Spatio-Temporal Constraints on Social Networks Workshop
Final Report

Upham Hotel
Santa Barbara, California
December 13–14, 2010

I. Executive Summary

The impact of the Internet on human communication and the organization of social networks has been profound, greatly reducing the effects of distance and time differences. Strong spatial and temporal constraints persist, however, because of the importance of human contact and the spatial and temporal context of human actions. Recent research in social networks that places them in a meta-network context (multi-node, multi-link, multi-level) paves the way for exploring spatial and temporal effects. Yet to date only very preliminary efforts have been made to develop appropriate theory and models, or to calibrate them with the abundant data sources that are now available.

In December, 2010, the Center for Spatial Studies at the University of California, Santa Barbara held a workshop to explore spatio-temporal constraints on social networks, bringing together specialists drawn from the many disciplines with interest in these issues, including computer and information science, geography, mathematics, spatial statistics, and social network analysis. Forty participants from the academic, industrial, and governmental sectors attended. The workshop included plenary presentations by invited experts and small-group discussion of the issues, following a model developed more than 20 years ago by the National Center for Geographic Information and Analysis and refined in more than 40 such meetings. The two-day workshop was divided into four sections: an introduction, including 3-minute statements by each participant; five overview presentations and discussion; breakout group discussion; and a closing session.

The workshop assessed the current state of the art, identified and prioritized a research agenda, and began the development of an international community of collaborating scholars working on these issues. Specific questions addressed at the workshop included:

- What is the current state of knowledge with respect to spatiotemporal constraints on social networks and information flows, particularly from a meta-network perspective?
- How can theories of social network interaction be extended to incorporate the constraining effects of space, time, the Internet, and mass media?
- Can probability distributions be developed for networks or network metrics that are parameterized by spatial and temporal separation?
- What rich sources of data can be found to calibrate and parameterize these new models?
- What new metrics and models can be developed for assessing critical nodes, groups, and trails in and through networks that take spatio-temporal constraints, the Internet, and mass-media effects into account?
- Can we develop novel methods for visualizing the operation of spatio-temporal constraints and their effects on the flow of ideas and information through meta-networks?
- What methods of inference are appropriate for detection of spatio-temporal and network constraints in crowd-sourced data, and what are appropriate metrics of uncertainty?

Full details of the workshop, including the participant list, the position papers prepared by the participants, the presentations made during the meeting, reports of the various discussion groups, and the schedule of the workshop activities can be found at http://www.ncgia.ucsb.edu/projects/spatio-temporal/. In addition to the posting of workshop presentations on the website, other tangible outcomes of this workshop include this report, a paper on the role of social networks in emergency management in the *International Journal of Information Systems for Crisis Response and Management* (Li and Goodchild, in press), and a proposed special issue of the journal *Social Network Analysis and Mining* (SNAM) with contributions from workshop participants. Beyond these immediate and obvious outcomes, the ideas exchanged, the stimulating conversations, and the connections established during the workshop will inspire innovative research and facilitate collaborations in the long term.

This report follows the workshop's structure in summarizing the presentations and discussions in two sections; this is followed by the research agenda discussed at the meeting, a list of the participants, and acknowledgements. The summaries of the breakout-session discussions draw heavily on the summaries given by the session chairs and reporters. The summaries highlight the main points and are not meant to be a full account of what was said.

II. Presentations and Discussions on Day One

On Day One, the topic of the meeting was investigated from different perspectives: the social networking perspective by Dr. Kathleen Carley, the geospatial perspective by Dr. Michael Goodchild, the computational perspective by Dr. James Caverlee, the visualization perspective by Dr. Shih-Lung Shaw, and the social perspective by Dr. Matt Zook. Each participant also gave a short 3-minute perspective.

Overview I: The social networking perspective
Kathleen M. Carley, Carnegie Mellon University

Kathleen started off by defining network science as a discipline of studying the dots and trails in different networks to predict and explain behavior. There are two major components in a network: nodes and links. Examples of nodes include people, units of action, coalition partners, departments, resources, ideas or skills, events, nations, and states; while ties between nodes include friendship, transfer of resources, authority lines, association or affiliation, alliance, substitution, precedence, and proximity. Meta-networks can be pulled out of complex social networks to answer different kinds of questions. Take an event as an example: important questions that may be answered through social network analysis include: What happened? Where did it happen? Who were involved? How did it happen? Why did it happen? Techniques and tools have been developed to build social networks from text using data mining, to analyze links in social networks using statistics, and to assess different scenarios using agent-based "what if" analysis. Typical questions include: Who are the key players in a social network? How are they organized using techniques of fuzzy group clustering (*e.g.*, Fuzzy Overlapping Grouper, FOG)? Who are leaders and who are coordinators?

Networks are ubiquitous and socio-cultural data are everywhere. There are unstructured data in text (*e.g.*, interviews, news articles, blogs, and emails) and various online sources, semi-structured data in blogs, emails, and crowd-sourced data, and structured data in government and corporate documents and proceedings. Kathleen then discussed the three types of data using Sudan as an example. She continued to emphasize the importance of location in social network data by showing a map of the locations of crime around Washington in Google Earth. In terms of

geo-referencing information from text, there are different levels of challenges: some may be easy to georeference, while others are more difficult. In an article about the conflict in eastern and southern Sudan and Darfur, only about 40% of location references can be geolocated using gazetteers. Examples of social network analysis were given using the West Virginia High Technology Consortium (WVHTC) lunch data and origins and destinations in drug seizure. Incorporation of geographic information in the network measures improves the insights that can be gained. In addition to the spatial component, Kathleen also talked about the temporal component in social networks. The variation of social networks over time was demonstrated in graphs of email and face-to-face communications. Developments in Sudan were again used as an example to show the importance of temporal variation. From 2003 to 2008, several important events happened in Sudan and some important statistics were given in a table, such as the number of texts, the number of Sudan agents, the number of tribes and Sudan organizations, and so on. Based on social network data analysis, a comparison between 2003 and 2008 showed increasing complexity in tribal areas of conflict, especially in the areas of politics, administration, land, grazing, and water rights. Analysis on Sudan tribal networks also helped to identify the connections between apparently disconnected groups and the emergence of leaders within Sudan.

   Finally, Kathleen summarized the presentation by pointing out that sufficient metrics and visualizations for social networks exist; tools exist but they need to be speeded up. She identified three challenges in social network research: how to "drill down" to find information of interest from a combined spatio-temporal network perspective, how to identify locations in texts, and how to forecast the development of geo-temporal networks.

Overview II The geospatial perspective
Michael Goodchild, University of California, Santa Barbara

Michael started his presentation by proposing several big research questions on embedded social networks. Social networks are always embedded in geographic space and time. What can we learn from the embedding? What constraints does the embedding impose? What do we know about embedded systems that can inform research? The development of geospatial technologies enables us to collect geographic information much easier than in earlier times. For example, the Global Positioning System (GPS) makes the measurement of position a trivial issue. Remote sensing provides us with massive resources of imagery and ubiquitous, fine-resolution base maps in near real-time. Geographic information systems offer formal methods of representation of geographic phenomena and a wide variety of methods for spatio-temporal analysis and modeling. An example of an advanced emergency geographic information system developed by Loma Linda University showed the advantages of geospatial technologies in emergency management, as well as the challenge of interoperability. Interoperability requires easy exchange of data; however, existing methods for data exchange are primarily at the syntactic rather than the semantic level. Location referencing can be done both formally and informally. For example, there are five ways to describe the location of Michael Goodchild's house: (1) 119 52 14.4 West, 34 24 42.7 North, (2) 909 West Campus Lane, Goleta, CA 93117-4338, USA, (3) 3811560N, 236150E, Zone 11, Northern Hemisphere, (4) NE 1/4, Section 12, Township 23 Range 5 of the Second Principal Meridian, and (5) National Grid reference 11SKU36151156. Michael then identified weaknesses in current geospatial technologies: representation and analysis of time due to the legacy of static map-based information; representation and analysis of 3D information— existing GISs record elevation as a single-valued function of location and there is a lack of support for full 3D structures; representation and analysis of binary knowledge. Generally, there

are two types of spatial knowledge based on the number of involved locations: (1) Unary knowledge: knowledge about properties **z** present at locations **x** in space-time; when the knowledge is relatively static in time, it is usually expressed as maps, but this type of knowledge is increasingly becoming dynamic; (2) Binary knowledge: knowledge about the properties **z** of pairs of places in space-time $\mathbf{x_1},\mathbf{x_2}$; examples of this type of knowledge include distance, social affinity and interaction, travel time, flow, and proximity. This kind of knowledge is not ideally suited to mapping. Several examples of binary knowledge were given, including migration flows in the US, the famous map of Napoleon's March to Moscow in the campaign of 1812, and a generic flow model. Links are a particular example of binary geographic knowledge.

Michael then discussed how we can study social networks embedded in space and time, using major principles in geographic knowledge. We can infer processes from spatial and spatio-temporal forms. Geographic footprints on and near the Earth surface are very valuable for inferring possible processes that lead to such geographic forms. Two types of geographic context are very important. First, vertical context provides a variety of attributes associated with a particular location. If we want to study a particular attribute, we may ask the question: what else is known about this location. Second, horizontal context provides the same attributes associated with the neighboring environment. When studying a particular location, we may ask: what is known about nearby locations? This led to Tobler's First Law of Geography: nearby things are more similar than distant things. According to this law, two important geographic principles are spatial dependence and spatial heterogeneity. Spatial autocorrelation exists in nearby things, and there is usually an effect of distance decay: a general pattern observed in processes embedded in geographic space is that the interaction between two locations decreases as the distance between them increases. Spatial heterogeneity represents uncontrolled variance in spatial phenomena that are statistically non-stationary. Hence, a good spatial sampling design is required to effectively describe spatial heterogeneity. These two types of spatial patterns were demonstrated in Snow's Cholera map of 1854 and a map of cancer mortality by county in the US between 1950 and 1969. Finally, distance decay was demonstrated in the example of Wikipedia authorship. In Darren Hardy's work, 988,522 geotagged Wikipedia articles with 103,291 distinct locations were examined. He developed a "signature distance" metric, which is a weighted average of distances between author and article. For instance, for the entry about the University of California, Santa Barbara, 135 anonymous authors made 719 revisions, with a signature distance of 533km. For all geotagged articles, 64% of articles have a signature distance less than 2,000km. Given the distance decay function and the interaction between two locations, distance can be inferred.

Overview III The computational perspective
James Caverlee, Texas A&M University

James started his presentation using several impressive graphs to emphasize democratization of publishing and the huge amounts of data generated every day. According to Google's Eric Schmidt, every two days now we create as much information as we did from the dawn of civilization up until 2003. For example, the number of tweets per day increased to more than 45 million in January, 2010, within a mere four years of its launch. Hours of video uploaded to YouTube per minute increased from about 5 hours in June, 2007 to more than 35 hours in October, 2010. The number of unique visitors on Facebook increased from about 100 million in March, 2008 to more than 500 million in December, 2009. In this process, geographic information has been added to social media in FourSquare, Facebook, Google Latitude, and so on. Meanwhile, there is also democratization of computation as the amount of computational

resources per person rapidly increases. The increase of computational resources provides researchers with opportunities to run their analysis on a 1000-node computer cluster without worrying about data management, which makes it possible to do analysis that involves large quantities of data in a short period of time. Three examples were given to demonstrate the opportunity provided by big computation for analyzing big data. First, in a study that involves 3.5 million georeferenced Facebook users, probability of friendship is calculated as a function of distance normalized by population density which shows a strong distance decay effect. Second, in an extensive analysis of over 1.9 million MySpace profiles, language varies by location and age. Third, earthquake occurrence can be detected by monitoring tweets.

Although big data exist, there are also limitations. Researchers do not always get access to these data. You need to work for Facebook to get the user profiles. Alternatively, researchers can sample from Facebook, Gowalla, and Foursquare. However, challenges include potential bias in samples, uneasiness of sharing, and concerns over privacy. Although many social network websites expose a public API, they are primarily intended for partners and may ask for payment. Luckily, there are data that are inherently public, such as data generated on Twitter and Flickr. In addition to data access, there are also challenges in location granularity and location sparsity. In a large dataset of 1 million user profiles and 30 million tweets from Twitter, 21% list a location as granular as a city name; 5% list a location as granular as latitude and longitude coordinates; and only 0.42% of tweets contain geocodes. To overcome location sparsity, new methods for accurate and reliable geolocation of users need to be developed that require only public information, and can balance privacy and big-brother issues. One possible solution is content-based location estimation that considers the spatial distribution of words that are used in tweets. In addition to limitations on big data, there are also limitations on big computation. How to store big social, spatial, and temporal data? How to manipulate data and write efficient algorithms to analyze data? For example, traditional community-detection algorithms break down without significant infrastructure. How to put capabilities in the hands of the community? Take "crowds" on the real-time Web as an example. The real-time Web enables a fundamental shift from long-lived communities toward crowds that represent ad-hoc collections of users reflecting real-time interests; these communities are organic, short-lived, self-organized, and often implicitly defined. How can we identify and track online "hotspots" as they arise in real time in disasters, terror attacks, civil uprisings, and so on? How do crowds form and evolve? Hundreds of millions of users and highly dynamic and bursty interactions place huge demands on traditional methods. Regarding crowd discovery as clustering in time-evolving networks, James's group has developed a locality-based graph clustering framework with provable efficiency and quality guarantees.

James concluded his presentation by talking about possible research directions. While big data and big computation provide great opportunities to do research, where is the data in big data? What does the community need in terms of new algorithms and new toolkits?

Overview IV The visualization perspective
Shih-Lung Shaw, University of Tennessee

Shih-Lung started his presentation by citing visualizations of social networks from VisualComplexity.com. Some of the visualizations were based on the node-link network concepts and focused on topology and connectivity, including examples such as Facebook Visualiser, Facebook Tag Graph, and so on. Some examples on VisualComplexity.com visualize network growth and changes, such as Truthy, Social Collider, and so on; and some visualize time

and locations, such as the Twitter Conversations Map and the Invisible Cities. Then Shih-Lung focused on a space-time GIS approach to visualize social networks based on Hägerstrand's Time Geography and temporal dynamic segmentation. Hägerstrand's time geography examines human activities under various constraints in a space–time context. Space and time are connected through the concept of a space–time path, which depicts the sequence of an individual's activities at various locations over a time period. The possible locations that a person can visit within a given time window form a continuous space known as a space–time prism. Space–time paths can be divided through temporal dynamic segmentation. For example, in a typical day, a person's physical activities may include drive to work, walk to lunch, have lunch, walk back from lunch, drive back home, grocery shopping, and return home; while virtual activities include instant messaging with colleagues, and receiving a cellphone call from a spouse to do grocery shopping. In human activities, there are three types of constraints: (1) capability constraints: constraints that limit the activities of an individual because of his/her biological construction and the tools he/she can command; (2) authority constraints: constraints imposed by a domain within which things and events are under the control of a given individual or a given group; and (3) coupling constraints: constraints that deal with the interactions of an individual with other people or entities in a space–time context. As demonstrated in some graphs, space–time paths can be used to find locations, time, and potential interactions. These time-geographic concepts could be extended and applied to exploring and visualizing spatio-temporal constraints on social networks. Finally, Flickr data and Facebook data were visualized in a space–time GIS as a demonstration.

Overview V The social perspective
Matt Zook, University of Kentucky

Matt started his presentation by discussing the power and politics of code. According to Introna and Nissenbaum, "search-engine design is not only a technical matter but also a political one. Search engines are important because they provide essential access to the Web both to those with something to say and offer and to those wishing to hear and find." A comparison between two search results using the term "Tiananmen" provides a telling illustration of how different results can be obtained when one searches for images. The results from Google.com show images from the pro-democracy protests of 1989 while those from Google.cn (the only interface to Google available in China) show a series of smiling people and buildings. In the Google search engine, adding a comma to the end of search doesn't generally produce big changes, but for some terms, such as finance and health, the comma gives different results. Rather than getting the Google Brand (Google Health), other sites are featured with a comma in the keyword. Another example was about the misspelled name of a university. For about two weeks in October 2010, the University of Kentucky was labeled Transylvania University in Google Maps. In examples such as these, code helps to direct thinking in one direction rather than another. The second part of the presentation focused on the use of geographic representation to describe code, using several examples, including Charles' Booth's London Poverty Map between 1898 and 1899, the "Nations" of Facebook, and the Beer Belly of America. In addition to code and representation, participation is also an important factor in social networks. As shown in a screenshot from Scipionus.com Map that allowed people to make storm reports post Katrina, the visualization of the distribution of these notations provides a useful insight to the types of places that received attention from people using Scipionus. Several other examples were also given to demonstrate people's participation in online social networks, such as the Haiti map produced on OpenStreetMap, student riots and protests in London reported on Google Maps, and the locations

of facilities vital to US security in Google Earth. Matt then pointed out that geo-surveillance and sousveillance are another interesting topic in social networks. YouTube has been inaccessible in Iran since June 12, 2009, following the disputed Presidential election. Similarly, YouTube was inaccessible in Pakistan for 10 days in May 2010 due to concerns around the "Everyone Draw Mohammad Day" competition organized by a Facebook user. Code disciplines people by making them enact certain grammars of action and enforcing more pervasive modes of surveillance, automated management, and self-disciplining. An example from http://pleaserobme.com/ raised awareness about over-sharing and related privacy issues. The final topic in this presentation was merciless memory. Forgetting has its value. Pervasive surveillance and sousveillance have the potential to produce a society that never forgets—that has a permanent socio-spatial archive of trillions of events across a whole population, traceable through space and time; a detailed spatialisation of the history of everything, everywhere. Paradoxically everyware could well complicate life and introduce new technological hazards at the same time as it seeks to make life simple and reduce risk. According to Pickles, "Which lines are drawn, how we draw them, the effects they have, and how they change are the interesting questions." Finally, a table showed eight scenarios in a space–time–sociability context, based on the strength or weakness of sociability, the proximity or remoteness of space, and whether it is timely (synchronous) or timeless (asynchronous). In addition, there are other possible complexities in the technology of interaction (*e.g.*, face to face, analog, and digital), level of engagement, and multiple networks of sociability.

III. Breakout Sessions on Day Two

The second day was devoted to breakout discussion groups on specific topics. Participants were randomly assigned to the breakout groups. Reports from these groups to the whole assembly of workshop participants provided another opportunity for the exchange of ideas and the identification of priorities.

Breakout Session I: general questions

In the morning, participants were divided into four groups to investigate the same general questions on spatio-temporal constraints on social networks.

Questions
1. Does the domain need further clarification?
2. Where can we find the state of the art?
3. Are there missing voices?
4. What are some short- and medium-term research topics? What will be discovered?
5. What are the potential applications?
6. What infrastructure is needed?
7. What are the main impediments to progress?
8. What broader issues need to be addressed?

Group reports

Group 1: Hunt Room

Members: Carter Butts, Julie Cupples, Susan Enscore, Mike Gould, Brent Hecht, Shihlung Shaw, Loren Terveen, Arun Sen, Matt Wilson, Ed Pultar

1. Issues of representation—bringing together time, topology, and distance
   a. What are the appropriate distance metrics for modeling particular types of relations and relational processes? (shortest distance, cheapest, modes of transport)
   b. What are the right formalisms for representing the spatial properties of vertices and edges in particular settings? (When are points enough? Polygons? Space-time trajectories? Distributions thereof?)
   c. Are there ways to use topological properties to extend representation of spatially embedded systems? (*e.g.*, functional/schematic representations for spatially localized elements)
2. Impact of geography on social network structure
   a. What spatial interaction functions govern different types of networks?
   b. When are spatial effects spatially homogeneous? When not, how do they vary?
   c. When are place-based considerations needed to understand network structure?
3. Spatiotemporal effects on network-mediated processes
   a. How are information diffusion, adoption decisions, and influence affected by spatial context? (*e.g.*, adversarial interactions, technology use, volunteered geographic information (VGI; Goodchild, 2007), local expertise creation)
   b. How does spatial context affect mobilization and coordination? (*e.g.*, disaster response, community mobilization, local production)
   c. How is migration (at multiple spatial and temporal scales) shaped by social networks, and how are networks shaped by migration?
   d. How do tie structures within spatially defined communities affect social outcomes? (crime, economic development, individual attainment, risk)
4. Gathering information from spatially embedded populations
   a. How can geography affect or exploit the "human as sensor" paradigm?
   b. What combined spatial/network sampling strategies will allow us to measure spatio/temporal/social properties in hard-to-reach populations?

Group 2: Coach Room

Members: Jack Dangermond, Danielle Forsyth, Bill Goran, Carsten Kessler, Charles Perez, Martin Raubal, Chrisoph Schlieder, Waldo Tobler, Mike Worboys, Alan Glennon

1. Domain clarification
   a. Real social networks related to spatial-temporal
       i.   Separate from social networking tools
       ii.   Offline presence is a consideration
2. State of the art: Geo/temporal social networks
   a. New considerations of expertise and behavior
       i.   Multi-level meta networks
       ii.   Ubiquity of computing, sensors and handhelds
       iii. A lot more data
       iv.   Maps accessible
       v.   New mechanisms for network modeling
       vi.   Global perspectives and linkages
   b. Domains of expertise where spatio-temporal is a key consideration
       i.   Social sciences: sociology, psychology, geography
       ii.   Organizational behavior, marketing, and economics

       iii.   Mathematical part of anthropology/communication theory
       iv.   Physics and computer science
       v.   Spatial socio-networks
  c. Key topical focus for journals (special issues): *Journal of Social Networks*, *Computational Mathematical and Organization Theory* (CMOT), *Management Science*

3. Missing voices
  a. Organizations
       i.   Other academic—chemistry, genetics, other mathematics,…
       ii.   Technology companies
       iii.   Consumer Web companies
       iv.   Enterprise
       v.   Non-consumer organizations (*i.e.* NGO)
       vi.   Resources
       vii.   Real time-series studies, time–space series
       viii.   Computational networks and services
       ix.   Data formalization—access and discovery

4. Short-term and mid-term research
  a. Geo/temporal/mathematical (including network) reasoning
  b. Integration of time/networks/space
       i.   Animated network maps
  c. Dynamics (understanding change)
  d. Uncertainty, bias, expertise, relevance, and confidence as they relate to applications in spatio-temporal networks
  e. Event reasoning and contextualization
  f. Network re-organization
  g. Visualization
       i.   Clarity in presentation of social networks in spatio-temporal context
       ii.   Accessibility for different user types
  h. Data models for sparse and inconsistent information
  i. Data model forms—formalization and documentation
  j. Optimization for desired outcomes

5. What will be discovered?
  a. Network understanding as it relates to time and space in a dynamic environment
       i.   Human behavior
  b. Information diffusion
       i.   Need a spatio-temporal theory of network diffusion
       ii.   Gravitation/repulsion balancing (spring) related to spatio-temporal attractiveness

6. What are the potential applications?
  a Infinite, especially time critical
       i   Understanding market dynamics
       ii   Catching people of interest, influencing those that are influencable
       iii   Marketing
       iv   Transportation
       v   Health
       vi   Emergency management
       vii   Collaboration, public participation, forming/re-forming communities

       viii   Site selection, geodesign, building design, space planning, service delivery

       ix   E-tourism (and other kinds of new kinds of communities)

7. What infrastructure is needed?
   a   Data co-laboratories for access and discovery
   b   Cross training and pollenation
   c   Community co-participation
   d   Centers (NSF Initiative)
     - i   Data.gov or NIH data-sharing models
     - ii   Data use license for citing purposes
       - i) Geo-temporal tagging/enrichment system
       - ii) Spatio-temporal capture and recording
       - iii) Data models and metadata for common exploitation
   e   Computation and communication infrastructure
8. What are the main impediments to progress?
   a   Lack of
     - i   Theory
       - i) No axioms
     - ii   Bridges between core disciplines
       - i) Formalization
       - ii) Understanding of the value of the combined capabilities on the field
   b   A lot of amateurs in the mix
   c   Data
     - i   Formalization
     - ii   Availability and discovery
       - i) Appropriately spatially/temporally enabled, with understood representation
     - iii   Privacy
9. What broader issues need to be addressed?
   a   See impediments
   b   Too limited by US intelligence and defense focus
   c   Global perspective
     - i   Cultural mediation process
   d   Privacy

Group 3: Board Room

Members: James Caverlee, Hasan Davulcu, Grant McKenzie, Kostas Goulias, Don Janelle, Maureen O'Mara, Leah Vaughn, Matt Zook, Jamie Olson, Kitty Currier

1. Domain clarifications
   a   Nested (overlapping) social networks
   b   Study a range of social networks
   c   Predictive effects of: age, lifecycle, cohort
   d   Evaluating reliability and validity of data using spatial-temporal constraints (*i.e*. how to infer the real number of cholera cases, from tweets?)
2. Community mapping
   a   How to identify roles / levels / leaders / followers in social networks? How to sample?
   b   How to summarize? Is there a "master model" of roles and norms?

c   How to measure and improve the "efficiency" of social network processes for effective organization/spreading of information?
d   How to spatially/temporally/culturally "enable" a social network to accelerate rates of diffusion?

3. Potential applications
 a   Work on real problems and use cases (*e.g.* public health, disaster relief, epidemiology)
 b   Temporal-spatial dynamics (*e.g.* expansion, contraction) of real communities and data
 c   Outreach: How to influence opinion/behavior? —education

4. Spatial-temporal grounding of social networks
 a   Spatial range of social networks:
  i   Spatial vs. aspatial
  ii   Family, community vs. ideology
 b   Geographic grounding to reveal/exploit spatial patterns, context, and topology (linear, circular, hierarchical)
 c   Temporal range of social networks:
  i   Synchronicity vs. asynchronicity
 d   Relevance of history for the temporal dimension
  i   Time of significant events
  ii   Expose human patterning of achieving things
 e   We need network exploratory data analysis (EDA) tools to formulate and test hypotheses
 f   Trans-disciplinary network science:
  i   Ecology
  ii   Anthropology
  iii   Sociology
  iv   Psychology
  v   Ethics

Group 4: Garden Room

Members: Louise Comfort, Sarah Elwood, Russell Harmon, Huan Liu, Gonzalo Prokopec, Sally Schultz, Dan Sui, Connie White, Shashi Shekhar, Kristina Lerman, Linna Li

1. What broader issues need to be addressed?
 a   Must be able to model/analyze Cartesian and social/conceptual spaces within the same environment.
  i   These don't scale in the same ways
  ii   They are epistemologically and ontologically different
 b   We need appropriate (contextual) ways of expressing and measuring data quality and reliability
  i   Heterogeneity is not always error or mis-information
 c   What do we mean by a "network"? What characteristics define a "social network" for this research trajectory?
 d   Characterizing different types of social networks (*i.e.*, purpose, nature of interaction, social ties)
 e   This is a key precursor to being able to represent and analyze three principal elements—time, spatial, and social

 f Challenge of integrating data-handling techniques and infrastructure that have been used in social-networks research with those from GIScience, and blending top-down practices with those that are emerging from the grass roots, in crisis response.

2. Important questions

 a What conceptualizations have been used to differentiate networks, and are these sufficient for conceptualizing the social networks we see today?

 b How do we theorize precisely what is different about social networks in contemporary digital / material environments, as compared to the past?

 c How do contemporary social networks scale, in geographical spaces and social spaces, *i.e.*, interconnections between groups?

 d How do the data produced by various social networks change over time, and what are the implications for data quality, models, and analysis?

 e How do we conceptualize and analyze changes in social networks over time, ideally getting at changes in both their spatial and social dimensions together?

3. Missing voices?

 a Digital divide issues

 b What are the implications of using social networking data to examine social networks**?**

  i To do this, we must be able to characterize more precisely who are the social networking actors, and how they are linked to broader social networks?

4. What is the state of the art?

 a Complex systems research

 b Environmental policy (decision making under uncertainty)

 c Digital humanities (accommodating multiple epistemologies, representations)

 d VGI – as a manifestation of one type of socio-spatial network

5. Main impediments

 a Data are often quite partial, incomplete

  i systematic ways to make inferences under conditions of uncertainty

  ii "tolerance" for uncertainty is context-specific

 b Access to data sets (public and private)

 c Analytical procedures for getting at dynamic aspects of this phenomenon

 d Algorithms and data structures prepared to handle multiple dimensions

 e Synthesis and integration techniques for handling very different forms of data

6. Infrastructure needed?

 a Data and ways to share, exchange, and collaborate

 b Web environments for controlled experimentation with existing data

 c Education— to prepare scholars and practitioners

 d Identify appropriate funding sources

Breakout Session 2: development of research projects

In the afternoon, the participants were divided into breakout groups according to five major interests in research projects: theories, modeling and simulation, visualization and analysis, ethics, and data issues.

<u>**Group reports**</u>

The theory group: (some) theories in embedded networks

1. What theories are needed
   a   What is a network?
   b   How do we define edges and vertices over space? How do they change over time?
   c   Dynamics:
      i   Space: changes in physical space and changes of edges in space
   d   How does network theory adapt to the spatio-temporal domain? It adds or removes complexity?
      i   Effect of spatial dependence
      ii   "areas of influence"
      iii   Common demographics
   e   Exogenous effects on network formation and structure?
   f   Multi-level networks and organization
      i   Hierarchy of network structures: people, community, organizations
      ii   Temporally shifting networks (periodicity = market changes)
      iii   No theory of multi-level networks!
2. Ways of accounting for explicit space and time
   a   Look at other fields and applications:
      i.   Flow of information
      ii.   Disease spread
      iii.   Resource sharing
      iv.   Coordination/response
   b   Network design, based on resources at different locations
   c   Pattern formation and change.
      i   Point-process model plus network model to construct spatio-temporal networks.
      ii   Time and scale influences formation, how can we model that?
3. How are ties preserved? Is group maintenance and persistence dependent on spatial location? Performance of a social network embedded in space and time:
   a Task outcome
   b Perceptions of social networks
   c Organizational
   d Completion
   e Territory/effective control
4. Need for focused studies tailored to gathering network data to guide and inform theory

The modeling and simulation group: What is it (in this context)?

1. Modeling
   a Models for discovering uncertain facts in spatio-temporal information that point to commonalities or connections
   b Algorithm perspective—interested in things that might be important in modeling and simulation—and evaluating effectiveness at different scales
   c Policy—modeling large-scale functions that organizations perform—to assess the balance between resilience and entropy while keeping large-scale systems operational yet improving secondary and tertiary consequences
      i   Cascades of failure and recovery
   d Abstraction—multi-level components from instance to component to system views, using different levels of detail

 e Focus on reusability

 f Simulation—uses these models to project, hypothesize, evaluate, and improve underlying models, and to evaluate against goals. Networks of spatially situated actors and groups—using models to evaluate decision impact.

  i   Understanding right levels of detail

 g Model the embedding of social networks in spatially and temporally bounded environments

  i   Re-evaluating relationship approaches for embedding reification

 h Abstractions that apply to

  i   structure

  ii belief/concept

  iii outcomes/goals

  iv workflow

  v rules/constraints

  vi analysis (math)

  vii systems

2. Example: transportation

 a Metro region

  i   Infrastructure (*i.e.*, roads, mass transit, car pool)

  ii Want to motivate people to use green transportation options

  iii Social networks can change quickly

  iv Social networks are affecting the spatio-temporal model

  v Infrastructure being provided is not keeping up with the social networks

  vi Cost (time and money)

   i) Economic value/cost

   ii) Energy value/cost

   iii) Social value/cost

The visualization and analysis group:

1. Visualization

 a What can we learn from cartography?

 b Telling a story vs exploratory data analysis

 c Scale is a major issue, and we need generalization!

 d What properties do we want to keep? Topology?

 e How can we tie the geographic to the social space? we need to keep track of identifiers!

 f Dual view: geographic vs schematic

2. Analysis

 a We need a proper model for "spacetime" to

  i   handle streams

  ii detect change

  iii give up "slicing" things up

 b Throw away exact location in favor of a notion of nearness?

 c There is a need for line-based clustering

3. Applications

 a Visualizations and analysis methods are highly dependent on the application

 b Use spacetime as a secondary filter for analysis on a conceptual space

    c   We need to classify applications into groups and assign each the most useful analysis methods
4. Research questions
    a   What principles from cartography apply?
    b   How do they need to be transformed?
    c   To what extent is it visualization, and to what extent is it "just" calculations to answer questions?
    d   How can we reduce complexity?
    e   Maps not just to make it look nice, but to get the message across
    f   How can we visualize spatio-temporal change in social networks?
    g   How do I identify emerging and abnormal patterns in spatio-temporal social networks?

The ethics group: ethics/legal and social

1. Important research issues:
    a   When is informed consent called for in these applications and what would it actually look like?
    b   When is it appropriate to gather information from or about people without their knowledge? To re-use data for another purpose or aggregate with other data?
    c   Challenges posed by cross-institutional differences in ethical norms and acceptable practices
    d   Legal structures and regulations not geared to the issues raised by these data and practices
    e   Ethical concerns will vary with the nature of data representation—consider numerical vs. image—level of abstraction matters.
2. Ethical complications across the research process
    a   How do you operationalize data categories? For example, who is a "terrorist"?
    b   How do you collect data?
        i   Informed consent
    c   What assumptions drive the analysis? For example: spatial presence may not indicate behavior or involvement in a network
    d   How will data be shared?
        i   Aggregation, disaggregation, unintended consequences?
        ii   Can ethical considerations be embedded in metadata?
3. Questions for further study— ethics:
    a   In what ways are existing codes of ethics inadequate to deal with challenges posed by the new forms of data we are talking about working with?
    b   What practices are effective for ensuring ethics consultation and consent in this context? (High volume, time sensitive)
    c   What practices, restrictions, standards may protect against unethical or malicious third-party uses of collaboratively created spatial data?
4. Broader questions for social researchers:
    a   What is the relationship between digital and non-digital social networks?
    b   In what ways are digital social networks biased?
    c   What does this mean for the conclusions we may draw from these forms of data?
    d   What are the assumptions about human motivations and behaviors that underlie social networking services?

e   What does this mean both for the social networks that emerge through them and for the way we study them?
      i   Location-based services built around consumption;
      ii   Spatial proximity does not necessarily imply social proximity

The data issues group

1. Global issues
    a   How to get standards to represent the spatio-temporal context of interactions?
    b   How to get more spatio-temporal tagged data?
    c   Where to get these data?
    d   Which data do we need?
    e   What granularity of data is relevant?
    f   What amount of data is useful?
2. These questions lead to the requirement for a unique and global data service
    a   Data service that provides case-study models (templates) could help users to manage their data.
    b   The data service could also:
      i   evaluate the quality of data
      ii   improve data with tags
      iii   give some advice on the best scales to choose for different problems
3. Unified data-service platform
    a   Security issues
    b   Quality issues
    c   Socio-cultural issues
    d   Collection issues

IV. Research Agenda

The following research agenda was developed based on the presentations and discussions during the workshop. The most substantial research topics are listed below.

1. Data access and data discovery

Data access and data discovery are two important issues to consider because legitimate data are fundamental for scientific research and decision-making. How can we collect more spatial and temporal data of social networks from spatially embedded populations, both online and offline? How can geography promote the "human as sensor" paradigm in spatial data generation? How can we collect social-network data about involved people from existing sources? How can we identify useful social networks from masses of disparate data? Is there an automatic way to discover relevant information for a particular application over the Web when a range of tools and websites are used by different groups of people? What would be the essential features for a global data service that harvests data from the Web and other sources, evaluates data quality, and advises users about appropriate data according to their purposes?

2. Synthesis of social network data

Faced with information from various sources with different levels of uncertainty, another crucial question is how to synthesize information from various sources, ideally in real time. Provenance and uncertainty of different sources should be maintained in synthesis, which is still

a challenging issue in the Network Science, database and GIScience communities. How to conflate geospatial data with various accuracies, different levels of detail, and different generalizations is still an open question. This is less true for network data where transformation tools for going from one format to another exist. How can we quantify the uncertainty in a piece of information? Is it possible to produce automatically a probability distribution of data values from multiple sources? Semantic interoperability is also critical in integration of data from different social groups who use different concepts to think and different languages to communicate. In addition, what data type would be the most appropriate form for integration and synthesis of various sources to facilitate reasoning and decision making? For instance, should all information be linked by geographic locations and conflated into a base map to provide a holistic view of relevant knowledge? Or should all data be linked by the actors to provide a holistic view of the relevant knowledge? Is there a way of easily moving between a spatial and a network view without forcing the network to be a spatial layer and to allow the actors in the network to appear in multiple locations in the same data? Is there an automatic way to seek out the relevant data from large quantities of social network data at an appropriate scale? Furthermore, how can we combine data from different social-networking services, *e.g.*, Twitter, Facebook, or phone records?

3. Human motivations and behaviors underlying social networking services

Besides the use of alternative data sources, it is also very important to understand the emerging pattern of people who join social networks. A more general research issue concerns how a social network forms, is maintained, and evolves with time or as an event develops. What are the major types of motivation of people who join online social networks? Can traditional social network theories in sociology be used to explain human behaviors in online social networks? How can we promote participation and accelerate rates of information diffusion through socio-psychological incentives? How do exogenous factors affect network formation and structure?

4. Analysis and modeling of social networks

Social network analysis is promising for understanding the current situations of involved people, but we do not have adequate methods to conduct analysis. First, it is helpful to distinguish between real social networks and implied social networks. Real social networks are the complex relationships of people at various levels, and implied social networks are inferred from human activities. Any data we collect using either online social-network services or cellphone records are only an approximation of real social networks with simplification or even misrepresentation. Therefore, it would be useful to investigate the uncertainty of collected social-network data. What is the percentage of relationships that are represented in a particular social-network data source? How should we identify different roles of people in social networks, *e.g.*, leaders and followers? How do we identify emerging and abnormal patterns in spatio-temporal social networks? What models can be developed to simulate social networks embedded in spatially and temporally bounded environments, and to evaluate the impacts of possible decisions? Is it possible to develop a theory for multi-level networks? How can we integrate existing social-network graphs with location-based services to analyze the geographic distribution of participants of a social network at a particular time-space point? What models can be developed to discover hidden commonalities and connections in spatio-temporal social networks? What is the role of simulation in evaluating model effectiveness and the impact of possible decisions? When the social networks are evolving then trails exist. How can we

integrate existing trail graphics and analytics with maps? Would such integration support logistic forecasts for populations?

5. Representation and visualization of social networks and spatial data from a combined perspective

Representation of complex multi-level social networks is also a big challenge. Is there a way of using spatial information to effectively generalize large complex social networks, or to represent sparse and inconsistent information in a way that makes the resulting analysis actionable? How can cartography and geovisual analytics contribute to representations of spatially embedded social networks? How can social network visualization services contribute to the representation of logistics and spatial utilization? When large volumes of data are available, what would be an effective mechanism to use to sample? What are the important properties to maintain when visualizing spatially embedded social network data: topology of nodes and links, geographic locations, relative spatial distance, or other attributes? Similarly, what are the important geographic relations to retain when visualizing the spatially embedded social network: exact location, region, land-mass features, or other attributes? What techniques are useful to describe the relationships between networks in a geographic space and that in a social space? What would be an effective way to represent the temporal dimension in social networks? How can we use changes of edges in network graphs to represent changes of networks in physical space? How should we visualize social networks at different scales? How can we visualize the interactions between different networks? Because simple social networks are usually linear data, can we develop techniques for line-based clustering? Note dynamic meta-networks are not linear. Network data are usually represented as graphs composed of nodes and links. What are other possible representations for network data: polygons, trajectories, or other spatial forms? Spatial data are usually represented as maps. What are other possible representations of spatial data: feature clusters, proximity networks, *etc*.? Visualization and analysis of social networks and geo-spatial data are highly dependent on the applications. How can we classify applications that involve both social network data and spatial data into different categories and develop useful analysis and visualization methods accordingly that are valuable for a combined network-spatial perspective?

6. Spatial and temporal considerations in social networks

Social networks are always embedded in space and time. However, that embedding may not be relevant to the analysis. How can we incorporate spatial interaction functions into different types of network particularly when space and time vary significantly? For example, spatial dependence and distance decay are valid in many processes in geographic space; but are less relevant when all people in the social network are in the same room. While it is known that social network links decay with distance, and that new technologies do not completely overcome this decay; what is not known is the circumstances where the technologies overcome the decay (*e.g*., where in a task cycle). Moreover, the predictive value of knowing the network links fall off with distance cubed seems low. What can be predicted with a better understanding of the relationship between space and networks? What are the types of activities in social networks that are strongly constrained by space and time? What types of social networks emerge from different spatial environments? How does the social network change the way that the space is used? For what type of groups is group maintenance and persistence dependent on spatial locations? What would be a proper space–time model to handle nodes and links in social networks? The meta-network is one such possibility. What more is needed? How can we explicitly represent space

and time in social networks? What more is needed beyond what is in dynetml? How can we use space and time as constraints to infer missing network data, select specialized social network subgroups, and forecast change? Which aggregation functions other than unioning the data are critical to use as the "size" of the spatial and temporal "chunk" changes? What would be an appropriate way to describe virtual social networks in the digital world where Euclidean distance may not be a proper measurement for nearness? Can a compendium of physical distances be created that is more meaningful from a "contact" perspective than Euclidean distance? How are network-mediated processes (*e.g.*, information diffusion, VGI) influenced by spatial and temporal relations (*e.g.*, nearness in space or time)? How can spatial and temporal constraints be used to validate data quality?

7. Interaction between different networks

Social networks are embedded in geographic space and time and nested with other networks, such as knowledge, activity, transportation networks and information networks. In some cases the way change in one network impacts another is, at least partially understood. For example, as individuals interact in the social network they share information leading to changes in the social network leading to changes in the activity network. These networks tend to co-evolve leading to, in established groups, correlations between the social network defined on interaction and that defined on shared knowledge or shared activities. In other cases, correlation may exist between movements in one network and activities in another network. For example, movements of cars between two locations on the road network may reflect a connection of people at the same two locations in a social network. Another example is that the activities in online social networks (*e.g.*, comments on an event) may reflect activities in the physical world (an event that people are currently experiencing, such as a football game or a fire). The study of the interactions is made possible with the development of location-based services and the large volumes of geospatial data generated by them. However, novel methods and new techniques need to be developed to capture the correlation between networks and to represent and analyze spatio-temporal dynamics of social networks. MRQAP is a standard approach for assessing the correlation of two networks based on the same nodes, *e.g.*, the network of who interacts with whom with the network of who shares what information with whom. Are there other approaches or metrics needed when the networks are based on different types of nodes? What metrics can be used to describe correlation between different networks and to identify motifs within these networks that have special correspondence (as opposed to a statistical correlation)? How can we use knowledge about social networks to guide policies regarding other networks? For instance, when new transportation infrastructure is needed, where would be the optimal location when the intensity of social network connections is taken into account? A strong connection between people at two locations may indicate more traffic between these two locations on the road.

8. Ethics: privacy and equity

In social-network studies, there are two major ethical issues: privacy and equity. When is it appropriate to collect information from people and to study people without their knowledge? When is informed consent necessary for initiating research? When locations of residents are revealed to the public, this information may invite future crimes, especially for disadvantaged people. Is there a way to preserve spatio-temporal patterns of social networks and to protect privacy at the same time? What types of generalization and aggregation from statistics and cartography can be adapted to achieve the two objectives? What would be an appropriate level of abstraction for a particular application? How does the level of abstraction limit the types

of network questions that can be answered? In data-sharing projects, what practices and restrictions are necessary to prevent malicious uses of spatial data and spatially embedded network data? Another issue concerns the various degrees of information accessibility between different groups of people. How can we reach people without access to mobile phones, computers, and the Internet? How can we disseminate relevant information to people who don't use online social-networking services? Online social networks are only a small portion of the representation of real social networks. How can we collect data on social networks that are not represented in the digital world? What spatial sampling strategies will allow us to measure spatial, temporal, and social properties in hard-to-reach populations?

9. Potential applications, especially in time-critical situations

Many applications depend on social network data: data generated by social networks and data about social networks, especially in time-critical situations. Examples include public health, epidemiology, organizational redesign, workforce assessment, and disaster relief, *etc*. The research topics listed above need to be addressed in specific applications. Take emergency management as an example. When an emergency happens, how can we find and access relevant information generated by different social groups? In addition, there is always a trust issue in accepting information generated by volunteers. An individual may be faced with a choice between available but potentially unreliable information synthesized by volunteers, and authoritative yet possibly unavailable information from government agencies (Goodchild and Glennon, 2010). Should he or she wait for the official information from government which may be slower, or should he or she trust VGI that is asserted without validation? When different social networks are combined to perform a task towards a common goal (*e.g*., emergency response), how can information technology be used to reorganize social networks based on geographic proximity, expertise, skills, *etc.*, to perform tasks better? How does the number of volunteers increase as an event develops? What are the socioeconomic characteristics of habitual volunteers in generating and disseminating information? People rarely act independently in emergencies, but rather evacuate within groups. Parents may go to school to pick up their children before leaving town. How can we use correlation between different networks to plan an optimal rescue plan such as an evacuation route? Furthermore, how can we get the required help for people in need and meanwhile protect their privacy?

V. Participants

Name	Affiliation
Carter Butts	University of California, Irvine
Kathleen M. Carley	Carnegie Mellon University
James Caverlee	Texas A & M University
Louise Comfort	University of Pittsburgh
Julie Cupples	University of Canterbury
Jack Dangermond	ESRI
Hasan Davulcu	Arizona State University
Sarah Elwood	University of Washington
Susan Enscore	CERL
Danielle Forsyth	Thetus Corporation, Portland, Oregon
Alan Glennon	University of California, Santa Barbara

Michael Goodchild	University of California, Santa Barbara
Bill Goran	CERL
Michael Gould	ESRI
Kostas Goulias	University of California, Santa Barbara
Russell Harmon	Army Research Office
Brent Hecht	Northwestern University
Carsten Kessler	Muenster University
Kristina Lerman	University of Southern California
Linna Li	University of California, Santa Barbara
Huan Liu	Arizona State University
Jamie Olson	Carnegie Mellon
Maureen O'Mara	National Geospatial-Intelligence Agency
Charles Perez	UT Troyes
Gonzalo Prokopec	Emory University
Edward Pultar	University of California, Santa Barbara
Martin Raubal	University of California, Santa Barbara
Sally Schulz	National Geospatial-Intelligence Agency
Arun Sen	Arizona State University
Shih-lung Shaw	University of Tennessee
Christoph Schlieder	University of Bamberg
Shashi Shekhar	University of Minnesota
Daniel Sui	Ohio State University
Loren Terveen	University of Minnesota
Waldo Tobler	University of California, Santa Barbara
Leah Vaughn	The Healthpolicy Group
Connie White	Jacksonville State University
Matt Wilson	Ball State University
Mike Worboys	University of Maine
Matt Zook	University of Kentucky

VI. Acknowledgements

References

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

Goodchild, M.F. and Glennon, J.A. (2010). Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth, 3*(3), 231–241.

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): 49–59.