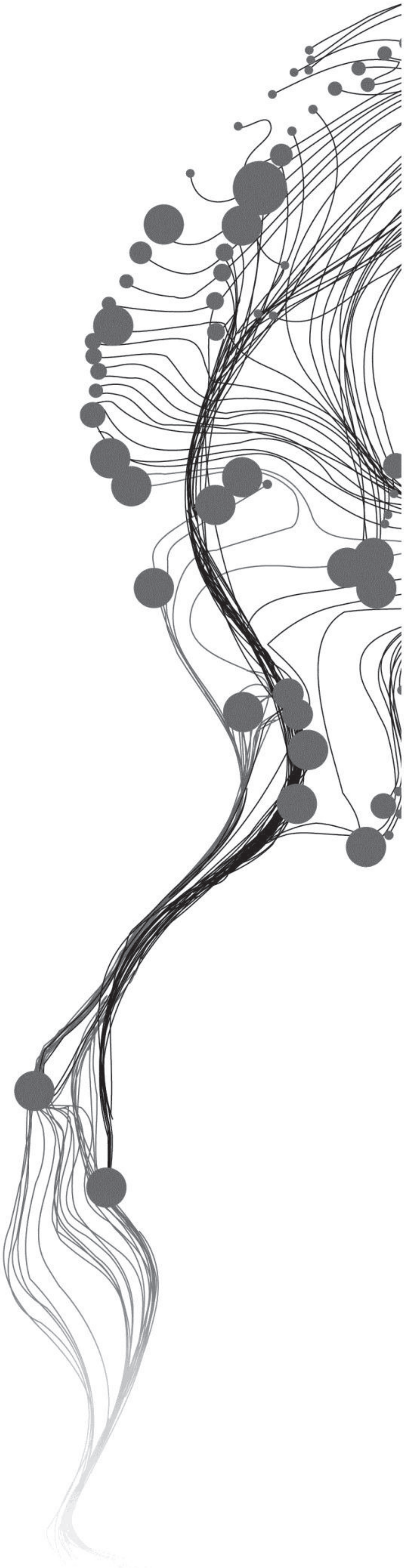
VISUALIZATION OF SOCIAL MEDIA DATA: MAPPING CHANGING SOCIAL NETWORKS

DING MA
FEBRUARY, 2012

SUPERVISORS:
Prof.Dr. M.J. Kraak
Dr.Ir. R.L.G. Lemmens



VISUALIZATION OF SOCIAL MEDIA DATA: MAPPING CHANGING SOCIAL NETWORKS

DING MA
Enschede, The Netherlands, February, 2012

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

SUPERVISORS:
Prof.Dr. M.J. Kraak
Dr.Ir. R.L.G. Lemmens

THESIS ASSESSMENT BOARD:
Chair: Prof.Dr.Ir. M.G. Vosselman
External examiner: Prof.Dr. C. Robbi Sluter (Federal University of Parana, Brazil)

Dedicated to my parents

ABSTRACT

Recently, countless social networks have been built via social media. Of those two kinds of networks are most popular: user-centric social network which develops from online relationships around a user (e.g. one's friends in Facebook or followers in Twitter etc.), and object-centric social network which develops from online interactions around a social object (e.g. photo in Flikr, video in Youtube or hashtag in Twitter etc.).

In order to understand these social networks, people already visualized them based on all kind of criteria, however, seldom based on geography. Since increasing number of geo-information in form of place name, GPS coordinates etc. exist in social media data generated by user, location becomes a criterion to help people physically understand these networks. This can be strengthened by including time as well to understand, the spatio-temporal dynamics of social networks. This for, both individual movement with changing friend composition (user-centric network), and spatial diffusion of information (object-centric network), also need to be investigated and explored.

The aim of this research is to visualize spatio-temporal dynamics of social networks. The starting point is Peuquet triad framework. This allows one to approach social network data from a spatial, temporal and attributes perspective, and uses it as the basis to analyze related user tasks. Based on the data framework and user tasks, a multiple linked view visualization environment combining social node-link diagram and map based visualizations together is proposed to reveal the spatio-temporal characteristics of changing social networks. Two case studies are used to illustrate this approach: one is my Facebook friend network (user-centric network) and another is trending topic (Japan earthquake) network in Twitter (object-centric network). The designed prototypes for the two case studies consisting of the implemented graphic representations and designed working environment were evaluated by the focus group method. Finally, conclusions and recommendations are presented.

Keywords: social networks, social media data, user-centric social network, object-centric social network, Facebook, Twitter, triad framework, social node-link diagram, map based visualization

ACKNOWLEDGEMENTS

First and foremost I would like to offer my deepest gratitude to my first supervisor, Prof. Dr. Menno-Jan Kraak, for his supports and patience throughout the process of the research. Without his inspiring guidance, I cannot finish the work. And I also take this chance to thank my second supervisor, Dr. Ir. R.L.G. Lemmens, for his valuable advices and comments.

My sincere thanks go to Dr. Tiejun Wang, for his help and care all the time during this one and half year.

I would like to thank Dongpo Deng, for the valuable suggestions you offered me.

Special thanks to Xia Li, who always give me support wherever she is.

I also want to thank Dr. Corné van Elzakker and Dr.Ir. Luc Boerboom, thank you for your help and coordination of my usability test.

I would like to express my gratitude to all my friends, happy to be with you in this study period. This experience would be a priceless treasure in my whole life.

Last but not least, my deepest thanks go to my parents, for your endless love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Motivation and problem statement

1.1.1. Background and Motivation

Social media, as Kaplan & Haenlein (2010) defined, "is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user generated content". It contain many kinds of online social platforms, ranging from blogs and microblogs (e.g. Twitter), content communities (e.g. Youtube) to social networking sites (e.g. Facebook) etc. From social media data, there are not only the mass media (text, audio, photo, video etc.) that people posted onto the web, but also inherently-built social networks that were not previously possible in both scale and extent (Barbier and Liu, 2011). Undoubtedly understanding such social networks can provide us useful insights of ways that the social communities are formed and interact, therefore there is a need to convert this dataset into meaningful information for people to understand.

Visualization can be deemed as an effective way to satisfy this demand for helping people understand social networks and convey the result of analysis (Freeman, 2004). In most cases, visualizations of the social network are node link graphs, where nodes represent individual actors (e.g., persons, organizations) and links represent relationship ties (e.g., communication, financial aid, contracts) between actors. These graphs focus on evaluating the centrality and influence of actors by the criteria such as degree, betweenness, closeness etc. and the community structure by ones such as cohesion, clustering etc. (De Nooy et al., 2005; Freeman, 2004; Wasserman, 1994). Consequently, they have become an important capability in many domains, such as business (Cross and Parker, 2004), expert assessment (McDonald and Ackerman, 2000) and criminal investigation (Chen et al., 2004) etc.

However, besides mentioned social relation measures, space and time should also be the important criteria to take into account. Although the effects of space and time limitations have greatly reduced by Internet and communication technology on social networks, space and time still matter because of the spatial and temporal context of human actions (L. Li and Goodchild, 2010). To be specific, from spatial perspective, each social entity has location information as an important property and combining this information with social networks we can gain more insights of the unknown patterns of the community (Wellman, 1996). For example, physical proximity means more ties to other people (Cummings et al., 2006) as well as more interactions with them (Mok and Wellman, 2007). Such patterns also exist in social networks like MySpace, Facebook (Escher, 2007). And moreover, with the increase of location-enabled mobile devices, social media make location have more efforts on building social networks. For example, from Twitter, people can send explicit or implicit geo-located tweets (GPS coordinates or geo-name in text) to interact with followers; from location-based social networks (e.g. foursquare.com and the "Places Check-in" feature on Facebook) people use location data to facilitate their socialization; from Volunteered Geographic Information (e.g., Wikimapia, Google MyMaps), people associate with others with home town, point of interest (POI), work place, geo located digital documents, etc. (Khalili et al., 2009). From temporal perspective, social networks from social media always vary over time. Integrating time in social networks can help people detect valuable information like change, trend, duration etc. Take one's network as an example, considering time we could get that how the number of friends change, what are the trends of network size or structure, how long the relationship keeps with one or a group of friend(s) etc. Also one's movement (location changes over time) like migration and travel can trigger changes on the social network. The interplay between mobility and the new network patterns has to be addressed (Timo, 2006). Therefore it is necessary to involve space and time for deepening our understanding on the social network.

In the visualization field, however, there is a gap between spatio-temporal data representation and traditional social node-link graph, since to date very few studies considering spatio-temporally integrating social networks and meanwhile keep the original features of the networks. It is evident that finding the link between these two can help us to address spatio-temporal problems of social networks, such as how does one event develop all over the world, how is one's composition of friends or friends' spatial distributions changing over time with one's movement etc.. In this case, the research aims at designing a visualization environment based on both geo-visualization methods and social node-link graphs to implement exploratory process of the social network data.

1.1.2. Problem statement

At present, social networks from social media data are more location-aware and dynamic. Following this trend, people are not only interested in understanding the static social structure by traditional node-link graphs, but also want to combine space and time to explore dynamic patterns of relationships and then deepen their understanding of the network. Researches have been conducted to this end and applied in different fields, such as travel (Timo, 2006), gang violence (Radil et al., 2010) etc. Also one example on VisualizationComplexity.com shows 1500 people use Twitter for communications at different places worldwide (Rafelsberger, 2008). What they have done have already brought social networks in a spatio-temporal context and then detected some spatial-temporal patterns of the network. However, existing researches cannot deal with spatio-temporal characteristics and social network properties at the same time. Therefore the problem of the research is (see Figure 1-1):

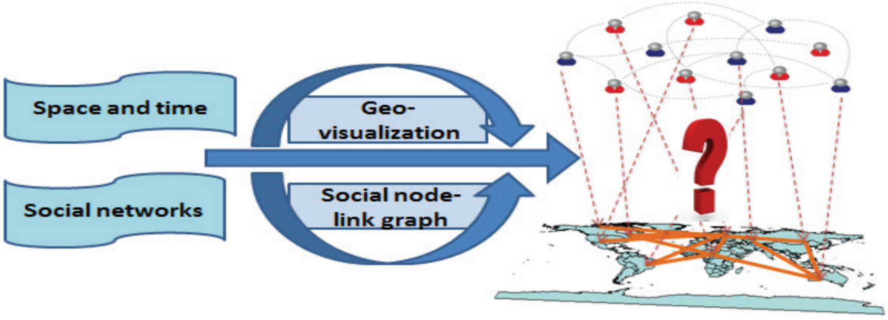


Figure 1-1: The problem statement

"Can we develop a visualization environment to incorporate social network graphics with geo-visualization methods to not only reveal the social networks' spatio-temporal characteristics but also keep the features in traditional social node-link diagram?"

1.2. Research identification

1.2.1. Research objectives

The main objective of the research is to design a visualization environment that allows the representation and exploration of social networks that have been extended with geo-components that change over time.

Based on the main objective, the sub-objectives are as follows:
1. To get an overview of existing visualization methods to represent social network data.
2. To extend the social network data with geo-components and select suitable graphic representations.
3. To design an effective prototype that allows visual exploration of the spatio-temporal social network data.
4. To evaluate the designed prototype.

1.2.2. Research questions

1. What are social networks and how do social networks evolve with the advent of social media?

2. Which existing visualization methods are suitable to depict social network data?
3. How to extend social networks with space and time?
4. Which graphic representations can be used for representing spatio-temporal social network data?
5. How can we represent all characteristics of social networks in a 'map'?
6. What are the required functionalities of visual interactive environment for spatio-temporal social network?
7. How to implement analysis and exploratory in the environment based on the use case(s)?
8. Which usability method to use to decide upon the effectiveness of the designed environment?

1.3. Innovation

As illustrated, space and time are new criteria for the social networks from social media data. However, existing visualization methods are limited to represent the spatio-temporal characteristics of the social networks. To this end, the research aims at extending social network data from geo-information perspective and then accordingly expanding the functionality of existing geo-visualization environment to explore the extended social network dataset.

1.4. Related work

Over the years, social relations and interaction patterns are visualized in node link graphs (Aggarwal, 2011; De Nooy, et al., 2005; Wasserman, 1994). The resultant network graphs frequently alter the geometric relations present in the real world in order to emphasize the connectivity and overall view of the networks (Khalili, et al., 2009). Among the graphs those nodes and links are not geographically encoded.

Recently, the spatio-temporal characteristics of social networks have been researched (Barthélemy, 2011; Mok and Wellman, 2007; Timo, 2006; Wellman, 1996). Efforts also paid on what the effects of space and time are in social networks from social media (Escher; Khalili, et al., 2009; Takhteyev et al., 2010). Undoubtedly space and time should be integrated in social networks for gaining more insights, however, traditional network graphs are limited to address the spatio-temporal problems (Shekhar and Oliver, 2011). Geo-researchers have made efforts to map networks integrated with space or space-time (Escher, 2007; Khalili, et al., 2009; Radil, et al., 2010; Shaw and Yu, 2009; Takhteyev, et al., 2010; Timo, 2006). For example, Timo (2006) developed a concept, which can allow us exploring the relationship between social networks and travel over time and space; Radil et al.(2010) spatialized network data by embedding social network graph in 2D map to understand the overall context of gang violence; Khalili, et al.(2009) considered the geography on the social network of randomly selected Flikr members. And one example which name is Twitter Conversations Map (Rafelsberger, 2008) found on VisualComplexity.com and from this map we get the conversation among 1500 users at different locations. Nonetheless, none of them can handle both spatio-temporal characteristics and internal properties of social networks simultaneously. Therefore it can be seen that there exists a gap between spatio-temporal data representation and traditional social node-link diagram. The geovisualization environment in this case can be used to link these two since it can integrate different visualization approaches from different disciplines to provide theory, methods and tools to support visual thinking and exploration about geospatial patterns (Dodge et al., 2008; Kraak, 2003a). Moreover, it has been applied in the field of social science (Kwan and Lee, 2004) and furthered to handle spatio-temporal network data in 2D map and space time cube(Kraak, 2010; Yang, 2011).

This research will be based on related work and try to design a visualization environment to visualize and explore both spatio-temporal characteristics and social structures of social networks.

1.5. Methodology

(1) Literature review
The literature review will be carried out on:

- The concepts of social networks and social media
- The evolution of social networks in the era of social media
- The existing methods of representing social network data
- The concept and models of spatio-temporal data
- The existing methods of representing spatio-temporal data

(2) Analyse and extend social network data
By understanding basic features and spatio-temporal characteristics of social networks, the triad geo-data framework model will be used to extend social network data from the geo-information perspective.

(3) Design a conceptual model to represent spatio-temporal social network data
A conceptual framework will be deduced from the study of literature review, in which the suitable graphic representation methods and function tools are selected.

(4) Design the prototype by using two case studies

(5) Test the designed prototype and evaluate the usability

(6) Discuss the results and draw conclusions and recommendations.

1.6. Structure of the thesis

Chapter 1 introduces the background, research objectives, research questions and methodology of the research.

Chapter 2 introduces the basic concepts of social networks and illustrated how the social networks developed in the era of social media.

Chapter 3 reviews the existing visualization methods for social network data. The review starts from introducing social network into Peuquet Triad framework and then based on the framework summarized the existing methods in both network and geospatial domain.

Chapter 4 designs a conceptual model for representing the spatio-temporal social network data. A user task space is proposed and based on the task space, suitable graphic representations and function tools are selected.

Chapter 5 describes the design the prototypes based on the conceptual model by means of two case studies: Facebook friend network for user-centric network and Twitter trend topic for object-centric network. The development of the designed prototypes consisting of the implemented graphic representations and design working environment is also described.

Chapter 6 illustrates the evaluation of the designed prototypes. It describes how the focus group method used in the usability test and what results and feedbacks from participants were obtained.

Chapter 7 draws the conclusion of the research and outlines the recommendations for the future work.

2. SOCIAL NETWORKS

2.1. Social networks

2.1.1. Basic concepts

Social networks are defined as "a set of people who share a common interest and have connections of some kind" (Wasserman, 1994). Therefore they are generated from the collection of connections among people. Ever since people communicated or exchanged something with others, social networks occur although they are invisible.

2.1.2. Review of social network researches

Social network analysis is a key area in sociology. By adopting from network data model, social network data can also be stored and viewed in a node-link form in which nodes represent individual actors (people, organization) and links represent relationship (kinship, language, trade, exchange etc.) or interaction (communication, exchange etc.) between a node-pair. It aims to analyze the structure of relations between actors in a social network that enables people to understand and communicate a wealth of information inside a social network (Scharl and Tochtermann, 2007; Valente, 2010). Over the past decades, researchers in this field have developed many creative theories, methods and techniques to study the patterns of connections in this complex system. One classic example is the theory of small world phenomena by Milgram (1967), who hypothesized that each actor in a social network is linked to any other with a maximum of 6 intermediaries; many mathematicians and statisticians evaluated the value of some criteria (centrality, degrees etc.) of the network to detect important individuals, relationships and clusters ; and also social networks were applied in many application fields such as business marketing (Anderson et al., 1994), human resource management (Collins and Clark, 2003) public health (Rothenberg et al., 1998) and scientific citation (Barabâsi et al., 2002) etc.

2.2. Social networks in the era of social media

2.2.1. Social media

With the advent of Web 2.0 and computer technologies, social media as the internet-based social interaction applications make billions of people create and exchange the content generated by themselves to facilitate their socialization (Hansen et al., 2009; Kaplan and Haenlein, 2010). Nowadays, it becomes a complex collection which contains email, mobile short text messages, social sharings, blogs and podcast, collaborative authoring, discussion groups, social networking sites and location-based services etc. (shown as Figure 2-1).

COLLABORATIVE AUTHORING	
Wiki	Wikipedia, Wikia (Lostpedia), pbwiki, wetpaint
Shared documents	Google Docs, Zoho, Etherpad
BLOGS AND PODCASTS	
Blogs	LiveJournal, Blogger, WordPress
Microblogs and activity streams	Twitter, Yammer, Buzz, Activity Streams
Multimedia blogs and podcasts	Vlogs (video blogs such as Qik), photo blogs (Fotolog, FAILblog.org), moblog (mobile blogging such as moblog.net), podcasts (iTunes, NPR)
SOCIAL SHARING	
Video and TV	YouTube, Hulu, Netflix, Vimeo, Chatroulette
Photo and art	Flickr, Picasso, deviantART
Music	Last.Fm; imeem; Sonic Garden
Bookmarks, news, and books	Delicious, Digg, Reddit, StumbleUpon, Goodreads, LibraryThing, citeulike
SOCIAL NETWORKING SERVICES	
Social and dating	Facebook, MySpace, BlackPlanet, Tagged, eHarmony, Match
Professional	LinkedIn, Plaxo, XING
Niche networks	Ning (e.g., classroom 2.0), Ravelry, Grou.ps
ONLINE MARKETS AND PRODUCTION	
Financial transaction	eBay, Amazon, craigslist, Kiva
User-generated products	Instructables, Threadless, TopCoder, Sourceforge, Codeplex
Review sites	ePinions, Amazon, Angie's List, Yelp
IDEA GENERATION	
Idea generation, selection, and challenge sites	IdeaConnection, Chaordix, IdeaScale, Imaginatik
VIRTUAL WORLDS	
Virtual reality worlds	Second Life, Club Penguin, Webkinz, Habbo
Massively multiplayer games	World of Warcraft, Lord of the Rings Online, Aion
MOBILE-BASED SERVICES	
Location sharing, annotation, and games	Foursquare, Gowalla, Loopt, MapMyRun, Geocaching, Letterboxing, SCVNGR

Figure 2-1: Types of social media listed with example services (Hansen et al., 2009)

2.2.2. Social media data

Social media data generally is the data we generated through social media. To be specific, social media data contains mainly four types of information as Figure 2-2 shows: profile, people, interaction and content. Profile is personal information (name, birth, sex, education etc.) users provide on the web like Facebook personal webpage, Twitter Bio etc.; people can be friends on Facebook, the followers on Twitter, the subscribers on Youtube etc.; interaction refers to the visits to the friends' 'wall' (personal webpage), the press on 'like' button (Facebook), the comments or views to a blog and the re-tweets to a tweet etc.; content refers to the text and media that user-generated covering message, tweet, photo, video and even location.

Figure 2-2: Social media data (source: Author)

2.2.3. Social networks from social media data

Today, new network science concepts and analysis tools can already make the hidden ties that link each of us to others become more visible and machine readable (Hansen, et al., 2009). From the social media data, the friends we make and the content we 'like' or comment or 'retweet' can all be recorded as connections among people and/or objects. Therefore, since the mode of the formation of connections has been dramatically changed, social networks built through social media are in detail and scale never before seen (Barbier and Liu, 2011).

By means of social media, nodes and links of social networks are different from common ones to some degree (Hansen, et al., 2009; Smith et al., 2009). Nodes can be people or objects. Objects means besides representing people, nodes can also be other entities such as web pages, digital media and even physical locations or events; links can take form of relationship or interaction. The relationship only connects two people; the interaction can connect two people, or people and content like digital media. Specifically, the relationship between people can be multiplex. For example, Twitter has three types of relationships: following, reply, mention. The interaction between two people can be sending a message or visiting the personal webpage; the one between people and content can be pressing a 'like' button to one's photo, retweeting one's tweet or commenting on one's blog.

There are two types of social networks from social media data nowadays known and used by most of people: user centric network and object centric network. User centric network is the social network that develops around one user and his/her friends, such as Facebook, MySpace and LinkedIn etc. Object centric networks, on the other hand, develop around interactions from one digital social object—such as Flickr, which has formed communities around photo-sharing and Twitter, which can organize collective conversion by tweets and retweets around one trending topic (#hashtag).

2.3. Social networks in space and time

2.3.1. The geo-component in social media data

With the development of location-acquisition technique, social media become increasingly geographic(MacEachren et al., 2011). In social media data, geographic information takes forms of text and GPS coordinate pair. The former one exists in user's profile which shows where the user is from and the posted text-based information like status (Facebook) or tweet (Twitter) which may contain geo-names or other location-related content; the latter one is becoming popular in social media in recent years with the advances of geo-tagging technology both in PC and mobile device. Not only do people post location-related information through computer, phones and cameras equipped with low-cost GPS chips equipping can also allow people record locations while taking pictures and videos and post them onto the social media platforms (e.g. Flikr, Youtube etc.). Moreover, location itself can also be a criterion for people interacting with each other. Foursquare and Facebook let people only post check-in points for interacting with others. The table below illustrates the geo-components contained in 5 most popular social media sites.

Geo-components / Social media sites	User profile	Content
Facebook	Hometown; Current city (city level)	Geo-tagged photo/video/status, place name in status, check-in point
Twitter	User's location (city level)	Geo-located tweet or place name in

		tweet
Flikr	Hometown; Current location (city level)	Geo-tagged photo
Youtube	User's location (city level)	Geo-tagged video
Foursquare	User's location (city level)	Check-in point, geo-tagged status/photo

Table 2-1: geographic component in social media data

2.3.2. Understanding social networks from social media data in a spatio-temporal context

The geo-components of social media offer the potential to physically comprehend social networks built up in this virtual space. For instance, Escher (2007) plotted Facebook friends' networks on the map by using the place name indicating where friends come from to indicate that our online friendships have a 'local focus'; Takhteyev et al.(2010) analyzed the geography of Twitter networks and found that physical distance can influence social ties over Twitter space. Furthermore, this idea can be strengthened by including time. Understanding the spatio-temporal dynamics of social networks undoubtedly can provide people powerful new insight. Goodchild and Janelle (2010) demonstrated that the spatial-temporal dynamics in social science can refer to two aspects: individual movement and information diffusion. This social-spatial thinking can also be applied into the social networks from social media data. Specifically speaking, for the user-centric network, we can deepen our insights by providing the view of the user's changing location and the corresponding changes of his/her friends' composition and spatial distribution with the elapse of time; for the object-centric network, by monitoring the process of its spatio-temporal changes, we can detect the way how the information spread by region and also other information like where are the origin and destination of the information etc..

2.4. Summary

This chapter briefly introduced the concept and current situations of social networks. Two most popular social networks from social media data are presented: user-centric network and object-centric network. By introducing the geo-component in the social media data, the chapter also discussed the opportunity of physically understanding social networks, and strengthened this idea by including time.

3. VISUAL REPRESENTATIONS OF SOCIAL NETWORKS

3.1. Introduction

Visualization can be seen as an effective way to help people understand social networks and convey the result of analysis (Freeman, 2004). This chapter aims at reviewing social network visualizations and focus on the node-link form. As mentioned in last chapter, there is a demand for understanding social networks in a spatio-temporal context. The chapter firstly introduces social network data into spatio-temporal data triad framework and then conducts the review based on this framework to clarify to what extent existing methods in both network and geospatial domain have been done so far. Related implementations in social media case are also involved. At the end of the chapter geovisualization methods are introduced and the considerations for social networks by using these methods are also proposed.

3.2. Peuquet Triad framework for social network data

Same as other network data, social network data contains two elements: nodes and links (also called segments). These data elements can be stored and represented as points and lines. As illustrated, since we want to study spatio-temporal characteristics of the social networks, social network data can be introduced in a triad framework proposed by Peuquet (1994). As shown in lower left figure, three components are distinguished: attribute, time and location and meanwhile relations exist in every two components. Specifically, these three components in social network data can be denoted as the right image in Figure 3-1, attribute refers to observed or collected qualitative values or quantitative values both in nodes and segments; time refers to temporal instants or intervals; location here not only refers to geographic location because when studying social networks in a non-geographic space, nodes still can be judged or measured by where they locate on the network. For example, in some graph layout of social networks, the more central place a node lies at in the network, the more important the node is. We define this location as graph location. This data framework will be used to logically organize the review of the visual representations of social network data throughout this chapter.

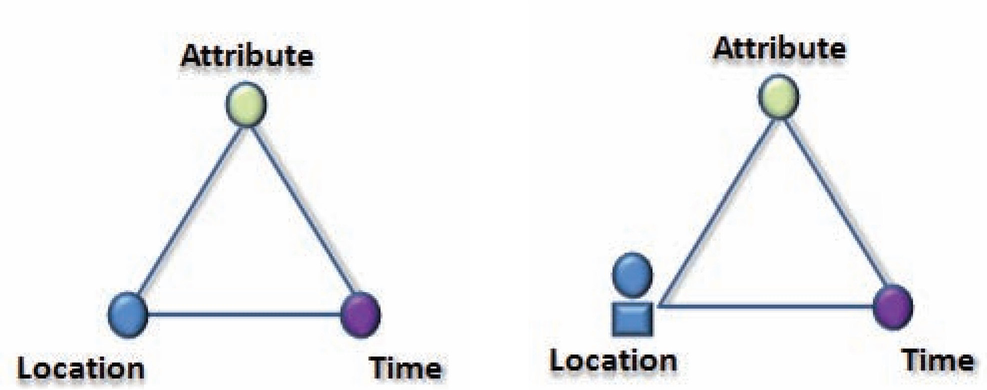


Figure 3-1: Triad framework for social network data

(Green circle: attribute; purple circle: time; blue circle: geographic location, blue rectangle: graph location). Left: Peuquet (1994) triad framework; Right: the triad framework for social network data.

3.3. Social network data visualization

Social network visualization simply is a graphical representation of a social network which translates abstract information in a social network into geometric representations(Ing-Xiang Chen, 2010). It becomes an important tool to allow people visually gain insights of the structure and dynamics of social networks (Carlos D. Correa, 2011). There are two distinct forms of representations have been used to display social network data, one is node-link diagram and another is matrix-oriented method. Traditionally, social networks are visualized as a node-link diagram since it is the most direct and intuitive way to visualize networks. The research focuses on the node-link form of social networks.

Visualizing social network data is not trivial due to its complexity. To meet the visualization need, one can represent data based on each component of the triad framework (X. Li and Kraak, 2008). Therefore, based on this theory social network data can be represented from three perspectives which are attribute perspective, temporal perspective and locational perspective. From the attribute perspective, visualization needs to denote the qualitative or quantitative values of both nodes and links. This can be done by applying graphic variables (size, color, shape, texture, value and orientation) based on Bertin's (1983) theory. For example, one can use different size of nodes to represent the degree of importance or use different color of links to represent different types of relationship. From the temporal perspective, temporal elements can be represented linearly. Basically, there are three approaches to visualize changing networks. The first method is using graphic variables to depict changes; Second one is adding time as a new dimension so that visualizing network in a 2.5 or 3 dimensional way (Gaertler and Wagner, 2006); third one is applying series of snapshots or graph animation techniques. From locational perspective, we need to distinguish the two terms: graph location and geographic location. For graph location, nodes and links are placed according to their connectivity (e.g. the most connected people is always placed in the central position) or other characteristics (e.g. in a hierarchical diagram, children are placed lower than their parents) through numerous network layout algorithms; for geographic location, nodes are fixed by physical location on the earth, map in this case is always used as the base.

In practice, visualization methods for social network data are always take forms of combinations of two or three perspectives. Existing social network visualizations mostly combine attribute and locational perspectives for static network, and combine all perspectives for dynamic network. However, there is a gap for considering geographic location since most of social network visualizations only consider graph location. For taking geographic location into account, physical network visualization method can be introduced. But in this case there is no existing method for combining all three perspectives. Geovisualization is considered to have the potential to meet this gap. The following content will discuss the above situations.

3.3.1. Social node-link graph

Traditionally, social networks are visualized in the graph by using a number of network layout algorithms. It can combine two or three perspectives to both represent static network and dynamic network. Since these methods visualize the networks on the graph, the location is used as the graph location.

3.3.1.1. For static social networks

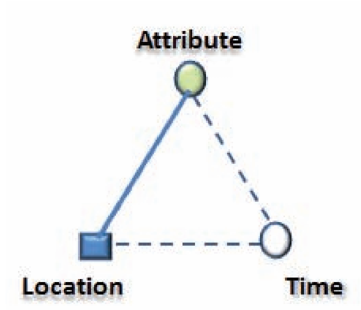


Figure 3-2: Static social network data with graph location:

(1) Graph layout
Random layout

A random layout is to simply place the nodes at randomly computed positions inside a user-defined region. Díaz et al.(2002) categorized 3 models of random graphs which are binomial random graph, random grid graph and random geometric graph (Figure 3-3).

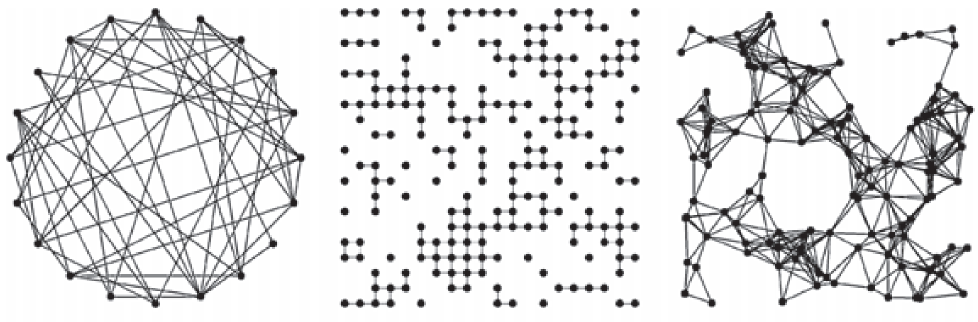


Figure 3-3: Random layout (Díaz, et al., 2002) Left: Binomial random graph; middle: random grid graph; right: random geometric graph.

Advantages:

1) It can efficiently draw the social network graph in linear time.

2) It could be useful to visualize very large network graphs in cases there is no need for an aesthetic, readable drawing.

Disadvantages:

1) It may lead to difficulty on producing useful results when nodes increasing dramatically.
2) Links may cross heavily in a complex social structure.

Force-directed layout

The force-directed layout is to simulate the nodes and the links as repelling objects (Coulomb's law) and springs (Hooke's law) in the network graph. When nodes and links firstly generated, this algorithm assigns forces among nodes and links to pull them together or push them apart and repeat iteratively until the situation comes to the equilibrium state, which refers to all graph nodes and attractive forces between the adjacent nodes run to convergence (Ing-Xiang Chen, 2010).

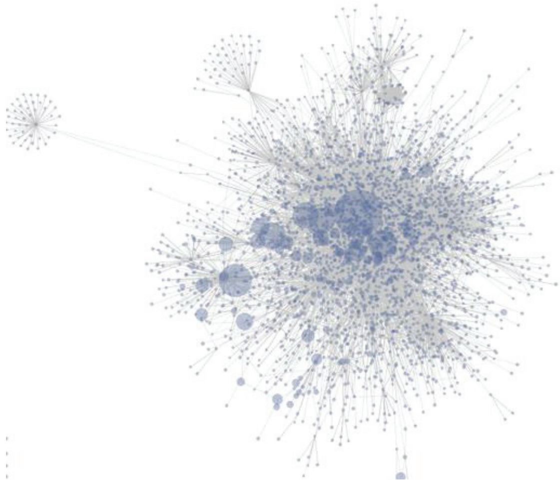


Figure 3-4: Force-directed layout (source: Wikipedia)

Advantages:

1) It is easy to implement.
2) It can be quite effective for spatially grouping connected communities
3) Allow aesthetics rules to achieve good and clear graph output.

Disadvantages:

1) It would be time consuming to obtain a stable state especially when dealing with large scale networks.
2) A "hairball" view is always obtain for most of networks with a moderate size (Carlos D. Correa, 2011).

Circular layout

Circular layout is the most traditional methods used to draw graphs, which places all the nodes on the periphery of a single circle and links connecting these nodes passing within the circle (Bertin, 1983). The circular layout can not only deal with one single circle layout, but also can place multiple circles together. In general, these layouts can provide a compact presentation, focusing on individual nodes and links.

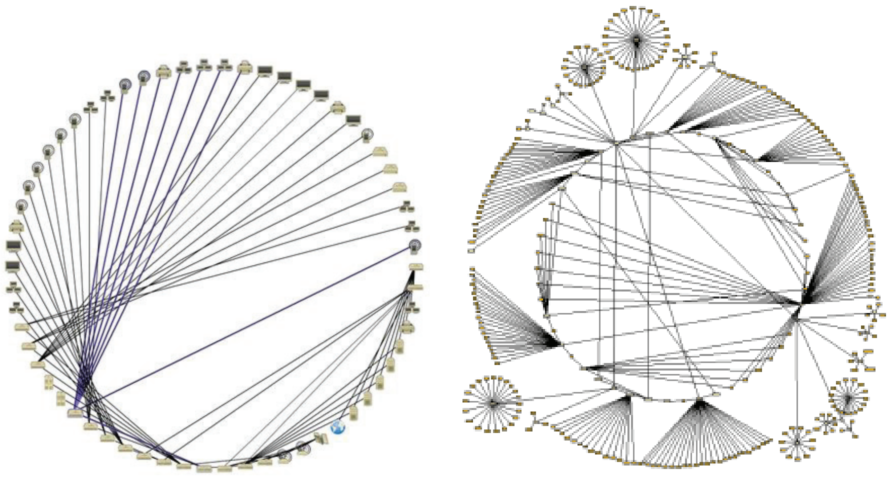


Figure 3-5: Circular layout; (source: Internet). Left: single circle layout; right: multiple circles layout

Advantages:

1) A node cannot be occluded by another node or by a link.
2) Links will not obscure each other.

Disadvantages:

1) Strong regularity can make other information not clear.
2) More links would lead to a heavily crossed view.
3) Not suitable when drawing large scale networks.

(2) Tree layout

The tree layout is originated from the idea that placing the nodes with the root node in the centre, and then other nodes connected to the root node form a circle around that (Reingold and Tilford, 1981). A classical tree layout follows a top-down hierarchical mechanism (Figure 3-6). Thus nodes that are at one level away from the root become the children of the root and so on and the space between nodes reflects the number of nodes in the sub-tree generated from that node (Walker and John, 1990).

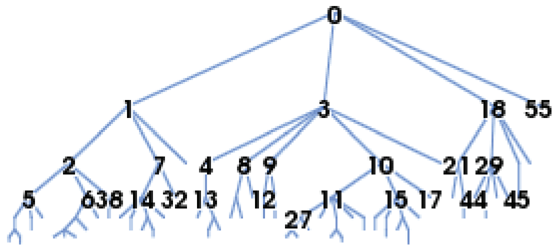


Figure 3-6: Standard tree layout (URL: http://www.kitware.com)

Tree layout forms can be varying regarding different domains of information. Some examples are:

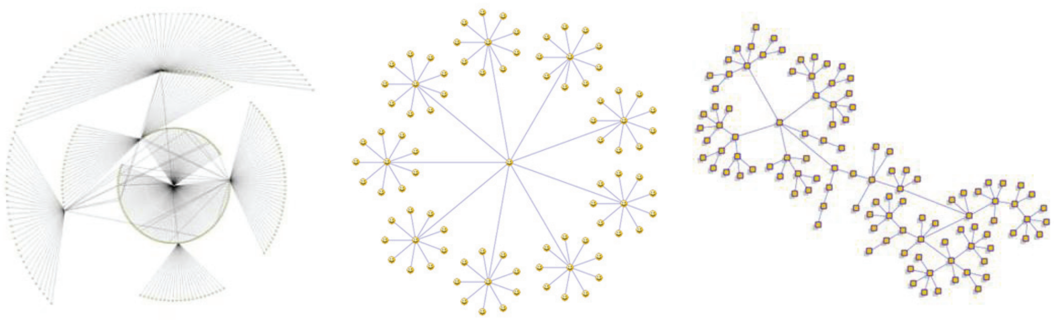


Figure 3-7: Examples of the variation of tree layout; source: (Hong et al., 2009; Technologies, 2003) Left: radial layout; middle: balloon layout; right: wedge layout

Radial layout:
Wills (1999) described the main idea of this layout algorithm as follow: "Given a focal point for A, and any node R, the structure of spanning tree needs to meet that the conditions that the distance from A to R in the tree should be the shortest path among each two points in the graph". In short, this algorithm is to place nodes in a circle and links are drawn as secant lines through the circle (Carlos D. Correa, 2011). Radial layout is suitable for the network which contains large amount of nodes with tree structure and relatively small dense links.

Balloon Layout: places sub-trees in circles around the parent node in a balloon-like pattern.

Wedge layout: "places sub-trees in separate sectors (wedges) around the parent vertex in a circular fashion. The angular width of the sectors is assigned according to the sizes of the sub-trees" (Technologies, 2003).

Advantages:

1) The inherent hierarchical nature enable the tree layout to be more structural and straightforward and provide more contextual information and facilitate network analysis (Ing-Xiang Chen, 2010).
2) It intuitively describes the social distance from a centred node.
3) It can use the space effectively and suitable for large networks.

Disadvantages:

1) This layout may not suitable for the networks with complex structure.
2) The effectiveness of this layout relies on too much on the centred node (Hong et al., 2009).

3.3.1.2. For dynamic social network

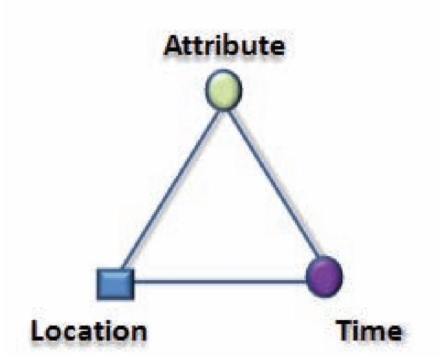


Figure 3-8: Dynamic social network data with graph location

Network's evolution and dynamic are always the topic that captures the attention of visualization researchers. In order to capture the development and changes of a network, the temporal perspective is added and the changes of attribute and geometric location over time are able to detect. As mentioned, time can be represented as variables, one dimension and series of snapshots. Here we use the work done by Erten et al. (2004) as a instance. Regarding the method of different graphic variables, they applied different sizes for nodes and different colours for links to display the different states of cumulative citation network from 1994 to 2002; for 3D view, the evolution of networks displayed vertically by using 3 different time slices where each time slice represents 3 consecutive years; for the screenshots from the animation technique, the evolution of collaboration networks in the period 1994-2002 shown in nine time slices where each time slice represents one year period.

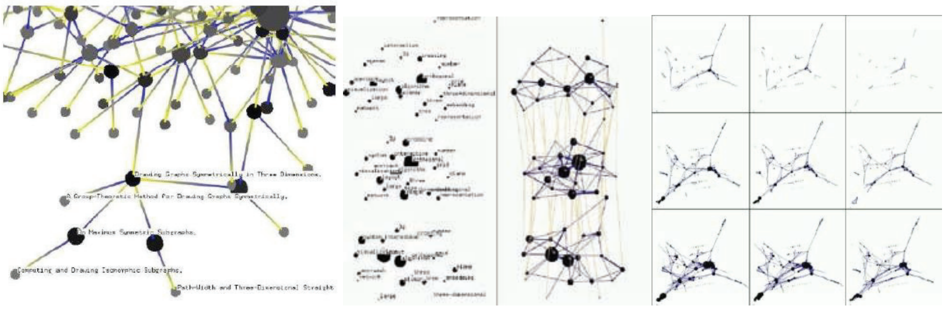


Figure 3-9: Dynamic social network visualization methods (source: Erten et al. (2004))

Related implementations from social media

- **Static network visualizations**

TouchGraph

TouchGraph is a modern Facebook visualization that is based on node-link representation and uses force-directed layout algorithm. It allows users to see how their friends are connected, and who has the most photos together. Users can also explore their own personal networks by graphing photos from anyone's album, or view the connections between members of a group. As figure 10 shows, network members are presented using both their self-provided name and, if available, a representative photograph or image. And this visualization also has interactive functions (zoom and spacing) to help user overcome the line intersection problem.

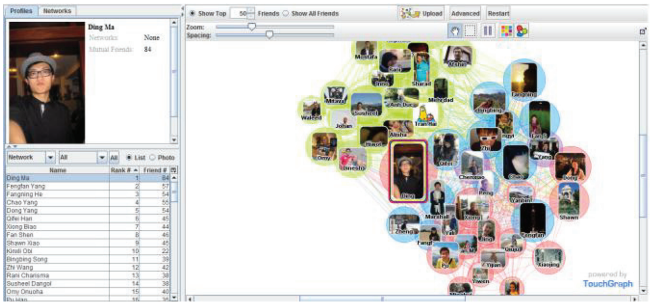


Figure 3-10: Visualize Facebook social relationship by TouchGraph

(URL: http://www.touchgraph.com/facebook)

Mentionmap

Mentionmap displays networks from Twitter by using node-link diagram and tree layout algorithm. It utilizes Twitter API to allow people search for any user they want and have a visual on the network generated from this user. The end node from the user can be a user or hashtag. The lines drawn between nodes can indicate people how many times users mention each other. Additionally, clicking a user will display their network of mentions as well as details from their profile. It can help people to discover who they interact the most and what they're talking about. It can also be an interesting way to find relevant people to follow.

Figure 3-11: Mentionmap
(URL: http://apps.asterisq.com/mentionmap/)

- **Dynamic network visualizations**

Virginia earthquake in the hashtag cloud

This visualization represents Twitter hashtag network changing over time. The trending topic was about a minor earthquake in Virginia. The three views as Figure 3-12 shows are the tracks of the development of this topic based on a real-time analysis of hashtag activity on Twitter. The node's size represents the number of retweets of the hashtag of the activity. Links are created if hashtags occur in the same tweet.

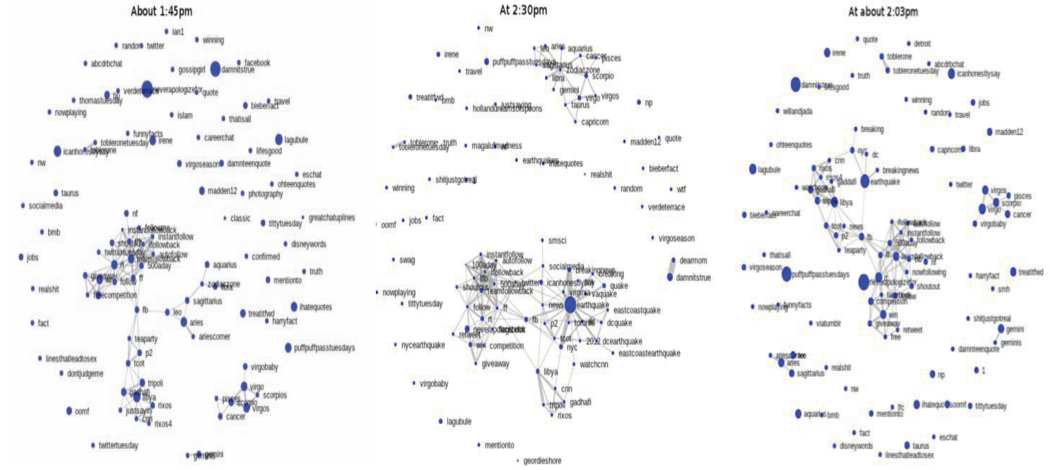


Figure 3-12: Dynamics of Twitter hashtag network

(URL: http://twimpact.tumblr.com/post/9335320611/last-days-virginia-earthquake-in-the-hashtag-cloud)

3.3.2. Visualizing the geographic location of social networks

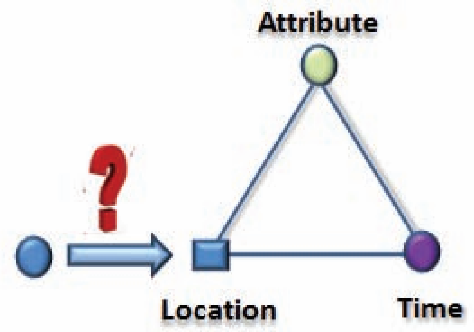


Figure 3-13: How to represent location information of the social networks?

Considering representing geographic location information (Figure 3-13), qualitative graphic variables can be applied to the traditional social network graph. For example, we can use different colours to indicate different places. However, since the limitation of human visual system, the maximum number of the qualitative variables for points and lines is seven (Kraak and Ormeling, 1996). Moreover, the graph cannot deal with the movement of nodes. Therefore social graph lacks ability of representing geospatial information.

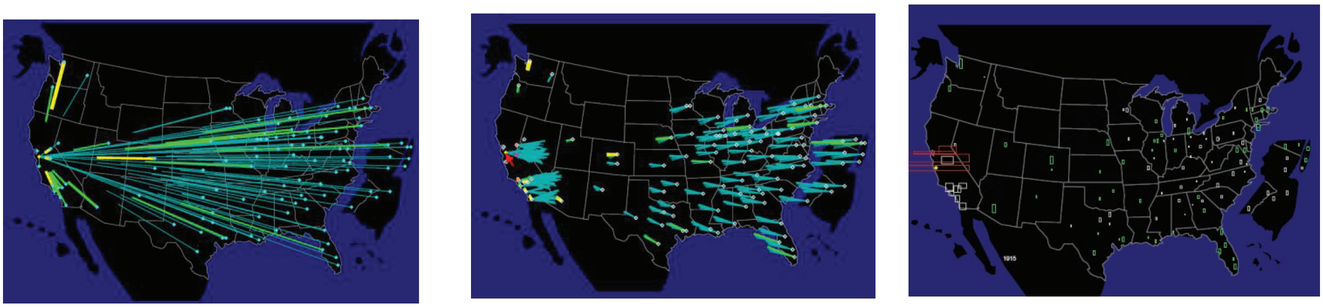
Basically, for representing geographic location of social networks, map is the first option. Nodes are fixed on the map by its location information. In this case, the visualization of social network could be similar with other geographic network representations. For introducing geographic in social networks, geographic network visualization methods should be studied. The most basic ones are link map and node map which were proposed by Becker et al (1995).

Link map

This uses line segments between each pair of connected nodes in two dimensions. Colour and line thickness can be used to represent the attribute information about the links. For larger networks, the crowd of links may cause the clutter problem of the map. This can be partially overcome by drawing the interesting links or by only drawing partial links.

Node map

This method displays each node on the map by node oriented information using simple glyphs or symbols such as circle or rectangle. Different orientations of symbol can represent different kinds of information. Different symbols convey different degrees of dimension of information, e.g. circles can be used for one dimensional information, rectangles for two and more complex ones can refer to higher dimensions. Additionally, colour and size and can also be used to add to show other interesting attributes.



a) b)

Figure 3-14: Geographic network map (source: (Becker et al., 1995)) a): link map; b) node map

Current researches also pay efforts on spatializing network node-link layout and then implement relevant analyses, namely consider visualizing spatial or spatial-social networks from a spatial perspective. Radil et al. (2010) used spatialized social network data to investigate relations among gang rivalry, territoriality, and violence in Los Angeles. Guo (2009) proposed an integrated interactive visualization framework to effectively visualize network structures to discover the patterns and relations from country-to-county migration data in the U.S.

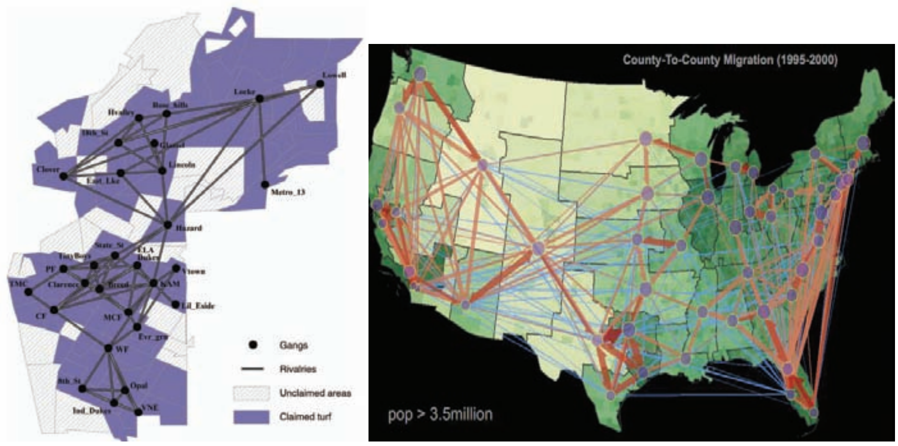


Figure 3-15: Current research of mapping network data (source: (Guo, 2009; Radil, et al., 2010))

Related implementations from social media

Mapping Facebook friendship

This visualization aimed at mapping the connections of 10 million Facebook users. Locating each user on the map is according to the coordinates of the city they are from. The highlighted area shows a high density of Facebook users.

Figure 3-16: Mapping Facebook friendship

(URL: http://www.facebook.com/notes/facebook-engineering/visualizing-friendships/469716398919)

However, visualizing social networks on the map is only from locational perspective. To date, existing social network visualization methods cannot combine time with other two perspectives together when location component refers to geographic location. In this case, time cartography could be an option for representing the dynamic social networks. The following content will review the methods of time cartography and discuss the potential chance of mapping the changing social networks.

3.3.3. Time cartography for dynamic social networks

3.3.3.1. Single static map

Single static map is the simplest visualization method for spatio-temporal data. It uses graphic variables and symbols to depict changes when representing an event (Kraak and Ormeling, 1996). The differences in size, color, orientation and shape etc. can all describe the changes in one static map.

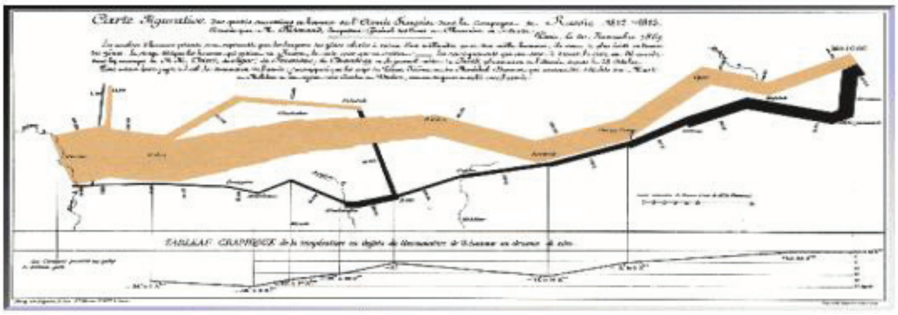


Figure 3-17: Single static map

(URL: www.itc.nl/personal/kraak)

- **For dynamic social networks**

Single static map is suitable only when compare 2 states of network change. Graphic variables in this case are capable of describing the change degree of both node and link. But it is hard to visualize the appearance or disappearance of node or link and to map the changes if the nodes or links are too many.

3.3.3.2. Series of static maps

The static map series is to put individual maps in the temporal sequence by a spatial sequence to depict a process of change (Kraak and Ormeling, 1996).

Figure 3-18: series of static maps (source: lecture handout of Kraak 2011)

- **For dynamic social networks**

It is an optimal option when comparing different states of network in a short series. And it also requires that the time of network change is discrete.

3.3.3.3. Animation map

Animation map can represent the dynamic characteristics of geodata in animated view or show spatial information dynamically in a sequence of static maps. It seems like a very suitable mode when visualizing spatio-temporal data.

- **For dynamic social networks**

It is able to visualize the changing social networks in a long series and people can use interactive tools to manipulate the process of playing. However, it is hard for people to intuitively compare two states of network where a big time interval exists. And considering the limitation of human visual system, it is difficult to make people perceive the location change and attribute change of a number of nodes simultaneously.

3.3.3.4. Space-time cube

In the late of 1960 Hagerstrand introduced Space-Time-Cube to represent people in time and space. It is the most prominent element in Hagerstand's space-time model (Kraak, 2003b). The STC combines time and space in a natural way: time can be represented as continuous or discrete in the Z-axis. The X and Y axis indicate the 2D space. The position of the object in STC is a point. It means that an object exists in one position at one time point. When visualizing an event in the STC, the trajectory of an object is always displayed as a line which is called Space-Time-Path. This geovisualization method would benefit greatly from interactive options when manipulating the viewer's perspective of the cube(Kraak, 2003a).

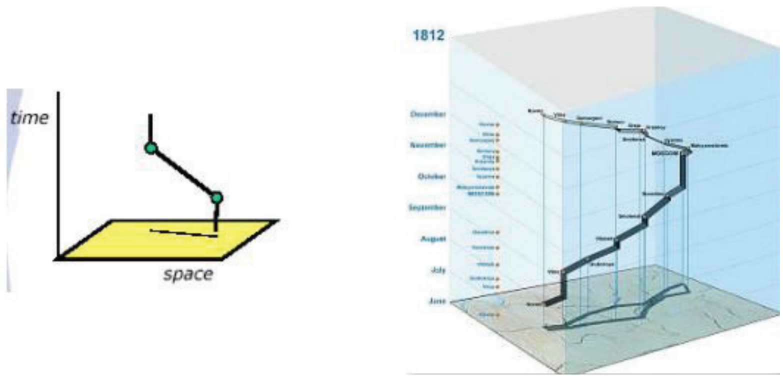


Figure 3-19: Space-time Cube (source: lecture handout of Kraak 2011)

- **For dynamic social networks**

It is capable of depicting the changing social networks both in discrete and continuous time. Every two time slices of network can be compared. And it also can visualize the movements of each node. However, if only 2 or 3 states of network need to be compared, STC is not an optimal option where series maps are more suitable.

3.4. Summary

This chapter reviewed the existing visualization methods of the social network data. Since representing social network data is not trivial, the chapter started with introducing the social network data into the Peuquet triad framework and broke down the complexity of social network data from attribute, time and location perspective. Location here referred to graph location and geographic location. By using these two types of locations, the chapter review the graph and map visualizations for social network data in both static and dynamic cases. Based on the review in the last two chapters, next chapter will design a conceptual framework to visualize the spatio-temporal characteristics of the social network data.

4. CONCEPTUAL MODEL DESIGN

4.1. Introduction

Based on the description of social networks in chapter 2, and existing visualization methods of social networks in chapter 3, this chapter will discuss a conceptual model by using visualization theories to help user understand spatio-temporal characteristics of social networks. Starting point is the common approach (see Figure 4-1) for solving problem with visualizations discussed by Li and Kraak (2008). Based on this approach, the research designs the conceptual model which involves three components: user tasks, data framework and visualization framework. Since the data framework has been proposed in previous chapter, this chapter will mainly illustrate the design of the user tasks and the visualization framework. Furthermore, because this research focuses on using geo-location as the criteria to visualize social networks from social media data, the 'where' component in the data framework in this conceptual model only refers to geographic location. For user task design, I propose a social network task space containing questions formed by the data framework (where, when, what, whether) combined with social network data elements (node, link and sub-network). According to different user tasks in social network task space, different graphic representations are selected to meet each need; a working environment is design based on the function tools to make user effectively query and manipulate the selected graphic representations.

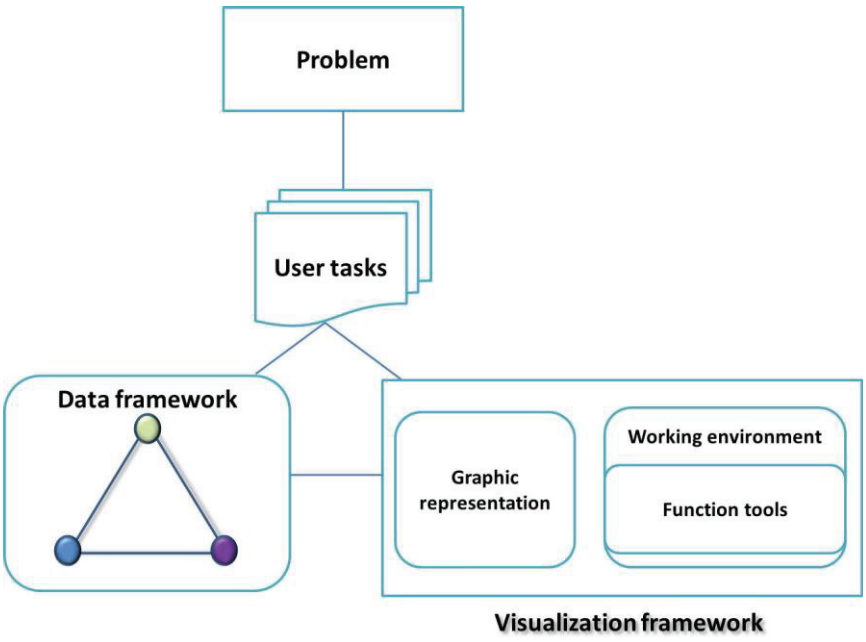


Figure 4-1: The conceptual model based on an approach to visual problem solving (source: Li and Kraak (2008))

4.2. User tasks design

4.2.1. Theoretical basis

User tasks can be translated into different questions from user side. Each question can be defined by the type of questioning component and its certain reading level as Bertin (1983) claimed. He introduced 3 reading levels for each type of question: elementary, intermediate and overall. Elementary level refers to a question concerns a single data element, intermediate refers to a group of elements taken as a whole, overall level refers to all elements constituting the object.

Specifically for spatio-temporal data, the triad framework which Peuquet (1994) proposed offers three components (where, what, when) to structure questions. Three basic questions used for querying the components of spatio-temporal data with different reading levels and their relations are as follows:

- When + where→what: Describe the objects or set of objects that are present at a given location or set of locations at a given time or set of times.

- When + what→where: Describe the location or set of locations occupied by a given object or set of objects at a given time or set of times.

- Where + what→when: Describe the time or set of times that a given object or set of objects occupied a given location or set of locations.

Concerning the questions of the object itself, Mennis et al. (2000) extended the Peuquet's framework into a 'pyramid' model where considers object on a higher level as a knowledge component and former three elements (what, where and when) as the data component. Based on this pyramid model, Xia Li (2010) added one question component 'whether' which takes object into account (querying the existence of the object). The basic questions for spatio-temporal data in this case are as follows:

- What + when + where→whether: Describe the existence of an object in a certain situation.
- Where + what + whether→when: Describe the time of an event.
- What + when + Whether→where: Describe the location of an event.
- When + where + whether→ what: Describe the character of an event.

Those four question components above are the basis of questioning and researching spatio-temporal data. Furthermore, if users want to describe dynamics in an event or phenomena, time can no longer be a single moment. By considering the changes, 'when' component is denoted as a set of times and accordingly 'what', 'where' and 'whether' can be referred to attribute change, location change and existential change (Andrienko et al., 2003; Blok et al., 2005; Xia Li, 2010).

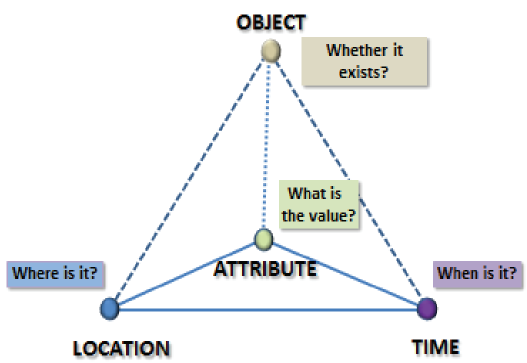


Figure 4-2: The pyramid spatio-temporal data model and related question components (source: Xia Li (2010))

The reviewed work is the foundation of user tasks in this research. In summary, Bertin's theory basically defined user tasks which are composed of components with different reading levels. As for spatio-temporal data, user tasks can be executed by four components: where (location), when (time), what (attribute) and whether (existence). As a result, all the questions from user side can be interpreted by these four 'atoms' in both static and dynamic situations.

4.2.2. Social network user tasks

As discussed above, we can form user questions for social network data by the four question components: where, when, what and whether. Because social network data contains different types of element, user's questions can target to any of them, for instance like 'what' questions, one would like to query one person's name (node), or the relationship type between two people (link) etc. For each type of social network data element there are corresponding questions organized by those four question components.

A social network task space is proposed as figure 4-3 shows to effectively link four user questions atoms with social network data elements. As mentioned, questions should relate to different types of social network elements. The "where", "what" and "whether" questions correspond to the elements of social network data. Since 'when' question is uniform for every data element and different levels of time (a single time moment or a set of times) can imply the whether the question is asking 'change', question "when" in this task space is described as a third dimension.

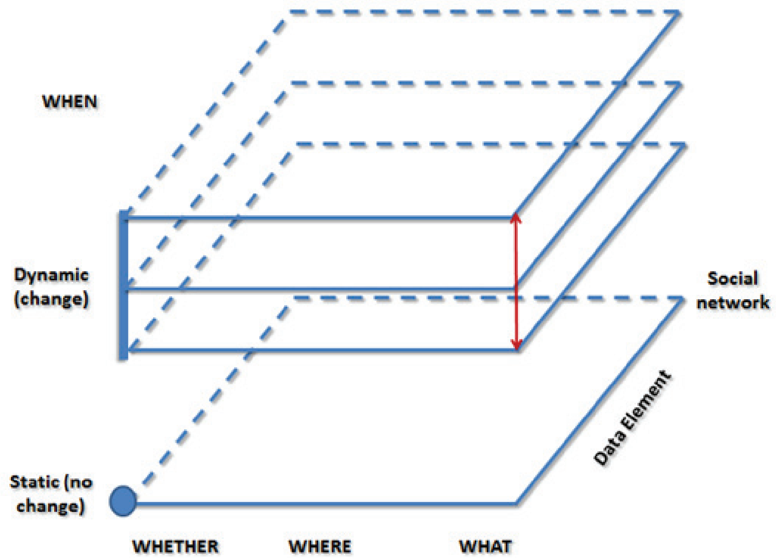


Figure 4-3: A social network task space from four question components (source: Author)

The task space can be elaborated as the figure 4-4 shows. As for the data element, social network data has three types of element: node, link and sub-network (Table 4-1). In this research I regard sub-network as one element type of social network besides node and link since people will not only be interested in nodes or links but also both nodes and links together. The sub-network should include at least two nodes and one link; the third dimension of time make the task space can not only include tasks for static social networks but also for the dynamic ones. Specifically, the "when" component include two reading levels of time on the third dimensional axis: single time moment and a set of times; with respect to one single time moment, "whether" refers to the existence of object; "where" is geo location; "what" refers to the attributes of object; with respect to a set of times, the "whether", "where" and "what" questions can be related to the changes of existence, geo location and attribute respectively. In addition, it is need to mention that different reading levels also can be used for social network data elements. For instance, as for sub-network, user can be interested in one single sub-network (elementary) or several sub-networks (intermediate) or the whole network (overall).

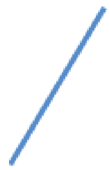
Original network	Node	Link	Sub-network
			

Table 4-1: Social network data element

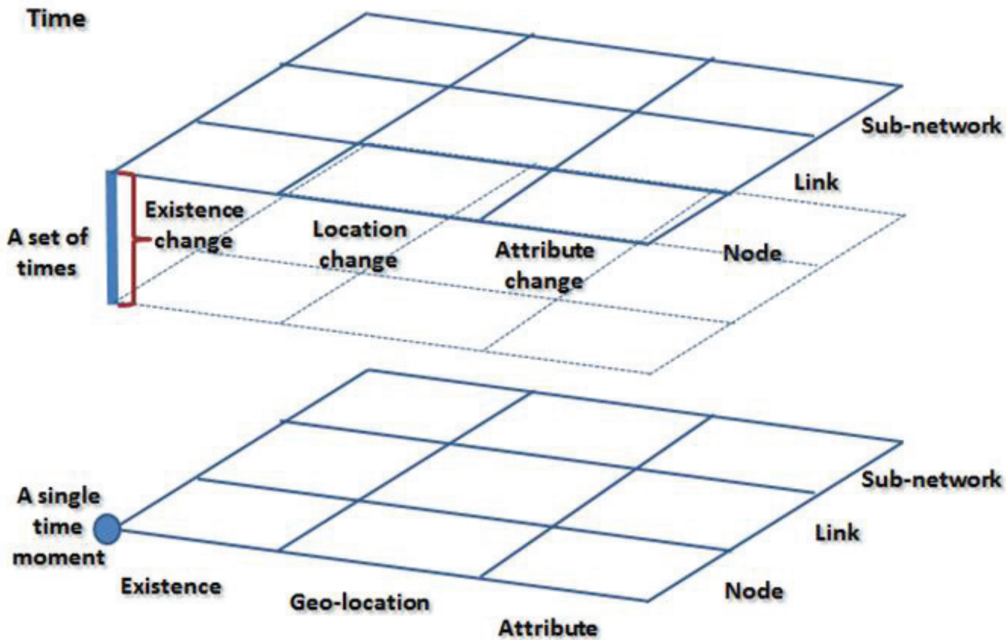


Figure 4-4: Elaborated social network task space (source: Author)

Specifically, each block in this task space holds information that users will ask in both situations of no change and considering changes. Based on this task space, user tasks will be categorized into static social network tasks and dynamic social network tasks. Just as the name implies, static tasks are focusing questioning the state of network element(s) at a certain time while dynamic ones are focusing the change of network element(s) during one time period. The detail will be discussed in the following 2 sections (4.2.2.1 and 4.2.2.2).

4.2.2.1. Static social network tasks

Since temporal information is known, “when” component used here is only as a constraint of question but not a question target.

Node:

- What + when+ where→whether: Describe the existence of a node or nodes at a certain situation.
- What + when + whether→where: Describe the location or a set of locations of a node or nodes at a certain situation.
- Where + whether + when→what: Describe the attribute(s) of a node or nodes at a certain situation.

Link:

- What + when→whether: Describe the existence of a link or links at a certain situation.
- When + whether + when→what: Describe the attribute(s) of a link or links exist at a certain time.

Sub-network:

- What + when + where→whether: Describe the existence of a sub-network or sub-networks at a certain situation.
- What + when + whether→where: Describe the location or a set of locations of a sub-network or sub-networks under a certain situation.
- When + where + whether→what: Describe the attribute(s) of a sub-network or sub-networks under a certain situation.

	Node	**Link**	**Sub-network**
Time	One single time moment	One single time moment	One single time moment
Existence	The existence of one node or nodes which meet the conditions	The existence of one link or links which meet the conditions	The existence of one sub-network or sub-networks which meet the conditions
Geo location	Node: one certain geographic location (place name, coordinates)	No relevant information	One certain location or spatial distribution (depends on scale)
Attribute	Qualitative attributes; Quantitative attributes	Qualitative attributes: type of link, directed or undirected Quantitative attributes;	Qualitative attributes: Quantitative attributes: number of nodes/links, connectivity

Table 4-2: Static question component for each social network element

4.2.2.2. Dynamic social network tasks:

Node:

- What + when + where→whether: Describe the existential change of a certain node or nodes over time.
- Where + what + whether→when: Describe the time period when a certain node or nodes meet a certain condition.
- What + when + whether→where: Describe the movement of a certain node or nodes over time.
- When + where + whether→what: Describe the attribute(s) change of a certain node or nodes over time.

Link:

- What + when→whether: Describe the existential change of a certain link or links over time.
- What + whether→when: Describe the time period when a certain link or links meet a certain condition.
- When + whether→what: Describe the attribute(s) change of a link or links over time.

Sub-network:

- What + when + where→whether: Describe the existential change of a certain sub-network or sub-networks over time.
- Where + what + whether → when: Describe the time period when a sub-network or sub-networks meet a certain condition.
- What + when + whether→where: Describe the spatial distribution change of one sub-network or sub-networks.

When + where + whether→what: Describe the attribute(s) change of one sub-network or sub-networks over time.

	Node	Link	Sub-network
Time	A set of times	A set of times	A set of times
Existence change	appearance/disappearance of node or nodes;	appearance/disappearance of link or links	appearance/disappearance of sub-network or sub-networks
Geo location change	movement	No relevant information	Spatial distribution change (node or nodes movement, geospatial range growth/contraction, scale of geo concentration increase/decrease)
Attribute change	Quantitative attributes change	Quantitative attributes change Qualitative attributes change	Quantitative attribute change

Table 4-3: Dynamic question component for each social network element

It is important to note that not all the question components on the left side of the arrow need to be used for every question. Either one or two components on the left side of the arrow can also form a question. For example, the question "Whether did John have friends in China in 2007?" is the case that follow the scheme "when + where→ whether" in which the 'what' component is not necessary to be involved.

4.3. Visualization framework

4.3.1. Graphic representations

After designing the user tasks of social network data, how to execute user tasks through visualization should be taken into account.

4.3.1.1. Graphic symbols and variables

Social network data in this study is only considered in a node-link form. Therefore, point and line are the symbols both considered in graph and map to represent social network data. The corresponding graphic symbols to different social network data elements are as Table 4-4 shows.

Social network data elements	Graphic symbols
Node	Point
Link	Line
Sub-network	Points and line(s)

Table 4-4: Graphic symbols for social network data element

Furthermore, Bertin's (1983) defined six graphic variables (size, value, grain, colour hue and shape) are used to depict data characteristics as figure below. Choosing different visual variables depends on the type of data (Kraak and Ormeling, 2003). Generally, shape, orientation and colour hue work for qualitative differences, while value (lightness) and size do for quantitative differences. The research will use size of point to represent different node's quantitative values (e.g. in a user-centric network, it can refer to number of friends), shape of node for different types of node (e.g. user or friend in a user-centric network) and colour of line for different types of links.

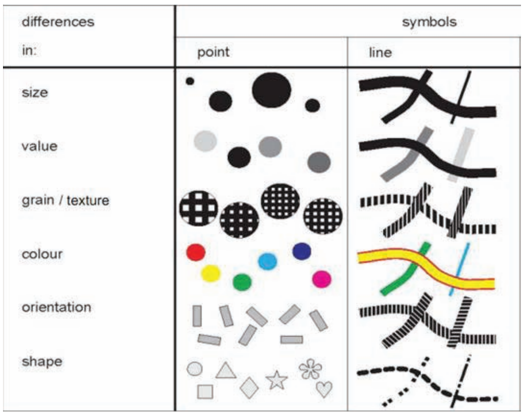


Figure 4-5: Bertin's visual variables for point and line symbols (source: (Bertin, 1983))

4.3.1.2. Choosing suitable graphic representation methods

After choosing symbols for social network data, suitable graphic methods are need to be considered in this phase. Earlier literature mentioned that selecting suitable representation methods depends on both data and user tasks. As summarized in last chapter, social network data constructed by nodes and links can be represented in two ways: graph and map. Selecting which method to use should be based on user tasks.

User tasks play a role of the selection of graphic representation because eventually the selected representation needs to answer questions. The table down below illustrates how graph and map can be used to answer user's whether, what and where questions in both static and dynamic cases. It can be seen from the table that neither graph nor map can solely answer all the questions. Therefore it can be concluded that in order to fulfil the needs from users based on the task space, both graph and map need to be taken into account.

Graphic methods / User tasks	Graph	Map
Static social network tasks		
Whether (existence)	Every node and link can be seen in the graph.	Have difficulties when the scale of location data is small (e.g. only country or city level), nodes and links would be possibly overlapped.
Where (location)	Can apply graphic variables (e.g. colour) to indirectly	Can easily show explicit location information for every

	location information for each node, but the number of colour is limited due to human visual system and cannot present spatial distribution of (sub) network.	single node and spatial distribution of (sub) network.
What (attribute)	Graphic variables	Graphic variables
Dynamic social network user tasks		
Whether (existential change)	Every appearance/disappearance of any node/link/sub-network can be detected.	Can be detected if there are no overlapped nodes or links.
Where (location change)	Cannot deal with movement but can somewhat handle network spatial range growth/contract (e.g. number of colour become more/less)	Can both represent individual node movement and network spatial distribution change
What (attribute change)	Changing visual variables	Changing visual variables

Table 4-5: Comparison between graph and map in static and dynamic tasks

Social network task space here is used again to choose proper visualization method based on the summary of visualization methods discussed in the last paragraph and Table 4-5. As results, Figure 4-6 shows that this study links related visualization methods with the block(s) in the task space where user's question locates. The blocks with different colour represent different choices of visualization methods. With respect to static social network tasks, since graph is able to place and show every node and link explicitly, it can be in charge of existence-related (whether) questions; map inherently represents data's geospatial characters, therefore map is used for location-related (where) questions; regarding attribute-related (what) questions, graphic variables can be used to represent data's qualitative and quantitative attributes but nothing for further detailed information, when further requests are needed, table or diagram can be considered. As for dynamic tasks, time for graph and map is represented can be series of snapshots or animation. The following sections (from 3.1.2.1 to 3.1.2.4) will elaborate these visualization methods used in this research.

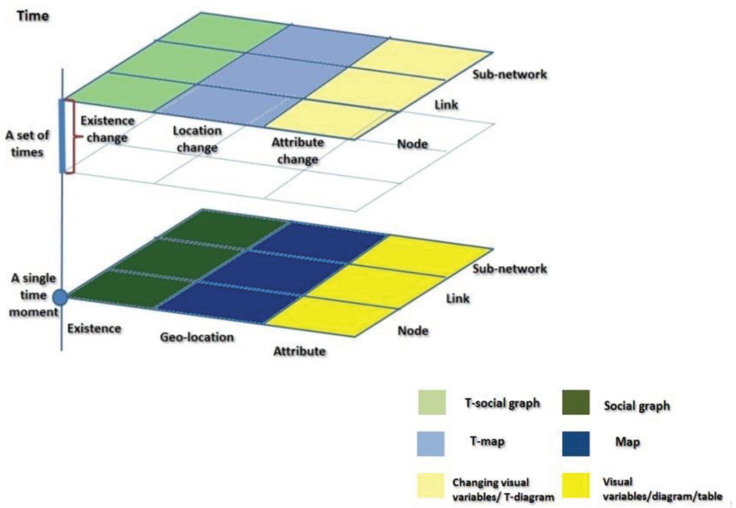


Figure 4-5: Selecting suitable representations for different type of tasks

4.3.1.3. Social node-link graph

In chapter 2, it has been discussed that there are many network node-link layout algorithms representing social network data. The choice is oriented by both data structure and user tasks. This means that the layout should on the one hand follow the network structure, and on another make every node and link clearly visible so that can effectively help user execute the tasks about existence. As results, circular layout is selected for user-centric network and tree layout for object-centric network.

- **Circular layout for user centric network**

User-centric network, as defined in chapter 2 "a network developed from online relationships around a user", consists of an ego (user), his/her 'alters' (friends, followers etc.) and the relationships between them. Therefore, in accordance with the inherent self-centred structure, one rational arrangement of the structure is circular layout, which follows a star topology placing all the "alters" on the periphery and "ego" in the centre of the circle. Meanwhile, this layout can represent all the relationships which are not only between 'ego' and 'alter' but also ones between "alters" as shown by figure below.

Figure 4-6: circular layout with a star topology for user-centric network (source: Internet)

- **Tree layout for object centric network**

Object-centric network is also a self-centred structure as user-centric network, which develops from online interactions based on a centred object. However, compared to the structure of user-centric network, object-centric network turns out to be a hierarchical view by concerning the formation of network and the mechanism of interaction. For example, in Twitter when discussing tweets around one hashtag, there are at least two kinds of tweets need to be considered: one is original tweet and another is the tweet interacting with the original one (retweet). The hierarchical structure in this case could be that the hashtag is the root (centred object) and the original tweets that link with root as the first level and the retweeted tweets that link with original ones as the second level and so on. Therefore, it cannot simply place non-centred nodes around the centred node otherwise the hierarchical structure will be lost. A common tree layout can be an option for this type of network since firstly this layout follows the hierarchical structure and second it arrange every node and link in a meaningful and clear way which make user easy to understand and detect the change.

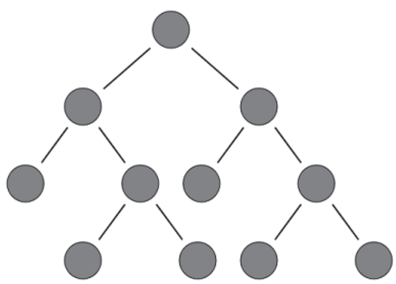


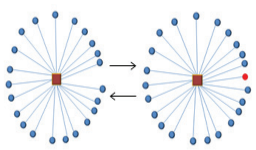
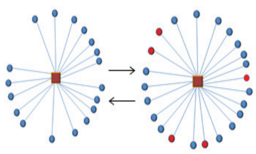
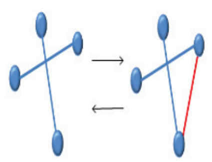
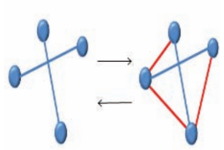
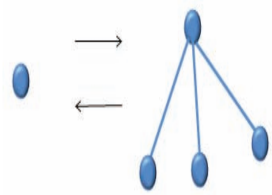
Figure 4-7: Tree layout for object-centric network (source: Internet)

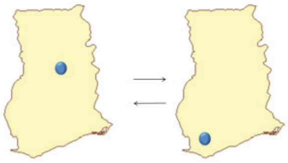
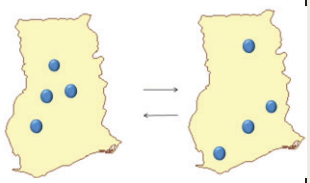
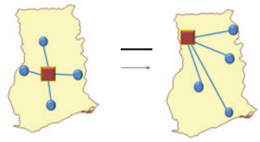
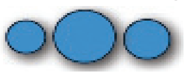
4.3.1.4. Map

The research uses a geographic correct map based visualization, on which displays the network under a geographical reference. The map based network visualization locates nodes by their geospatial information and connects the node-pair which has a certain relationship or interaction. It can facilitate user to see the network spatial distribution and understand the network characteristics and patterns by region. However, the problem of map based social network presentation could be that one location may have multiple overlapping nodes and if so the interconnected links cannot be seen in the meantime. It may be caused by map representing in a small scale or nodes which have the exact same geospatial information. The former problem can be solved by using function tool like zoom in to see nodes or links separately; due to the limitation of data and map representation, the latter one cannot be solved by map only, but since social diagram can arrange every node and link properly, linking map with social diagram could be the solution to this problem in this research.

4.3.1.5. Time representations for graph and map

Since this research also focuses on mapping the change of social networks, time representations for both social graph and map need to be considered. As noted above, different methods fit to different types of change, e.g. time for social graph is used for existential change etc. Table 4-6 summarizes the way how changes (existential, location and attribute) with each type of social network data elements are represented. There is no need to elaborate any literature about the changes with node(s) and link(s) in this table. For the sub-network, the location change for sub-network is more complex than other types of change. Specifically, location change for both types of sub-network includes movement and spatial distribution change (growth or contract); furthermore, spatial distribution change can be triggered by movement; in addition, spatial distribution change can also be caused by the existential change of node(s). The attribute change for sub-network in both social graph and map can be visualized by changing visual variables.

	Node	**Link**	**Sub-network**
Existential change: appearance or disappearance (represented by social node-link diagram)	Single node:  Multiple nodes: 	Single link:  Multiple links: 	
Location change(represented by map based visualization)		No relevant information	Movement without spatial range change

Location change(represented by map based visualization)	Single node:  Multiple nodes: 		● With centred node:  ● Without centred node:  Movement with spatial range change ● With centred node:  ● Without centred node:  Spatial range change with by the existential change of node(s) (only with centred node)  Spatial range change by both the existential change of node(s) and movement (only with centred node) 
Attribute change (represented by both)	Quantitative change 	Qualitative change 	Node-link graph:

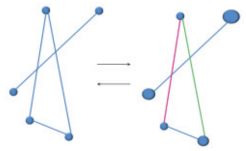
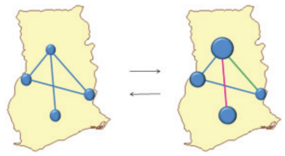
Attribute change (represented by both)			 Map: 

Table 4-6: Changing social network data element

According to the review of social network visualizations in chapter 3, single representation, series of snapshots and animation can be used to visualize change for social node-link diagram and map. However, single representation is not suitable for social network data because it inherently lacks ability in representing changes for many objects and cannot deal with the location change. The study is going to use series of snapshots and animation to represent changes summarized in Table 4-6. Series of snapshots are used in comparing different states of the network in several time moments; animation is used for grasping an overall trend or pattern of the changing social networks in a continuous time.

4.3.2. Working environment

As illustrated, user tasks of social network data in this research diversely involve questions from existential, locational and attribute areas. For a certain type of questions, there are certain visualization methods which can fit the questions better. The selections for suitable graphic representation have been made in last section, social node-link diagram is chosen to answer existence-related questions; map based visualizations are for location-related ones; visual variables and attribute tables are for attribute-related ones. However, in practice, a working environment is needed to make the above visualization methods be effectively meet user's end. The following literature will discuss which techniques and tools should be used in this case.

4.3.2.1. Coordinated multiple view (CMV)

For the consideration of integrating all the selected methods together to make user see and understand the data comprehensively and specifically, coordinated multiple view (CMV) technique is utilized in this prototype. This technique illustrated by Roberts(2008) that " an exploration environment where each of the views are linked together such that any user manipulation in one view is automatically coordinated to that of any other and environment". For this research, it is used as the working environment on one hand includes all the mentioned methods by multiple views, and on another allows user interact with data from different views by coordinating them with each other. Specifically, as Figure 5-1 shows, the multiple views are respectively the social node-link graph, map based visualization and attribute table or diagram. Moreover, in order to make this working environment can facilitate user to effectively interact with changing social network data, several function tools need to be introduced.

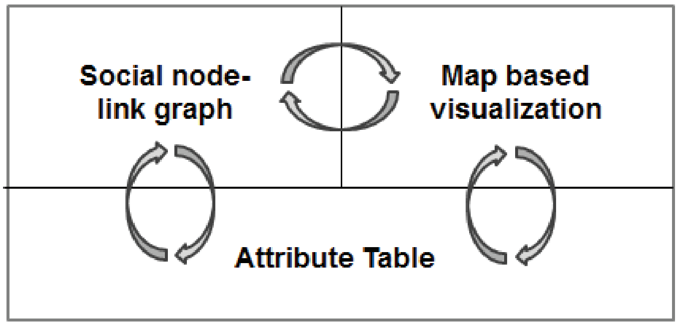


Figure 4-8: coordinated multiple view technique used in this research (source: Author)

4.3.2.2. Function tools

(1) Linking and Brushing

Linking and brushing are the tools forming an interactive selection process which aim at communicating the selected data in one view to the related data in other views(Dykes et al., 2005). Among the linked views, once selecting one or more objects in one view, the corresponding object(s) will be highlighted as well. It is the basis for coordinating multiple views and proved to provide user more information than from independent one. In this prototype, linking and brushing are used for coordinating all the three mentioned views.

(2) Query

Users are always interested in a subset of items and further their detailed information. Queries in this prototype can be conducted in two ways. One is firstly selecting the graphic object(s) in the view and then finding the result(s) in the corresponding information dialog window; another is firstly selecting interested information from the table and then seeing the corresponding objects highlighted in the map or social graph view with the support of linking and brushing functions.

(3) Time control

Since user tasks include static and dynamic ones, there is a need to make user interact with time. A time control tool here is proposed as Figure 5-2 illustrates to meet people's different temporal requirements by choosing different options. Each time choosing option can refer to a certain method. 'Static moment' can make user see the certain state of network data at one certain time by single static representation; 'multiple moments' can make user compare the different states of networks at different time moments by series of multiple static representations; 'a time period' can make user detect the overall change or trend of the networks over a period of time by animation.

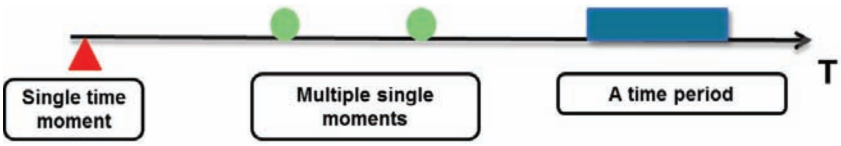


Figure 4-9: The time control tool with designed time choosing options (source: Author)

4.4. Summary

By following the visual problem solving approach, the chapter built up a conceptual model to represent spatio-temporal social network data. For the user tasks, the model designs a social network task space based on the “pyramid” data framework. Based on the user tasks and the nature of data, the graphic representations are chosen. For effectively using the selected graphic representations, a multiple view working environment with function tools like linking and brushing, time control etc. is proposed. The next chapter will discuss the prototype design based on this conceptual model.

5. PROTOTYPE DESIGN

5.1. Introduction

The conceptual model given by the previous chapter illustrated the visualization framework consisting of the suitable graphic representations and working environment for the social network data. Based on the conceptual model, the chapter describes the design of the prototype to visualize both types of social networks from social media data (user-centric social network and object-centric social network).Prototypes for the two case studies are designed respectively. One case study is Facebook user's friend network for the user-centric network, another one is Twitter trending topic network for the object-centric network. Finally based on the design, the related implementations and envisions will be discussed.

5.2. Prototype design for user-centric and object-centric social networks

As introduced in earlier chapters, user-centric and object-centric social networks, as two most popular types of social networks from social media data are going to be visualized in this research. The following content will illustrate the prototype design for both networks.

5.2.1. For user-centric social network

5.2.1.1. Case study data: Facebook friend network

Facebook, as the most popular social networking site, develops millions of social networks consist of users and relationships. For every individual user, he/she could make friends on the website so that the user and his/her friends and online relationships together form a user-centric network. The dataset in the case study is my friend network. For the network elements, they were derived by using Facebook application Name Gen Web which simply accesses and obtains the list of my friends and then outputs them in several formats (.dl, .gdf, .graphml). The output file saves node and link as figure 5.3 shows. Node element includes friend name (id), Facebook user id (uid), sex and profile picture web link; link element includes name of from node (source) and end node (target).


```
- <node id="TantaisChen">
      <data key="uid">100003033397136</data>
      <data key="sex">male</data>
      <data key="pic">https://fbcdn-profile-a.akamaihd.net/hprofile-ak-snc4/370719_100003033397136_756155642_s.jpg</data>
  </node>
- <node id="MimiCao">
      <data key="uid">100003192765363</data>                Node
      <data key="sex">female</data>
      <data key="pic">https://fbcdn-profile-a.akamaihd.net/hprofile-ak-snc4/370472_100003192765363_1334295992_s.jpg</data>
  </node>
- <node id="ChaoZhen">
      <data key="uid">100003329765391</data>
      <data key="sex">male</data>
      <data key="pic">https://fbcdn-profile-a.akamaihd.net/hprofile-ak-snc4/369331_100003329765391_979714977_s.jpg</data>
  </node>
- <node id="DanLi">
      <data key="uid">100003355443856</data>
      <data key="sex">female</data>
      <data key="pic">https://fbcdn-profile-a.akamaihd.net/hprofile-ak-snc4/370754_100003355443856_717369793_s.jpg</data>
  </node>
  <edge target="StevenMa" source="DingMa"> </edge>       Link
  <edge target="MimiCao" source="DingMa"> </edge>
  <edge target="DavidZhao" source="DingMa"> </edge>
  <edge target="JadeHan" source="DingMa"> </edge>
```


Figure 5-1: Data of Facebook friend network elements

The location data of both user and friends in this case study consists of two kinds of spatial information: hometown and current location (at a certain time t). As chapter 2 introduced, geo-component in Facebook is embedded as place name (up to city level) for hometown, current residential in the user profile (upper image in figure 5.4), or place name (up to street level) in status, photos, videos or check-in point (middle image in figure 5.4). Meanwhile, Facebook also has a map application organizing and summarizing all the location-involved contents for every user with an interactive timeline for every user to show where they have been and corresponding time information is also included (lower image in figure 5.4). All these geo-related information are the source for the acquisition of the location data. As for temporal data, the case study chose three time moments which are August, 2010, October 2010 and April, 2011 to represent the change of network.

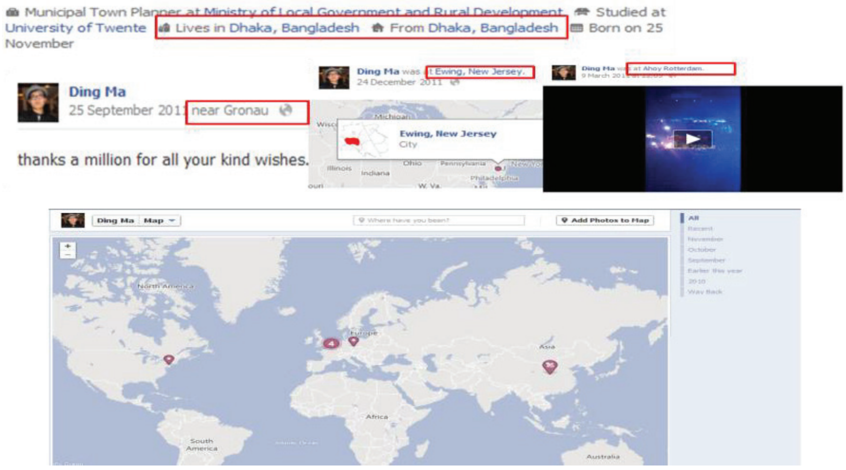


Figure 5-2: Location data in Facebook

5.2.1.2. Prototype design

Based on the case study data, the prototype as Figure 5-6 shows mainly consists of three components: map, social graph and attribute table. Specifically, for the map view, there are two kinds of map: hometown map and current location map. Hometown map locates user and friends according to their hometown information; current location map locates both user and friends based on their real locations at one certain time ('current' does not mean now). Social graph uses circular layout plotting user in the centre and every friend on the periphery to show the structure of the user-centric network. Visual variables are used in social graph (shape and colour hue) and two maps (shape and size) to represent different characteristics of the networks. Attribute table records the information of each node. As for the consideration of representing change, since in this case study there are only three moments of the network, using series of multiple static representations for both map and social graph could be a better option than animation. Therefore, options are only static moment and multiple moments for the time interactive tool.

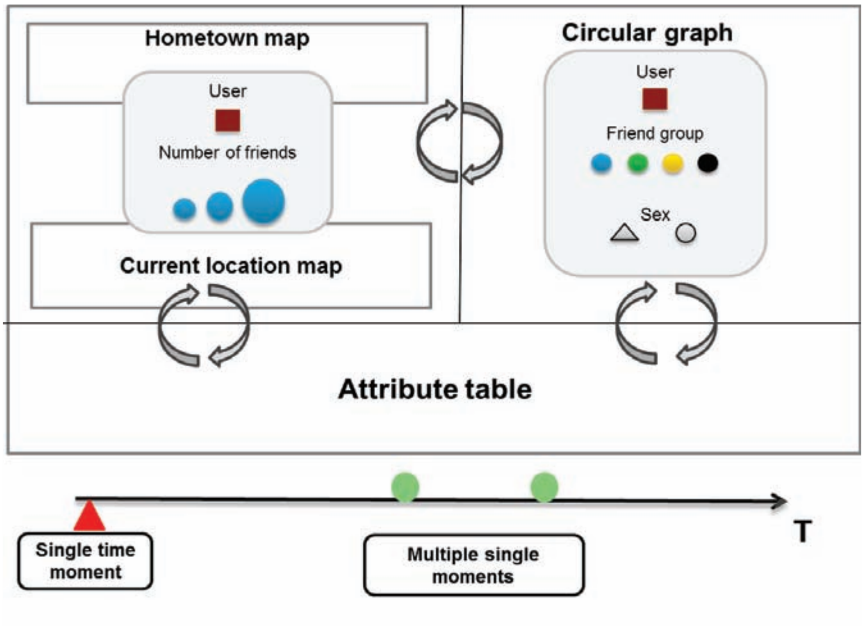


Figure 5-3: The designed prototype for Facebook friend network

5.2.2. For object-centric social network

5.2.2.1. Case study data: Twitter trending topic network

Twitter, as the most popular social networking and microblogging service, makes millions of people all over the world contribute and share their views, feelings etc. by means of tweets—short messages limited to 140 characters. Among billions of tweets posted in Twitter, there is a certain amount of tweets around one or more themes featured by hashtag(s), which is a descriptive keyword with a sign "#" in front (Hansen, et al., 2009). Hashtags could be an event, a person or even a holiday name. For any hashtag attracting a large number of tweets to talk about, it becomes a trending topic. Tweets around one trending topic can be understood as the object-centric network (the topic hashtag as the centre object).

The dataset used in the case study is about tweets around "#japanearthquake" hashtag, which was one of the top ten topics in the year of 2011 in Twitter. The tweets were collected from TwapperKeeper (http://twapperkeeper.com/index.html) which is a social media data service website helping social media user archiving their tweets by different hashtags. As for tweets, as Figure 5-7 shows, each tweet contains the hashtag "japanearthquake" and moreover, there are two kinds of tweet in the dataset: original tweet and retweet. Original tweet is way that people share their own ideas about the topic; retweet is the way that people share other's ideas. The retweet includes one user's original tweet and starts with "RT" and along with that user's name "@user name". The network elements therefore can be denoted that the centred node is the hashtag "#japanearthquake"; other node elements are composed of original tweets and retweets; link elements in this way contains two types of interactions: "original tweeting" interaction between original tweets and hashtag, "retweeting" interaction between retweets and original tweets. It needs to mention that in Twitter the retweet can also be "retweeted", however it is out of the scope of this research. The study only considers one level retweet situation, in other words, if one tweet includes more than one "RT", only the first "RT" and the related original tweet will be considered.

Figure 5-4: The example tweets collected in this case study

The location data in the case study come from two areas of the tweets as Figure 5-8 shows. For the geo-located tweets, GPS coordinates are used for the location data; for tweets without GPS coordinates, I assume the location information in the user's profile is where the tweet comes from. As for temporal information, each tweet contains a time stamp with the precision to second recording when the tweet had been posted onto the web. The time period of the dataset spans the early morning of the day (7:00 am to 9:00 am on March 11th) when the earthquake happened.

a) b)

Figure 5-5: Location data in tweets: a) Geo-located tweets; b) location information in user profile

5.2.2.2. Prototype design

Based on the case study data, the prototype contains social graph, map, attribute table and diagram (see Figure 5-9). Social graph uses tree layout displaying the hierarchical structure of the network where the hashtag is placed on the top and the original tweets and retweets are placed downwards at the first and second level; tweet map is map based visualization for the network locates hashtag on the place where the event happened and every tweet by their GPS coordinates or user's profile information to show the spatial distribution of the information around this trending topic; visual variables (shape, colour, size) are used in both representations for different qualitative and quantitative attributes of nodes and links; attribute table records the relevant tweet information; diagram is used here for depicting the number of original tweet and retweet through the time period; Time interactive options for this environment are static moment for single time and animation for change, since representing the change in the case study is for understanding the process of information diffusion, animation is more suitable for user to see the overall change or trend of the process.

Figure 5-9: The prototype design for object-centric network by the case study

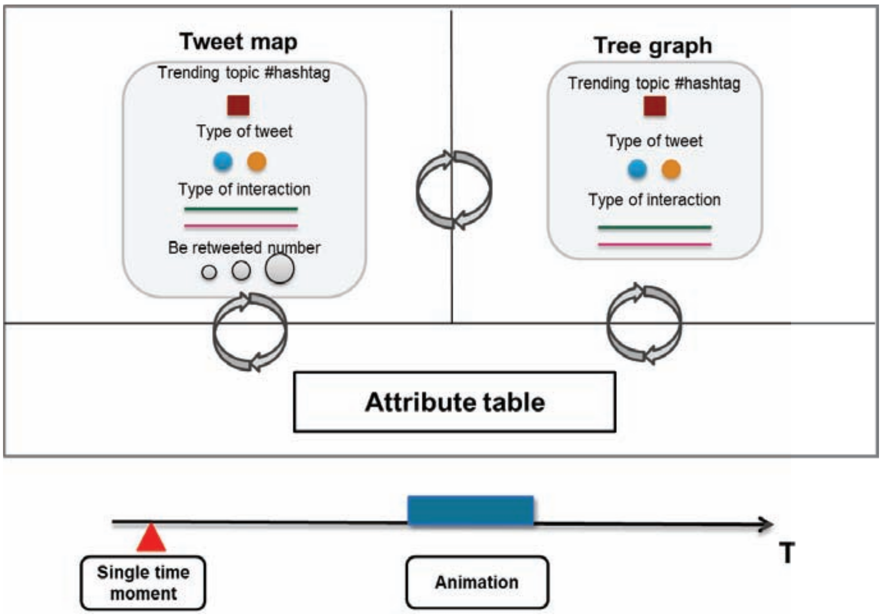


Figure 5-6: The designed prototype for Twitter trending topic network

5.3. Towards implementation of the prototype

5.3.1. Software selection

Table 5-1 shows the softwares selected for different usages to make an effective design for the development of prototype.

Selected software	Usage
ArcMap/ArcCatalog	Making hometown map, current location map for Facebook friend network, tweet map for Twitter trending topic network
NodeXL	Making circular social graph for Facebook friend network and tree graph for Twitter trending topic network
Adobe Macromedia Flash	Making animations for tweet map and tree graph
uDig	Providing the working environment with multiple views for map, social graph, table and diagram and involving function tools like brushing, linking, querying etc.

Table 5-1: Selected softwares and their usages at the prototype design stage

5.3.2. Implemented maps, graphs and time representations for each prototype

5.3.2.1. Facebook friend network

Circular graph

As introduced, circular layout is used for user-centric network. Here the Figure 5-10 shows the circular graph which depicts my friend composition on April, 2011. Colour and shape used here for differentiating the friend groups, sex, user and friend.

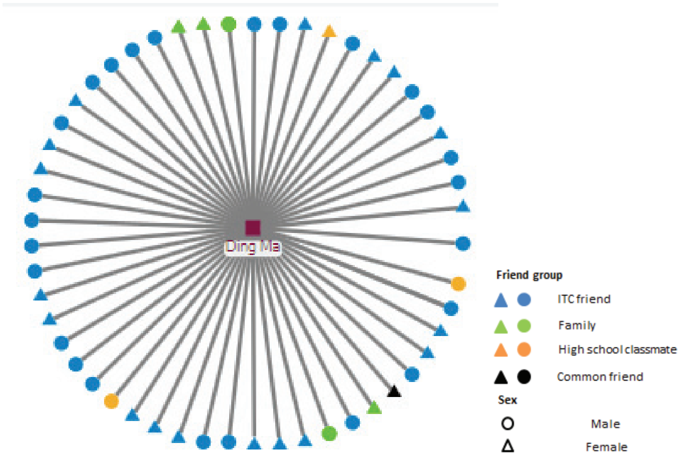


Figure 5-7: Circular layout for my Facebook friend composition

Hometown map

Hometown map is used to show where my friends are from. Figure 5-11 shows the hometown map depicting the spatial distribution of the nationalities of my Facebook friends on April, 2011. The map shows that the friends mainly come from south and middle part of China, and also there are friends from Europe, Africa and South America. Moreover, by differentiating the size of point symbols and referring to the legend, people can also see how many friends come from same place.

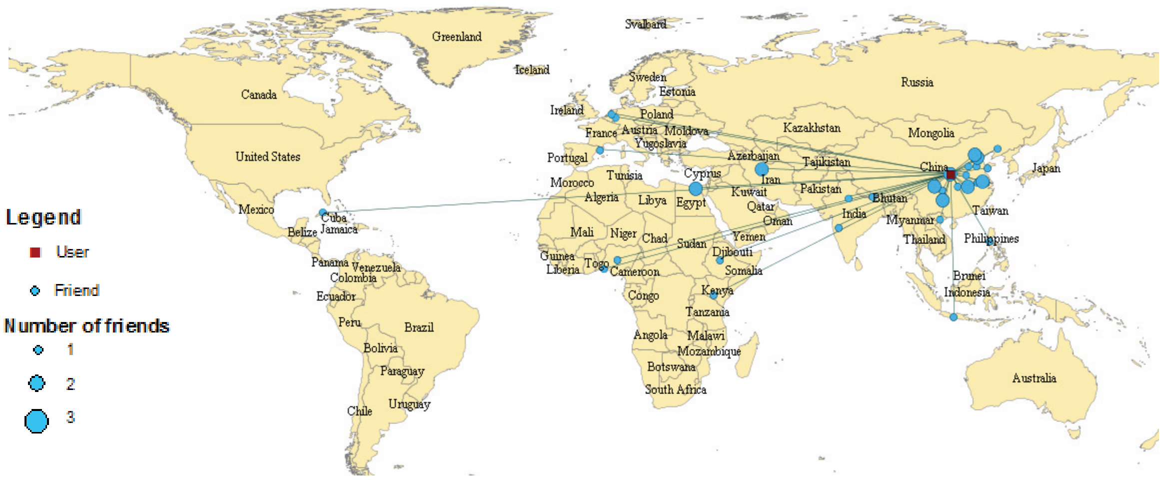


Figure 5-8: Hometown map

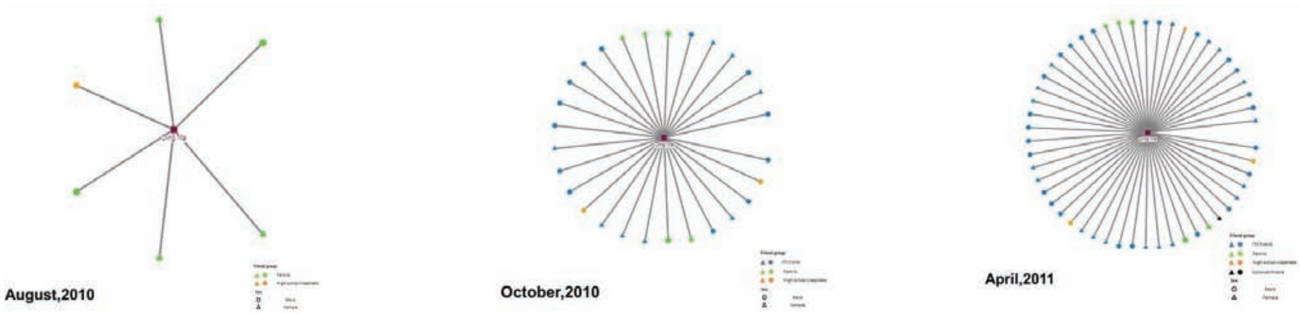
Current location map

Current location map shows the where the friends were at that moment. Same as hometown map, current location map also uses the proportional point symbols to describe the case that the friends come from same place. From Figure 5-12, it can be seen that more than ten of my friends were in The Netherlands at this moment.

Figure 5-9: Current location map

Series of multiple static maps and graphs for representing change

The case study uses multiple statics to represent the change of the network. The selected time moments are August 2010, October 2010 and April 2011 which are the special moments of my network. Figure 5-13a shows the series of social graph describing the friend composition change. It can be seen that the total number of friends increased (new nodes) greatly after August 2010 and also three new groups appeared (new colour); Figure 5-13b shows the series of hometown maps displaying the enlargement of the range of my friends' nationalities (spatial range change) and also can see whether the friends from the same place become more (point size change); Figure 5-13c shows the series of current location maps depicting user' movement and changing friends' spatial distribution.



a)

b)

c)

Figure 5-10: Series of multiple graphs and maps for representing change
a) series of multiple social graphs; b) series of multiple hometown maps; c) series of multiple current location maps.

5.3.2.2. Twitter trending topic network

Tree graph

Tree layout is used in this case study to follow the hierarchical structure of the network. As Figure 5-14 shows, the hashtag "#Japanearthquake" is on the top, the original tweets are on the first level and the retweets are on the second one. The size of the original tweet indicates the number of this tweet is retweeted.

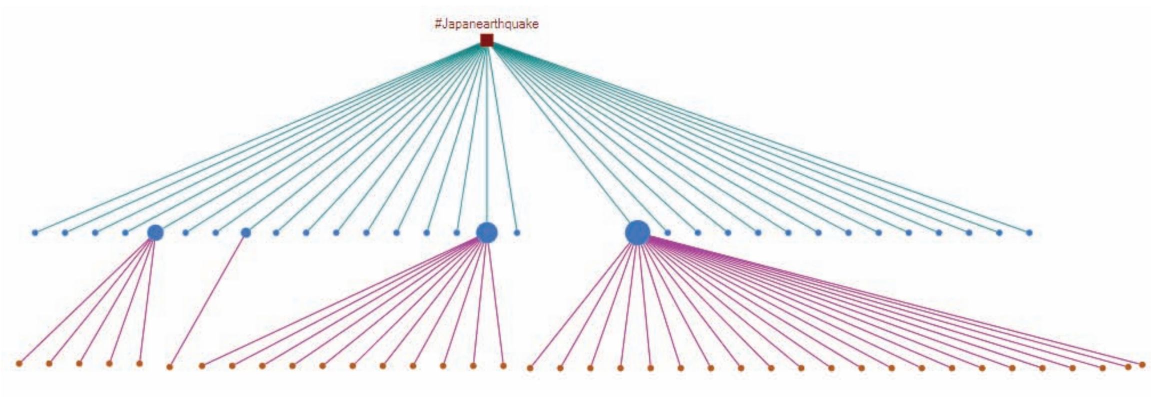


Figure 5-11: Tree graph for twitter trending topic network

Tweet map

Tweet map is used to show the information diffusion through the space. As Figure 5-15 shows, the message of trending topic "#Japanearthquake" at this time had spread all over the world. And also by differentiating the size of the original tweets (blue point), the important locations for spreading the information can be found.

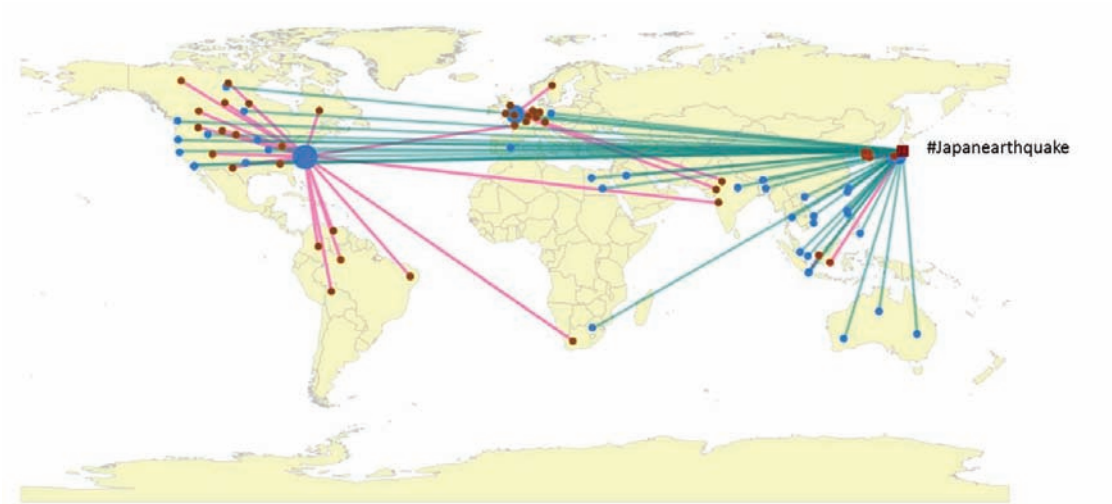


Figure 5-12: Tweet map

Animation of map and graph for representing change

In this case study, animation is used both in map and graph to see the overall change or pattern of the diffusion of this trending topic (Figure 5-16). As for the tweet map animation, the spatial pattern of the message spreading during the period can be seen; as for the animation of the tree graph, the tweet pattern which means how different types of tweet were interacting can be detected.

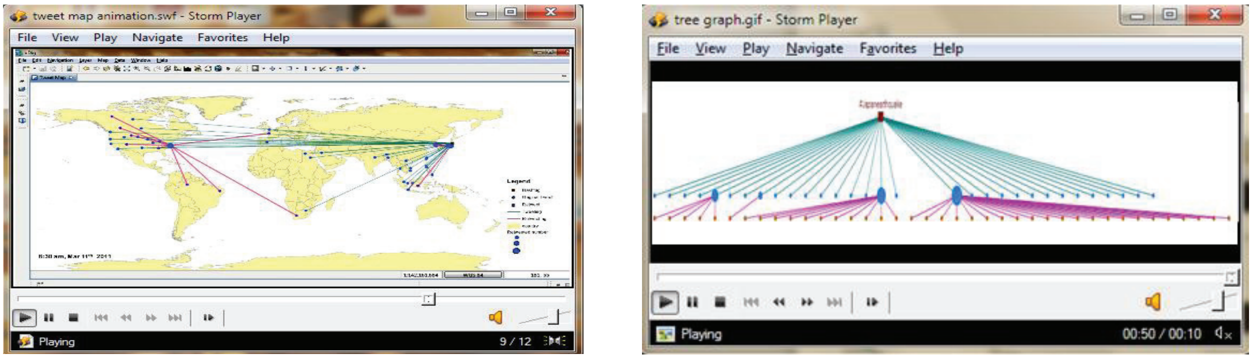


Figure 5-13: Animation of both map and graph in this case study

5.3.3. Working environment and functionalities

5.3.3.1. Overview

By using the implemented graphic representations can help user find the spatio-temporal characteristics of social networks to an extent. However, only using map or social graph is not sufficient when executing complex tasks. Therefore, how to utilize these representations effectively needs to be considered. As noted in chapter 4.3.3, a multiple view environment is proposed which contains social graph, map based visualization, attribute table and/or diagram, and a time control panel. With the functions like linking and brushing, these views are coordinated to support and be interactive with people to understand and explore social networks under a spatio-temporal context. As a result, the environment is designed and developed in uDig GIS environment, which is an open source desktop application framework and under the Eclipse platform to offer data manipulating solutions by using plug-ins and function tools. It can be seen from a full view of the working environment (see Figure 5-17) that several windows in the prototype represent social network data from different aspects, and the tool bar and time control tool provide the interactive functions. Due to the time limitation, this working environment is still in the design stage since the linking and brushing function between social node-link diagram and other two views, and the time interactive tool are not realized. The following content will illustrate these two envisioned functionalities of the working environment.

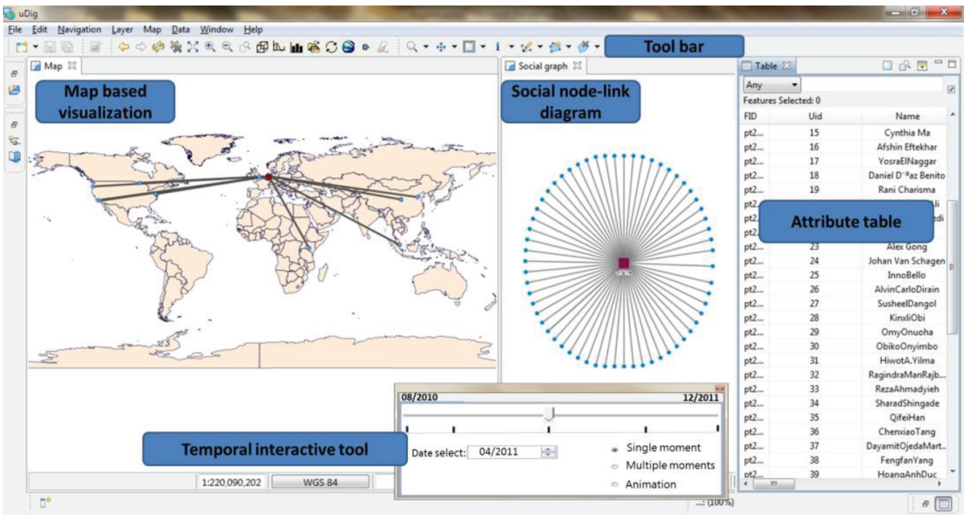


Figure 5-14: Overview of the working environment

5.3.3.2. Functionalities

Basic uDig functions

uDig can provide multiple windows for different visualization methods. In practice, user can choose which one is open or closed since not every question need to be answered by the combination of different windows. The tool bar offers some general GIS functions like: feature selection, zoom in/out, overview, information query, and data editor etc. Furthermore, the next two sections will introduce two additional functions designed to make the changing social network data become more operational and therefore facilitate user to explore the complex information.

Linking and brushing

Linking and brushing is designed to make views of the social graph, map and table act as one and dynamically filter the results thereby allowing user perform complex tasks. Using the function between map and social graph can greatly enhance the ability of each other. One case is which as discussed in chapter 4.3.2, when a-bigger-size point on the map referring to the overlapping nodes at the exact same location, reflecting them in the social graph can find each corresponding node clearly, and vice versa, user can also find the spatial information for the interested node in the social graph. Moreover with respect to the detailed information, we can link those points with attribute table. For example, user may want to know which group(s) are my friends in who were living in America on April, 2011. In this case, a linked view of social graph and current location map is provided. By clicking the points on the map, the corresponding points in the social graph will be chosen and then judging the group by colour, meanwhile the highlighted rows in the table offer the detailed information of these friends (see Figure 5-18).

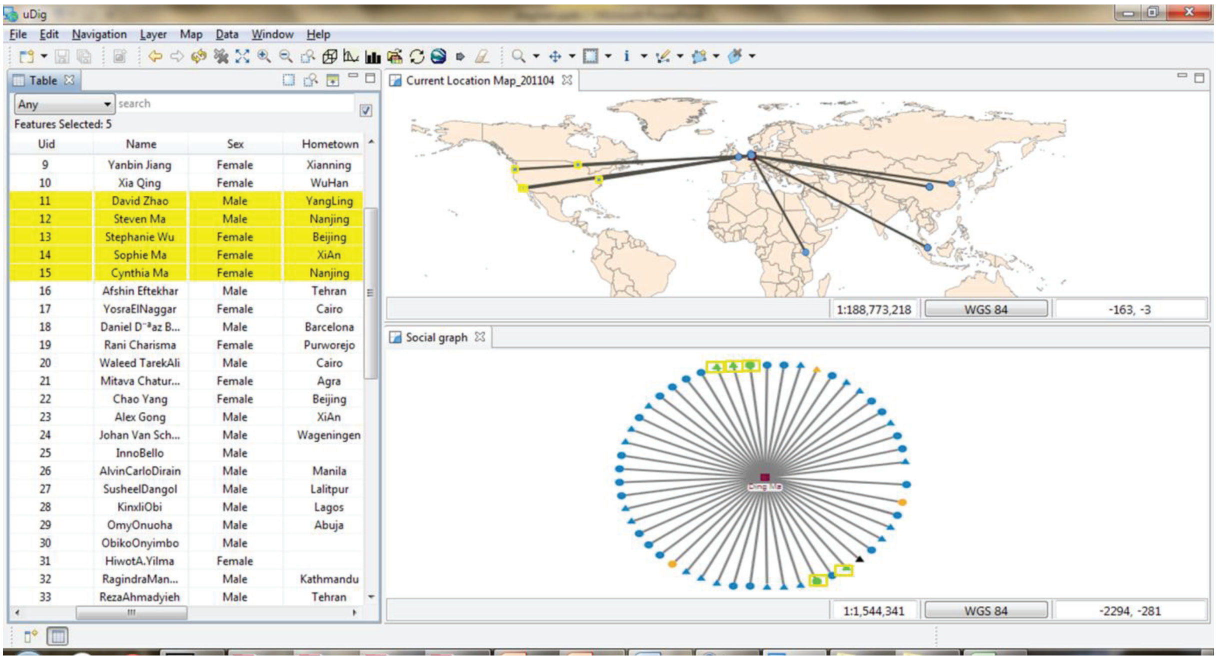
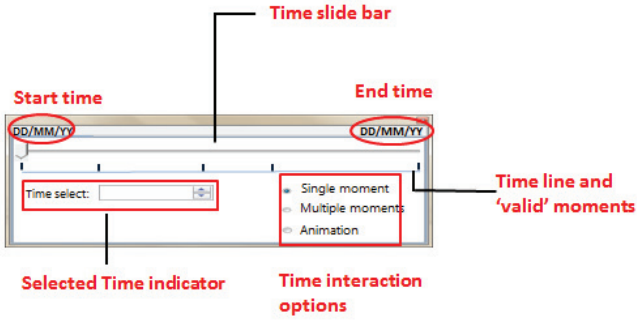


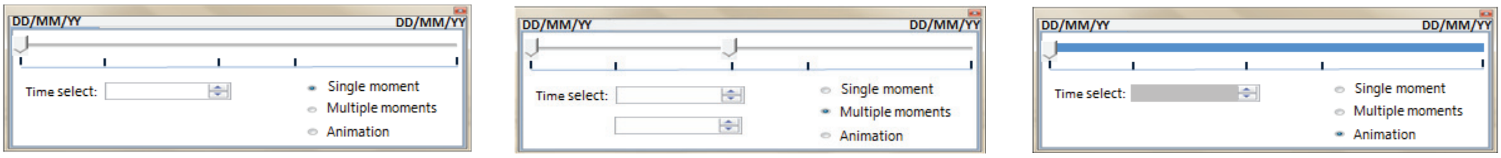
Figure 5-15: Linking and brushing for helping execute complex tasks

Time control panel

In order to fulfil the temporal tasks introduced in last chapter, a temporal interactive tool is designed. As Figure 5-19a indicates, this temporal interactive tool provides a time slide bar and clearly presents the 'valid' time moments along with the timeline, the selected time(s) or time period will be shown in the combo box, and further it allows user query data in one single time moment and a set of times by offering three options: single moment, multiple moments and animation. Figure 5.19b shows how three types of temporal interactive options by the tool looks like.



(a)



(b)

Figure 5-16: Time control panel

a) interface; b) examples of three time interaction options

By involving the time control panel, the working environment becomes more user-friendly when performing temporal tasks. User can choose one interesting moment, or two moments for comparing, or the whole time period to see the overall change or trend. Figure 5-20a shows the envisioned use time control panel in the Facebook friend network change; Figure 5-20b shows the one in Twitter trending topic network.

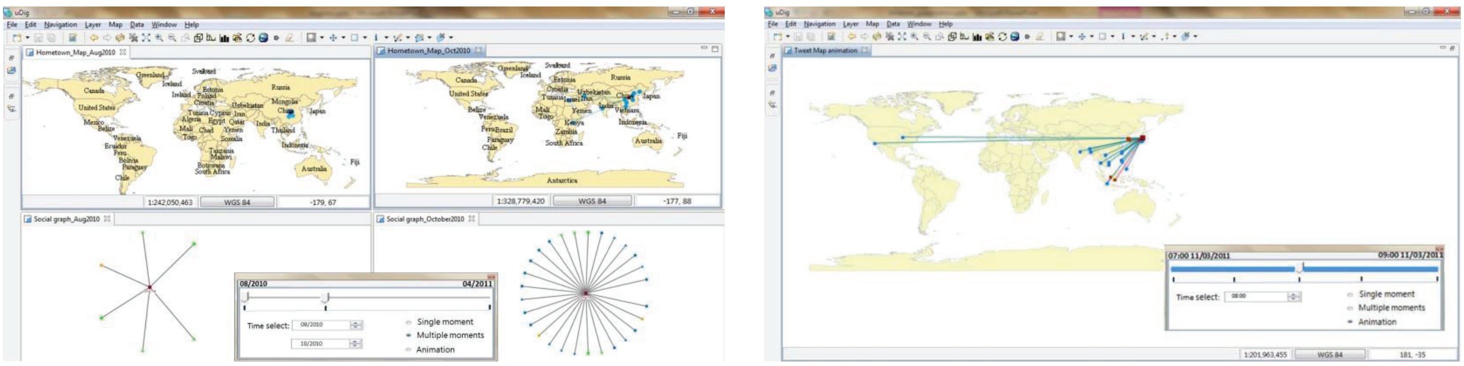


Figure 5-17: Envisioned use of time control panel.

a) In Facebook friend network; b) in Twitter trending topic network

5.4. Summary

Following the conceptual model, this chapter designs the prototype by using two case studies: Facebook friend network and Twitter trending topic network. For the development of the prototype, the map based visualization, social graph and time representation for each use case are all implemented and furthermore the working environment and its functionalities are designed. The usability evaluation of the prototype will be discussed in the next chapter.

6. EVALUATION

6.1. Introduction

Evaluating the designed prototype is one of the objectives of this research. In this chapter, the focus group as one of the usability methods will be used to evaluate the prototypes designed for the two case studies. Section 6.2 describes the method; section 6.3 illustrates how the usability evaluation was set up; section 6.4 discusses the obtained results and overall conclusion of the usability test.

6.2. The focus group method

A focus group can be understood as "a group interview" which is guided by a moderator discussing the predefined topics towards a product, service, concept and idea etc. A focus group session normally takes one to two hours collecting the opinions or ideas from a small group (5-8) of participants who come from similar background and would better be the domain-specific users or experts (Dykes, et al., 2005; Morgan, 1998). It is therefore a relatively simple, efficient and less time-consuming qualitative evaluation method (Blok, et al., 2005; Dykes, et al., 2005). In the geovisualization field, this method has been applied several times in the prototype design. For example, Monmonier and Gluck (1994) applied focus group to obtain user's reactions and ideas for the interface design; Blok et al(2005) used focus group to collect users' opinions to make adaptations to the "aNimVis" prototype etc..

In this research, the focus group method is used to evaluate the effectiveness of the designed prototypes for the two case studies: Facebook friend network and Twitter trending topic network. As the last chapter illustrated, there are two components of designed prototypes that need to be evaluated: the implemented maps, social graphs and time representations, and the designed working environment and its functionalities. The following content will demonstrate how the focus group session was designed and processed and finally summarize the received results from the test persons.

6.3. Usability evaluation

Figure 6-1 shows the workflow of the usability evaluation.

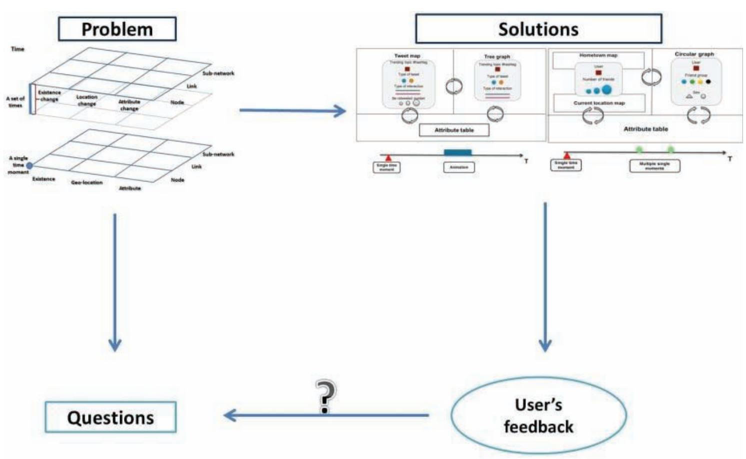


Figure 6-1: The overview of the set-up of evaluation

6.3.1. Objectives and tasks

6.3.1.1. Objectives

The main goal of the focus group session is to test whether the designed prototypes can be used to help user to understand the spatio-temporal characteristics of the social networks from social media data. Specific to the components of the prototype, the session focused on the usefulness of the representation of social network data and the designed working environment and its functionalities.

6.3.1.2. Tasks design

User task space is used here for designing the tasks. Since the interest of the research is visualizing changing social network, the designed tasks are all dynamic tasks. Table 6-1 indicates what each type of the social network data elements refers in the Facebook friend network and Twitter trending topic network. Because in both type of networks the existential change of link element is determined by the corresponding node, and the attribute of link element does not change either, there is no task for link element.

	Node	**Link**	**Sub-network**
Facebook friend network	User or friend	Online relationship	User and friend(s)
Twitter trending topic network	Hashtag, original tweet, retweet	Tweeting, retweeting	Hashtag and original tweet(s), original tweet and its retweet(s)

Table 6-1: each social network data element referred in both types of network

For each designed prototype, there are two questions, and test person has different roles to play. Below are the lists of the designed tasks for each prototype.

- Facebook friend network:

Suppose each test person is the ordinary Facebook user, he/she wants to know about changes over time in my Facebook friend network. The questions are:

(1) How did my friend composition change when I moved from China to The Netherlands?
(2) Were there any spatial changes (nationality, movement) of my ITC friends between October 2010 and April 2011?

- Twitter trending topic network:

Suppose each test person is a social scientist, he/she wants to detect how the message of Japan earthquake spread over the world in the early morning on 11 March, 2011. The questions are:

(1) During this period, which place(s) became important for spreading the information?
(2) What was the overall pattern of the diffusion of information? Were the patterns different between the first hour (7am to 8am) and the second one (8am to 9am)?

These questions are all related to the spatial-temporal dynamics of the social networks. Figure 6-2 shows the distribution of these questions on the task space.

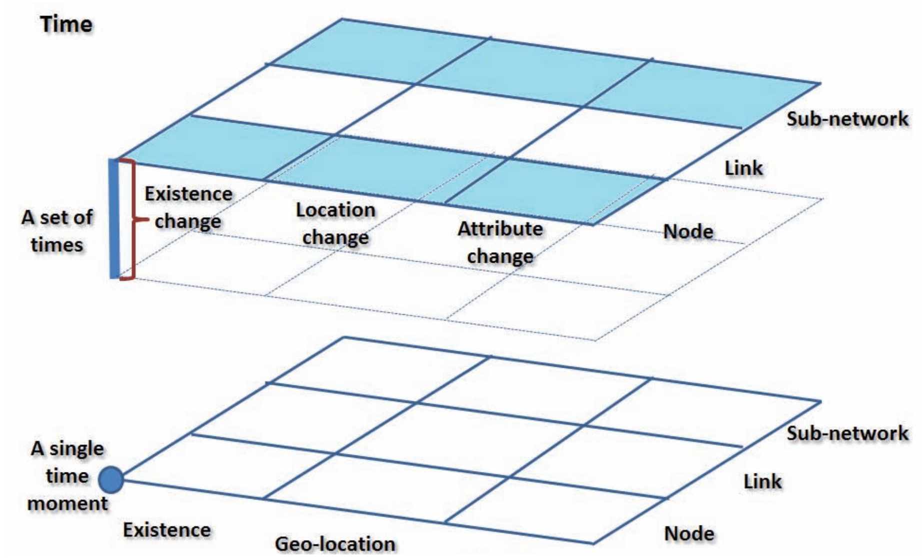


Figure 6-2: Tasks distributed in the task space.

6.3.2. Test environment

The session was held in a group decision room at ITC (Enschede, the Netherlands) on 8th February, 2012. A touch table was used to show the implemented graphic representations and designed working environment during the test period. Furthermore, a beamer and a whiteboard were prepared for the background introduction. The whole process was recorded by video for the further analysis.

6.3.3. Procedure and participants

The session started at 2:00 pm and ended at 3:30pm. During this one and a half hour, there were four steps for the session:

(1) Welcome the participants and present a 10 minutes brief introduction of the background of research and the objective of the evaluation.
(2) Introduce the two prototypes and introduce test to test with a 5 minutes demonstration, in the meantime user can familiar with the operations on the touch table.
(3) Group test of the two designed prototypes by answering the questions from the moderator for each case study, the duration of this step was about 1 hour and 5 minutes.
(4) 5-minute group discussion for giving suggestions and feedbacks.

There were seven participants totally: one MSc student, five PhD students and one teacher. All participants are from the Department of Geographic Information Processing. As mentioned earlier in this chapter, the focus group participants would better be the domain specific users or experts. Since every participant has the GIS and cartography background, the suggestions and comments from them are to a great extent useful and helpful for evaluating the prototype.

6.4. Results

The results were obtained from the discussion in step 3 and 4 of the session. During the one hour and ten minutes group discussion, all seven participants contributed their ideas, feelings, suggestions about the prototype. The results including suggestions and feedbacks from each participant are summarized in Table 6-2, which categorized the results into "wishes and suggestions" referring to the improvements of the designed prototypes and "usability aspects" referring to the overall feedbacks of the usability of the designed prototypes.

For testing the prototype of Facebook friend network, I prepared the circular graphs, hometown maps, and current location maps at three different moments (August 2010, October 2010 and April 2011) and also pictures of the designed working environment describing how those representations can be linked. When answering the first question, users firstly were instructed to open the map view to see my movement, and then they wanted to look the circular graph to compare the friend composition. At this time, one participant mentioned that "*it would be helpful if the map and social graph are linked*", it proved that the designed interactive environment is necessary when user want to use map and social graph together. When comparing the circular graph, they indicated that besides knowing every individual, the aggregated information was also required. For example, one participant said "*How can I know how many male ITC friends you have at this time?*" In this case, they gave suggestions of adding more information (e.g. classification) in the legend. Also, the suggestions of the visual variables in the social graph were given as listed in the table 6-2. When answering the second question, participants firstly investigated the hometown map to detect the nationality changes, participants all can find the range of nationalities was expanded. And for friends' movement, with the help of designed environment linking social graph and current location map together, participants saw several of my ITC friends moved back to their countries. In the meantime, they also discussed how the map can better visualize the social network data. One participant suggested that "using a pie chart can show different sext or groups of people in the same location". Another participant mentioned the limitations of map showing the social networks "*if the people are close, the map would become messy at that area*". To this concern, participants suggested to apply the visualization strategy like Shneiderman Seeking Mantra to outcome the limitations, one participant said "*we can see the messy points on the overview level, and then zoom in and filter, finally get the detail information on the social graph or attribute table.*"

For testing the prototype of Twitter trending topic network, I used two animations: one is tweet map animation and another is the tree graph animation. When answering the first question, the participants investigated the tweet map animation to find the important places. As results, animation answered this question clearly. Meanwhile many comments were made to the design of animation. One participant noted that the changing visual variable used in point symbol should be more obvious because the appearance of the lines can easily attracted people's eyes. And furthermore, one participant suggested that "*the countries which do not contain information should be assigned into white colour to make people more concentrated on the places where information has been spread*". And they also discussed that for the different types of lines, the applied animation techniques could be vary because tweeting and retweeting interaction are different characters in the information diffusion. When answering the second question, they first saw each animation separately and then watch them together. The overall patterns of both spatial and tweet were discovered. In this case, the group discussion focused on the interactive environment of the animation. They said "*it is better to combine two animations together to provide user a comprehensive view of how information goes on*". In the meantime, one participant mentioned that if two animations playing together would cause distractions. Then the group discussed that the designed environment can be improved by adding linking and brushing between two animations.

Wishes and suggestions
➢ Prototype for Facebook friend network *Implemented graphic representations* - More descriptions should be added to know the differences of hometown map and current location map - When more friends appear in the social graph, it is not easy to see how many female or male friends. Legend can include this information, such as adding more specific classifications.

- The size of node in social graph should be bigger for user to compare the shape. - Nodes in social graph can be aggregated when friends become more. - The size of nodes in the hometown map should more distinguishable. And for the bigger-size points which represent many people at same place, the point symbol could use a multivariate symbol to show more information. - The white colour can be used on the country polygon to indicate that the user does not have friend located in this country. *Designed working environment* - When clicking the user on the either current location map or hometown map, the corresponding circular graph shows up in a small window within the map but not besides the map. - A visualization strategy like Shneiderman seeking mantra can be applied. ➢ Prototype for Twitter trending topic network *Implemented graphic representations* - The design of colour and size of nodes should be more distinguishable. - The change of node should be made more obvious on the map. - Different animation techniques should be applied to the different types of lines. - Lines for retweeting interaction should be directed. - Colour can be used in polygon to differentiate whether the country having the information, and make people more concentrated on those places. *Designed working environment* - Animation option in the time control panel should let user import more than one animation. - The interactive environment between animations (linking, brushing) can also be designed.
Usability aspects
- Integrating map and social graph together can compensate each other's limitation on representing social network data. - The designed interactive environment is necessary especially when using multiple static snapshots. - The designed prototype can help people understand the spatio-temporal characteristics of social network data.

Table 6-2: The summarized results from the focus group session.

Overall, the participants agreed that using the combination of map and social graph can provide a comprehensive understanding of the social network data. For exploring the spatio-temporal dynamics of the social networks, they agreed that it is suitable to use multiple static snapshots and animation to represent the change of social network, and also the designed interactive multiple view environment is needed.

6.5. Summary

Based on the focus group method, results of usability of the designed prototypes for two case studies were gathered and discussed in this chapter. By doing the test of each prototype, suggestions and comments were received for further improvement of both prototypes. As a result, participants who are the domain experts concluded that the designed prototypes are useful to help people understand the spatio-temporal characteristics of the changing social networks from social media data.

7. CONCLUSIONS

7.1. Conclusions

The main objective of the research is to design a visualization environment to represent the spatio-temporal characteristics and also keep the social structures of the social networks from social media data. To meet the objective, eight questions in total are proposed in chapter 1 need to be answered by doing this research. Consequently, the follow content will conclude how these questions are answered by looking through the chapters before.

1. *What are social networks and how do social networks evolve with the advent of social media?*

Chapter 2 reviewed the basic concepts and the development of social networks. Social networks originally can be defined as "a set of people who share a common interest and have connections of some kind", which only include people and their relationships. With the advent of social media, the range of the social networks is enlarged. Besides representing the relationships among people, the social networks built from social media data can also refer to the interactions between objects, or people and objects. The object is the digital medium like photograph (Flikr), video (Youtube) and Hashtag (Twitter). Of those countless social networks, the research chose two most popular types of social networks to study: user-centric social network which develops from online relationships around a user (e.g. one's friends in Facebook or followers in Twitter etc.), and object-centric social network which develops from online interactions around a social object (e.g. a photograph in Flikr).

2. *Which existing visualization methods are suitable to depict social network data?*

Chapter 3 reviewed the existing visualization methods for the social network data. The social network data are mainly in forms of node-link or matrix. The research only focused on the node-link form. In this case, social node-link graph is used most for visualizing social networks. There are many network layout algorithms to visualize social networks. Basically, the network layout can be categorized into two forms: graph layout and tree layout. Examples in graph layout are circular layout, random layout and force-directed layout; tree layout originally is the follows a top-down hierarchical mechanism, its variations can be balloon layout, radial layout etc. In addition, map can also depict social network data by placing every node according to its location information and connect the node-pair by lines or arcs.

3. *How to extend social networks with space and time?*

Chapter 3 firstly introduced the social network data into the Peuquet triad framework which described social network data from an attribute, temporal and location perspective. In chapter 4, the research used the extended triad framework—"pyramid" framework which places social network data as the "object" on a higher level. Furthermore, the social network data in this model contains three elements: node, link and sub-network. For node and sub-network elements, attribute, location and time are all involved; for link element, only attribute and time are involved. In chapter 3, in order to summarize the existing visualization methods for the social networks in a logical way, the location component of the data framework referred to graph location and geographic location. In chapter 4, since the research is focusing on the geo-related phenomena, the location only referred to geographic location.

4. *Which graphic representations can be used for representing the spatio-temporal social network data?*

In chapter 4, a conceptual model was designed for representing the spatio-temporal social network data. Within the model, a user task space was proposed based on the "pyramid" social network data framework. From the user task space, tasks for spatio-temporal social network data consist of "whether" (existential) questions, "where" (location) questions and "what" (attribute) questions in both static and dynamic situations. The suitable graphic representations were selected based on the designed tasks. For existential-related questions, social graph is selected; for location-related questions, map based visualization is selected; for attribute-related questions, visual variables of point and line symbols are selected. As for representing change (dynamic situation), series of multiple static representations and animation are selected.

5. *How can we represent all characteristics of social networks in a 'map'?*

Map is designed in this research to represent location information of the social network data. As for other characteristics, the visual variables of point and line symbols could be very helpful to represent different nodes and links with different attributes. However, map has its limitation for that one location may have multiple overlapping nodes and if so the interconnected links cannot be seen in the meantime. It may be caused by map representing in a small scale or nodes which have the exact same geospatial information. The former problem can be solved by using function tool like zoom in to see nodes or links separately; due to the limitation of data and map representation, the latter one cannot be solved by map only, but since social diagram can arrange every node and link properly, linking map with social diagram could be the solution to this problem in this research.

6. *What are the required functionalities of visual interactive environment for the spatio-temporal social network data?*

In chapter 4, the working environment and its functionalities were designed. The research proposed a multiple view environment integrating social graph, map based visualization and attribute table together to represent the spatio-temporal characteristics of the social network data. In order to make the environment become interactive, the function tools like linking and brushing, querying and a time control panel are designed.

7. *How to design the prototype based on the use case(s)?*

Chapter 5 illustrated the designed prototypes for two case studies: Facebook friend network and Twitter trending topic network. For each prototype, the related graphic representations and the working environment are proposed according to the collected data. As for the graphic representations, the prototype of Facebook friend network used two map based visualizations: hometown map and current location map, a circular layout for the social graph and the related attribute table, and series of multiple snapshots for representing change; the prototype of Twitter trending topic network used tweet map as map based visualization, a tree layout for the social graph and the related attribute table, and animation for representing change. The design working environments for two prototypes are the same: a multiple-view environment with the basic GIS functions, a time control panel and the linking and brushing function tool.

8. *Which usability method to use to decide upon the effectiveness of the designed prototype(s)?*

Chapter 6 discussed the evaluation of the two designed prototypes by using the focus group method. The objective of the evaluation is to test whether the designed prototypes can be used to help user to understand the spatio-temporal characteristics of the social networks from social media data. Specifically, the evaluation session focused on the usefulness of the representation of social network data and the designed working environment and its functionalities. As a result, by doing the test of each prototype, suggestions and comments were received for further improvement of both prototypes; the participants who are the domain experts concluded that the designed prototypes are useful to help people understand the spatio-temporal characteristics of the changing social networks from social media data.

7.2. Recommondations and future work

Recommendations for further research are:

- This research only studied two types of social networks from social media data, however, there are more types of social networks with a higher complexity. So the research should be broadened to other networks as well in the future.
- In this research, the visual variables used in map and social graph are only colour, shape and size. Actually from the focus group session, there should be more options like value or even multivariate symbols (e.g. pie chart, histogram for point symbols) can be applied to represent the social networks more comprehensively.
- The evaluation method in this research only used the focus group. In fact, it should be better if a combination of usability evaluation methods is used. For example, if using focus group first to get suggestions and make adaptions to the design, and then applying questionnaire to evaluate, the prototype obviously will be more useful.
- In this research, due to the limitation of time, the working environment is not fully implemented. The future work should implement the designed functionalities like linking the social graph to the map based visualization, and establish the time interactive environment. Moreover, it is also necessary to think about applying a visualization strategy like Shneiderman's seeking mantra to the environment for more complex situations.

LIST OF REFERENCES

Aggarwal, C. C. (2011). *Social Network Data Analytics*: Springer-Verlag New York Inc.

Anderson, J. C., Håkansson, H., and Johanson, J. (1994). Dyadic business relationships within a business network context. *The Journal of Marketing*, 1-15.

Andrienko, N., Andrienko, G., and Gatalsky, P. (2003). Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages & Computing, 14*(6), 503-541.

Barabâsi, A. L., Jeong, H., Néda, Z., et al. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications, 311*(3-4), 590-614.

Barbier, G., and Liu, H. (2011). *Data Mining in Social Media. Social Network Data Analytics*, 327-352.

Barthélemy, M. (2011). Spatial networks. *Physics Reports, 499*(1-3), 1-101.

Becker, R. A., Eick, S. G., and Wilks, A. R. (1995). Visualizing network data. *Visualization and Computer Graphics, IEEE Transactions on, 1*(1), 16-28.

Bertin, J. (1983). Semiology of graphics: diagrams, networks, maps.

Blok, C. A., Ormeling, F. J., and Kraak, M. J. (2005). *Dynamic visualization variables in animation to support monitoring of spatial phenomena.* Universiteit Utrecht,ITC, Utrecht,Enschede.

Carlos D. Correa, K.-L. M. (2011). VISUALIZING SOCIAL NETWORKS *Social network data analytic* (pp. 307-324): Springer.

Chen, H., Chung, W., Xu, J. J., et al. (2004). Crime data mining: a general framework and some examples. *Computer*, 50-56.

Collins, C. J., and Clark, K. D. (2003). Strategic human resource practices, top management team social networks, and firm performance: The role of human resource practices in creating organizational competitive advantage. *Academy of Management Journal, 46*(6), 740-751.

Cross, R. L., and Parker, A. (2004). *The hidden power of social networks*: Harvard Business School Press Boston, MA.

Cummings, J., Lee, J., and Kraut, R. (2006). Communication technology and friendship during the transition from high school to college. *Computers, phones, and the Internet: Domesticating information technology*, 265–278.

De Nooy, W., Mrvar, A., and Batagelj, V. (2005). *Exploratory social network analysis with Pajek* (Vol. 40).

Díaz, J., Petit, J., and Serna, M. (2002). A survey of graph layout problems. *ACM Computing Surveys (CSUR), 34*(3), 313-356.

Dodge, M., McDerby, M., and Turner, M. (2008). *Geographic visualization: concepts, tools and applications*. US: John Wiley & Sons Ltd.

Dykes, J., MacEachren, A. M., and Kraak, M. J. (2005). *Exploring geovisualization* (Vol. 1): Pergamon.

Erten, C., Harding, P. J., Kobourov, S. G., et al. (2004). *GraphAEL: Graph animations with evolving layouts*.

Escher, T. Researching the Geography of Social Relations.

Escher, T. (2007). *Where are your friends? The geography of social networks*.

Freeman, L. C. (2004). *The development of social network analysis*: Empirical Press Vancouver, BC.

Gaertler, M., and Wagner, D. (2006). *A hybrid model for drawing dynamic and evolving graphs*.

Goodchild, M. F., and Janelle, D. G. (2010). Toward critical spatial thinking in the social sciences and humanities. *GeoJournal, 75*(1), 3-13.

Guo, D. (2009). Flow mapping and multivariate visualization of large spatial interaction data. *Visualization and Computer Graphics, IEEE Transactions on, 15*(6), 1041-1048.

Hansen, D. L., Schneiderman, B., and Smith, M. A. (2009). *Analyzing social media networks with NodeXL*: Morgan Kaufmann.

Hong, L., Meng, F., and Cai, J. (2009). Research on Layout Algorithms for Better Data Visualization.

Ing-Xiang Chen, C.-Z. Y. (2010). Visualization of Social Networks *Handbook of Social Network Technologies and Applications*

(pp. 585-610): Springer-Verlag New York Inc

Kaplan, A. M., and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons, 53*(1), 59-68.

Khalili, N., Wood, J., and Dykes, J. (2009). *Mapping the Geography of Social Networks*.

Kraak, M. J. (2003a). Geovisualization illustrated. *ISPRS journal of photogrammetry and remote sensing, 57*(5-6), 390-399.

Kraak, M. J. (2003b). *The space-time cube revisited from a geovisualization perspective*.
Kraak, M. J. (Producer). (2010) Aspects of physical and social network visualizations. retrieved from http://intranet.itc.nl/papers/2010/pres/kraak_asp.pdf
Kraak, M. J., and Ormeling, F. (1996). *Cartography: visualization of spatial data*: Longman Harlow.
Kraak, M. J., and Ormeling, F. (2003). *Cartography: visualization of geospatial data*: Pearson Education.
Kwan, M. P., and Lee, J. (2004). Geovisualization of human activity patterns using 3D GIS: a time-geographic approach. *Spatially integrated social science, 27*.
Li, L., and Goodchild, M. F. (2010). The Role of Social Networks in Emergency Management: A Research Agenda. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM), 2*(4), 48-58.
Li, X. (2010). *The time wave in time space: a visual exploration environment for spatio-temporal data.* ITC, University of Twente, Enschede.
Li, X., and Kraak, M. J. (2008). The Time Wave. A New Method of Visual Exploration of Geo-data in Time&# 8211; space. *Cartographic Journal, The, 45*(3), 193-200.
MacEachren, A. M., Robinson, A. C., Jaiswal, A., et al. (2011). *Geo-Twitter Analytics: Applications in Crisis Management*.
McDonald, D. W., and Ackerman, M. S. (2000). *Expertise recommender: a flexible recommendation system and architecture*.
Mennis, J. L., Peuquet, D. J., and Qian, L. (2000). A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science, 14*(6), 501-520.
Milgram, S. (1967). The small world problem. *Psychology today, 2*(1), 60-67.
Mok, D., and Wellman, B. (2007). Did distance matter before the Internet?:: Interpersonal contact and support in the 1970s. *Social networks, 29*(3), 430-461.
Monmonier, M., and Gluck, M. (1994). Focus groups for design improvement in dynamic cartography. *Cartography and Geographic Information Science, 21*(1), 37-47.
Morgan, D. L. (1998). *The focus group guidebook* (Vol. 1): Sage Publications, Inc.
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, 84*(3), 441-461.
Radil, S. M., Flint, C., and Tita, G. E. (2010). Spatializing Social Networks: Using Social Network Analysis to Investigate Geographies of Gang Rivalry, Territoriality, and Violence in Los Angeles. *Annals of the Association of American Geographers, 100*(GEOBASE), 307-326.
Rafelsberger, W. (2008). Twitter Conversations Map. Retrieved 20th August, 2011, from http://www.visualcomplexity.com/vc/project.cfm?id=600
Reingold, E. M., and Tilford, J. S. (1981). Tidier drawings of trees. *Software Engineering, IEEE Transactions on*(2), 223-228.
Roberts, J. C. (2008). Coordinated Multiple Views for Exploratory GeoVisualization. *Geographic Visualization*, 25-48.
Rothenberg, R. B., Sterk, C., Toomey, K. E., et al. (1998). Using social network and ethnographic tools to evaluate syphilis transmission. *Sexually Transmitted Diseases, 25*(3), 154.
Scharl, A., and Tochtermann, K. (2007). *The geospatial web: how geobrowsers, social software and the Web 2.0 are shaping the network society*: Springer Verlag.
Shaw, S. L., and Yu, H. (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.
Shekhar, S., and Oliver, D. (2011). Computational Modeling of Spatio-temporal Social Networks: A Time-Aggregated Graph Approach.
Smith, M. A., Shneiderman, B., Milic-Frayling, N., et al. (2009). *Analyzing (social media) networks with NodeXL*.
Takhteyev, Y., Gruzd, A., and Wellman, B. (2010). Geography of Twitter networks. *Manuscript submitted for publication. Available from http://takhteyev.org/papers/Takhteyev-Wellman-Gruzd-2010.pdf*.
Technologies, M. (2003). Graph Layout Algorithms. Retrieved 01-11, 2011
Timo, O. (2006). *Mapping social networks in time and space*: Working paper, Institute for Transport Planning and System, and Swiss Federal Institute of Technology Zurich.
Valente, T. W. (2010). *Social networks and health: Models, methods, and applications*: Oxford Univ Pr.
Walker, I., and John, Q. (1990). A node - positioning algorithm for general trees. *Software: Practice and Experience, 20*(7), 685-705.
Wasserman, S. (1994). *Social network analysis: Methods and applications*: Cambridge university press.

Wellman, B. (1996). Are personal communities local? A Dumptarian reconsideration. *Social networks, 18*(4), 347-354.

Wills, G. J. (1999). NicheWorks: Interactive visualization of very large graphs. *Journal of Computational and Graphical Statistics*, 190-212.

Yang, F. (2011). *Use of different visual representation in the space time cube to explore changing network datasets.* University of Twente Faculty of Geo-Information and Earth Observation ITC, Enschede.