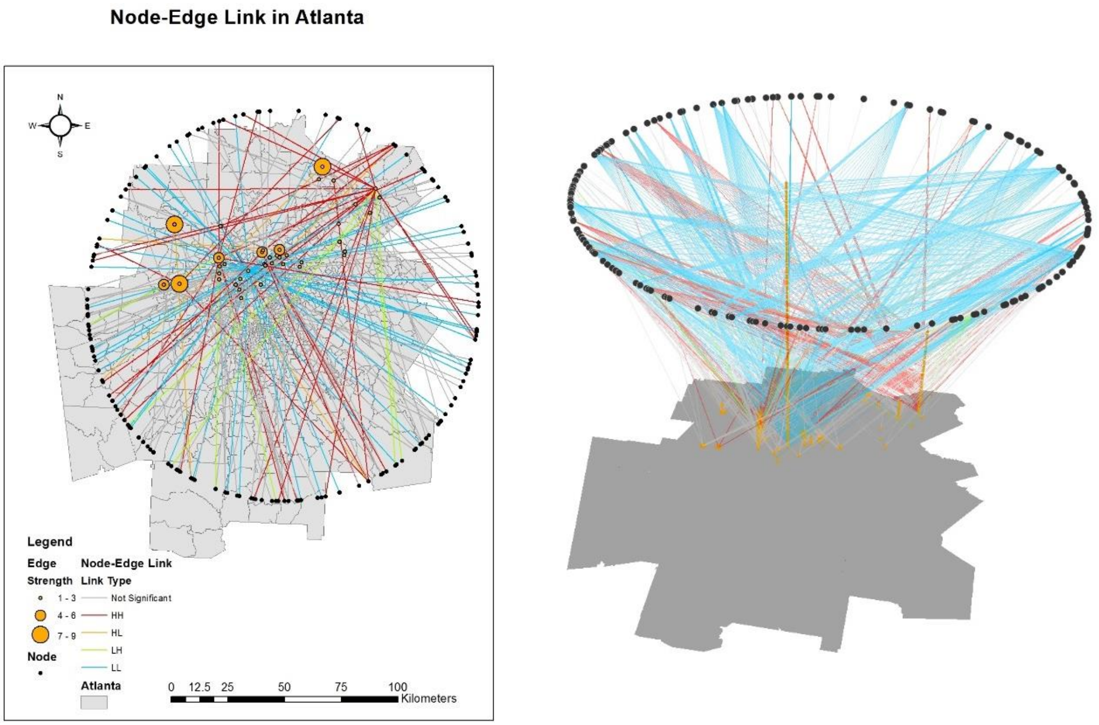


Figure 4-16 Visualization of Node-Edge Link in Atlanta (Partial)

Discussion and conclusion

Human activities are essentially spatial and social. Traditional surveying method of gathering social relationship is time consuming, resulted in incomplete and non-independent sample data. Location-based social media not only provides new platforms of communicating that influence the decision making of human geographical behaviors, but also serves as new channels for observing and analyzing human activities in spatial-social dimension at fine scales. This research provides a data collecting mechanism and social connection constructing method that are able to extract massive location-based social media activities in an authentic, transparent, repeatable and accessible way. Based on the proposed concept of the location-based social connection, human interactions can be visualized and analyzed in a spatial-social dimension, and spatial-social clusters can be identified and visualized in GIS. The proposed method can be applied to many areas, such as simulation and analysis disease spread, and traffic analysis, etc.

Reference

Adams, P. 1998. Network topologies and virtual place. Annals of the Association of American Geographers 88 (1):88.

Backstrom, L., E. Sun, and C. Marlow. 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. New York, NY, USA: ACM http://doi.acm.org/10.1145/1772690.1772698 (last accessed 10 December 2014).

Butts, C. T., R. M. Acton, J. R. Hipp, and N. N. Nagle. 2012. Geographical variability and network structure. Social Networks 34 (1):82–100.

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

Cho, E., S. A. Myers, and J. Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 1082–1090. ACM.

Crandall, D. J., L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. 2010. Inferring social ties from geographic coincidences. Proceedings of the National Academy of Sciences 107 (52):22436–22441.

Cranshaw, J., R. Schwartz, J. I. Hong, and N. M. Sadeh. 2012. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In ICWSM. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/download/4682/4967 (last accessed 6 April 2014).

Croitoru, A., A. Crooks, J. Radzikowski, and A. Stefanidis. 2013. Geosocial gauge: a system prototype for knowledge discovery from social media. International Journal of Geographical Information Science 27 (12):2483 – 2508.

Eagle, N., A. Pentland, and D. Lazer. 2009. Inferring friendship network structure by using mobile phone data. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA 106 (36):15274 – 15278.

Emch, M., E. D. Root, S. Giebultowicz, M. Ali, C. Perez-Heydrich, and M. Yunus. 2012. Integration of Spatial and Social Network Analysis in Disease Transmission Studies. Annals of the Association of American Geographers 102 (5):1004–1015.

Facebook. 2013. Facebook Developers. Facebook Developers. https://developers.facebook.com/ (last accessed 25 December 2013).

Google. 2013. Google App Engine — Google Developers. Google App Engine: Platform as a Service. https://developers.google.com/appengine/?csw=1 (last accessed 24 January 2014).

Hanneman, R. A., and M. Riddle. 2005. Introduction to social network methods. University of California Riverside.

Hipp, J. R., R. W. Faris, and A. Boessen. 2012. Measuring “neighborhood”: Constructing network neighborhoods. Social Networks 34 (1):128–140.

Humphreys, L. 2007. Mobile Social Networks and Social Practice: A Case Study of Dodgeball. Journal of Computer-Mediated Communication 13 (1):341–360.

Ji, S. Y., E. Niklas, and S. Lee. 2010. TimeMatrix: Analyzing Temporal Social Networks Using Interactive Matrix-Based Visualizations. International Journal of Human-Computer Interaction 26 (11/12):1031–1051.

Kwan, M.-P. 1998. Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-based Framework. Geographical Analysis 30 (3):191–216.

Padmanabhan, A., S. Wang, G. Cao, M. Hwang, Z. Zhang, Y. Gao, K. Soltani, and Y. Liu. 2014. FluMapper: A cyberGIS application for interactive analysis of massive location-based social media. Concurrency and Computation: Practice and Experience 26 (13):2253–2265.

Russell, M. A. 2013. Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More. O’Reilly Media, Inc.

Scellato, S., A. Noulas, R. Lambiotte, and C. Mascolo. 2011. Socio-Spatial Properties of Online Location-Based Social Networks. ICWSM 11:329–336.

Twitter. 2013. Exploring the Twitter API. Twitter Developers. https://dev.twitter.com/console (last accessed 25 December 2013).

Zhang, S., and X. Yao. 2011. Social-spatial structure of Beijing : a spatial-temporal analysis. 2011.

CHAPTER 5 CONCLUSION AND FUTURE RESEARCH

Conclusion

People are social creatures that inhabit in the physical world and interact with others in the social world. With the popularization of location-based social media, people can communicate their geographical locations and interactions of social events with others. Such location-based social media data provides theoretical and technical challenges in GIScience to model human activities in a comprehensive way.

To represent and analyze location-based social media activity, this research has extracted and examined real-world LBSMA data, and organized the data in an effective and efficient way that support the analysis of human activities in the spatial-social-temporal dimension. Specifically, this paper has developed the first conceptual model and associated logical model to represent LBSMA data in GIS, implemented a pilot computer system, and formulated applicable methods for visualization and utilization of LBSMA data in spatial-temporal and social dimension. Chapter 2 elaborates the proposed conceptual framework and associated pilot prototype where the location-based social media activities can be efficiently organized and effectively analyzed. A case study is conducted to prove the usefulness of the proposed data model and the developed tools. Chapter 3 and Chapter 4 introduce the methods of quantifying places and social connections in different aspects. In Chapter 3, a simulation methodology that mimics human activities in spatial-temporal dimension, i.e., random walking algorithm, is developed to characterize urban road networks. It is found that the random walking algorithm is able to quantify the spatial features of urban road networks and their surrounding areas in terms of their correlations to social-economic characteristics. This method can thus be applied in assessing important areas of urban areas when spatial connectivity or social-economic importance are of interests. Chapter 4 focuses on the spatial-social dimension of human activities where location-based social media activities are employed again to construct social relations at a fine scale. The constructed social connections can be re-defined by coupling with geographical information, i.e., the location-based social connections, in which social connection can be geographically visualized, and spatial-social clusters can be identified. This method is valued in the analysis of disease spread, information transmission, and localized advertisement.

However, there are some limitations in this doctoral research. First of all, the location-based social media activities cannot reflect the whole picture of the human interactions in the real world. A large number of people are not used to share their information online. In addition, people may have different preferences on distinct social media platforms. Studies that focus on one social media data may result in biased conclusions. Secondly, the definition of the one-to-one social connections in this research are based on the assumption that two people who are Facebook friends and physically presented in the same social events are friends in the real world. This assumption should be verified by comparing the detected social connections with the survey data. Finally, the geographical presences of persons are determined from all Facebook posts. However, if no further information are provided, only the photos that capture human faces should be used to extract the spatial presences of persons in different activities.

Future Research

Location-based social media provides unprecedented opportunities for geographers to observe human activities in the spatial-temporal-social dimension. Based on this research, new theories, analysis methods and data collection technologies can be further investigated or explored. For example, the definition of place can be re-considered where the virtual world and the physical world are tightly connected on the location-based social media. Places are not static objects but rather dynamic systems that can be defined in terms of contexts (Adams 1998). A new representation of the place that incorporates information in the geographic space and the virtual space is thus significant for the study of location-based social media activities.

To analyze location-based social media activity, effective approaches are required to fully understand the purposes and the dynamics of human interactions. For example, semantic analysis of online posts from Facebook or Twitter can generate meaningful information regarding the purpose and the content of human activities. The location-based social media activities provide preferences and social strengths of human interactions in the spatial-temporal-social dimension. Such information forms a complex system in which random walking algorithm can be improved to calculate more measurements of human activities by imitating human interaction under those spatial-social-temporal constraints.

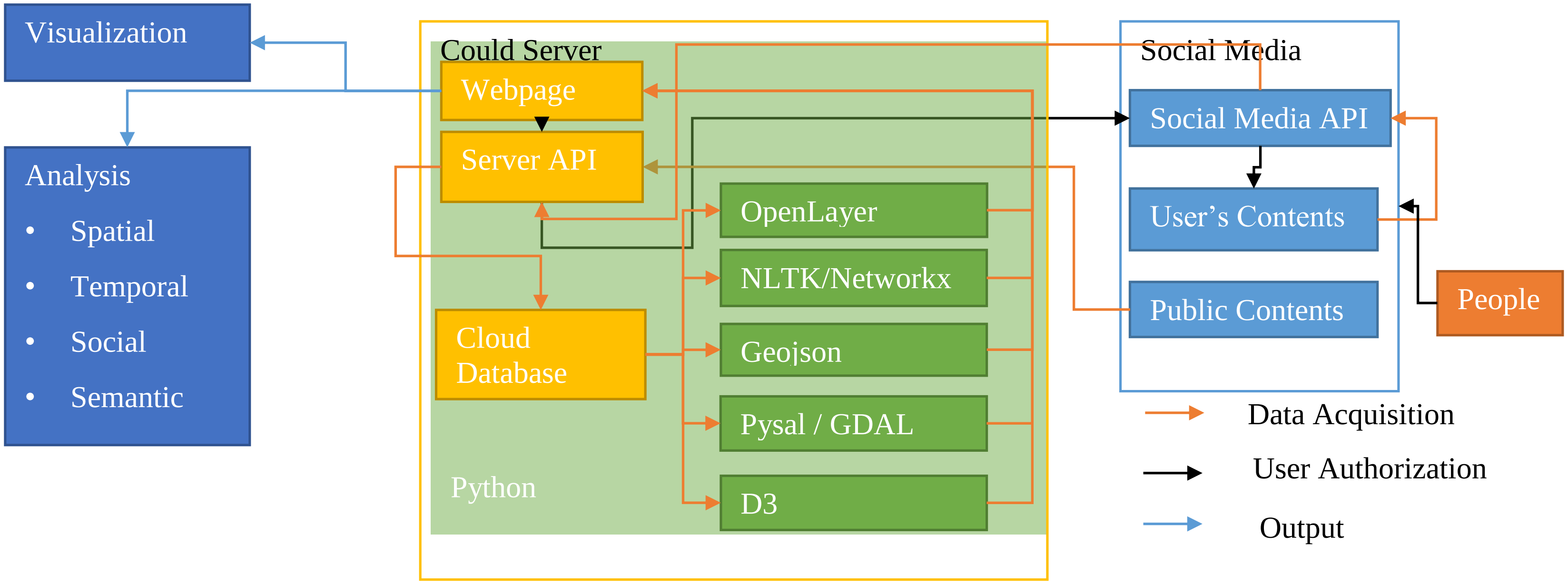


Figure 5-1The Next Generation of LBSocial.net

Finally, authentic and transparent data collection system is needed to extract social media data while protecting users' privacy. Since people constantly publish their interactions and activities on different social media, the next generation of the LBSocial (Figure 5-1) website (www.losbcial.net) will continue supporting the collection of location-based social media data that serve as one of the open source data for different studies. In addition, the spatial, social, temporal and semantic analysis can be performed online for non-expert users. By utilizing the cutting edge of visualization techniques and incorporating spatial, temporal, social and sematic tools, additional online visualizing and spatial analysis functions will be provided for non-GIS users to explore their interested social media data.

Reference

Adams, P. 1998. Network topologies and virtual place. *Annals of the Association of American Geographers* 88 (1):88.