Title
The Role of Geography in Social Networks: CouchSurfing as a Case Study

Author
Pultar, Edward

Publication Date
2011-05-01

University of California Transportation Center
UCTC Dissertation UCTC-DISS-2011-06

**The Role of Geography in Social Networks:
CouchSurfing as a Case Study**

Edward Pultar
University of California, Santa Barbara
2011

UNIVERSITY OF CALIFORNIA

Santa Barbara

The Role of Geography in Social Networks: CouchSurfing as a Case Study

A Dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Geography

by

Edward Pultar

Committee in charge:

Professor Martin Raubal, Chair

Professor Michael F. Goodchild

Professor Konstadinos Goulias

Professor Stephan Winter

September 2011

The dissertation of Edward Pultar is approved.

Stephan Winter

Konstadinos Goulias

Michael F. Goodchild

Martin Raubal, Committee Chair

May 2011

The Role of Geography in Social Networks: CouchSurfing as a Case Study

ACKNOWLEDGEMENTS

Many thanks go to all of my advisors for their inspiration and advice over the years, especially Martin. All of my family has always been very supportive of me staying in school. My parents and siblings are exceptionally wonderful influences for me. I thank multiple sponsors for funding this research the past four years: Department of Geography at UCSB, Konstadinos Goulias and the University of California Transportation Center (UCTC), Jack and Laura Dangermond, and UCSB's President's Dissertation Year Fellowship. Thanks to Srinath Ravulaparthy, Kate Deutsch, Nate Isbell, and the rest of the UCSB GeoTrans Laboratory for their help with count-based models.

Some material in this dissertation is based on the following:

Pultar, E., Raubal, M., and Goodchild, M. (2008) GEDMWA: Geospatial Exploratory Data Mining Web Agent. In *16th ACM SIGSPATIAL Conference*, eds. H. Samet, C. Shahabi, & O. Wolfson: Nov. 5-7, 2008, Irvine, CA, USA, 499-502.

Pultar, E. and Raubal, M. (2009) Progressive Tourism: Integrating Social, Transportation, and Data Networks. In Sharda, N. (ed.) *Tourism Informatics: Visual Travel Recommender Systems, Social Communities and User Interface Design*. IGI Global, 145-60.

Pultar, E., Winter, S., Raubal, M. (2010) Location-Based Social Network Capital. In *GIScience 2010 Extended Abstracts Vol.*, eds. R. Purves & R. Weibel: Sept. 14-17, 2010, Zürich, Switzerland.

VITA OF EDWARD PULTAR

June 2011

EDUCATION

Bachelor of Science in Computer Science, Minor in Mathematics, University of Utah, June 2005
Bachelor of Science in Geography, University of Utah, June 2005
Master of Science in Geography, University of Utah, June 2007
Doctorate of Philosophy in Geography, University of California, Santa Barbara, May 2011 (expected)

PROFESSIONAL EMPLOYMENT

2004-2007: GIS Specialist, Structural Geology International, Salt Lake City, Utah
2005-2007: Research and Teaching Assistant, Department of Geography, University of Utah
2007-2010: Teaching Assistant, Department of Geography, University of California, Santa Barbara
2010: GIS Engineer Intern, Google, Mountain View, California
2011: Guest Seminar Teacher, Universitat Jaume I, Castellón, Spain

ARTICLES IN REFEREED JOURNALS

Pultar, E., Raubal, M., Cova, T.J., and Goodchild, M.F. (2009) Dynamic GIS Case Studies: Wildfire Evacuation and Volunteered Geographic Information. *Transactions in GIS,* 13(s1), 85-104.

Pultar, E., Cova, T.J., Yuan, M., and Goodchild, M.F. (2010) EDGIS: A Dynamic GIS Using Space Time Points. *International Journal of Geographical Information Science*, 24, 329-346.

ARTICLES IN REFEREED CONFERENCE PROCEEDINGS

Pultar, E., Raubal, M., and Goodchild, M. (2008) GEDMWA: Geospatial Exploratory Data Mining Web Agent. In *16th ACM International Symposium on Geographic Information Systems, SIGSPATIAL ACM GIS 2008*, eds. H. Samet, C. Shahabi, and O. Wolfson: November 5-7, 2008, Irvine, California, USA, 499-502.

Pultar, E. and Raubal, M. (2009) A Case for Space: Physical and Virtual Location Requirements in the CouchSurfing Social Network. In *International Workshop on Location Based Social Networks (LBSN'09) at 17th ACM International Symposium* *on Geographic Information Systems, SIGSPATIAL ACM GIS 2009,* eds. X. Zhou and X. Xie: November 3, 2009, Seattle, Washington, USA, 88-91.

Pultar, E., Winter, S., Raubal, M. (2010) Location-Based Social Network Capital. In *GIScience 2010 Extended Abstracts Volume*, eds. R. Purves and R. Weibel: September 14-17, 2010, Zürich, Switzerland.

Pultar, E. (2010) The Synergy of Social, Transportation, Communication and Data Networks. In *Proceedings of the GIScience 2010 Doctoral Colloquium*, eds. J.O. Wallgrün and A.-K. Lautenschütz: September 18, 2010, Zürich, Switzerland, 63-68.

REVIEWED BOOK SECTIONS

Pultar, E. and Raubal, M. (2009) Progressive Tourism: Integrating Social, Transportation, and Data Networks. In Sharda, N. (Ed.) *Tourism Informatics: Visual Travel Recommender Systems, Social Communities and User Interface Design*. IGI Global, pp. 145-160.

AWARDS

Excellence in Teaching Award, UCSB Department of Geography, 2009

President's Dissertation Year Fellowship, UCSB, 2010

Jack and Laura Dangermond Fellowship Award for Excellence in GIScience, 2010

University of California Transportation Center (UCTC) Dissertation Research Grant

OTHER

Over 300 hours of volunteer work for the 2002 Winter Olympics in Salt Lake City

Intermediate Spanish (Castellano) conversationalist

Founded U of Utah's juggling club, president for 6 years while coordinating and performing juggling events at Salt Lake 2002 Winter Olympics, Utah Opera, TV shows, university functions, Utah Jazz, music concerts, elementary schools

Organizer of Club Juggling Club at University of California, Santa Barbara, which hosts annual juggling festival donating proceeds to Santa Barbara Rape Crisis Center

Volunteer juggling teacher and performer for Isla Vista Elementary School students

ABSTRACT

The Role of Geography in Social Networks: CouchSurfing as a Case Study

by

Edward Pultar

Social networks are ubiquitous in the mobile information society of the present day. Here the focus is on social networks that depend on the physical and virtual locations of their users in order to provide various services. In these contemporary social networks both virtual and physical presence is a requirement. This research examines travel behavior using an Internet-based website, CouchSurfing, which provides free lodging with local residents. Increases in computing power and accessibility have led to novel e-travel techniques and the users of such systems utilize an amalgamation of social networks, transportation networks, and data communication networks. Thus the focus is on how the geographical spread of people in a modern, digital social network influences the travel choices of each individual in the network. In this dissertation a general model is presented that describes traveler behavior using a cost-free lodging network. Also presented for this type of travel behavior is an information representation and visualization methodology utilizing time-geographic dimensions. Two surveys with human participants were completed as part of this research. One survey concerns the factors

that affect the number of times an individual participates in an activity in the network: being a host or guest. The other data set collected for this work examines social capital and how it is influenced by geography and the distinct roles a person can play within the social network. The chapters are tied together in their focus on travel using Internet-based social networking and lead to new conclusions about the power and potential for contemporary social networks. These conclusions include a conceptual model of travel using a hybrid social network, the value of playing different roles in a social network, and the importance of geography in a social network.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

A fundamental issue in geography is the role of information in patterns of human behavior. In recent years the flourishing of technology has brought many conveniences, connections, and frustrations for people across the world. Specific to this research, the digital age provides novel ways for humans to discover geography. Means of mobility have greatly increased over the decades, and we are able to travel farther and faster than ever before; however, travel has also become less necessary in some circumstances through the integration of virtual communication and meetings. Physical presence is still required in many instances, but are we seeing an increase or decrease in *time-space compression* (Harvey 1989)? Part of the interest in time-space compression is the effect that modern technology has on a human's use of time and capabilities for travel.

Modern, Internet-based social networks allow individuals to connect unlike ever before with large, geographically spread, culturally diverse, and yet, maintainable structures. The trends for popular general social networks have moved in recent years between Friendster[1], MySpace[2], Orkut[3], Facebook[4], and others. Their use has become commonplace in daily life with businesses worldwide encouraging their

---

[1] http://www.friendster.com
[2] http://www.myspace.com
[3] http://www.orkut.com
[4] http://www.facebook.com

shoppers to become their fans or friends in social networks. Through Groupon[5] and LivingSocial[6] merchants provide discounts to those that can arrive with acquaintances from their social networks. Combining social networks and mobile phones with a Global Positioning System (GPS) lets users "check-in" at places, build their status as a local of an establishment, and more easily locate the physical position of people in their social circle. Social networks are also affecting the entertainment industry with prestigious film awards being given to a movie telling the story of how a social network was created by university students[7].

At their core social networks have geographical components such as the location of a member and their friend links. Geographers have strong contributions to make in social network research and had a strong presence at the recent *Spatio-Temporal Constraints on Social Networks* (http://www.ncgia.ucsb.edu/projects/spatio-temporal/) two-day workshop hosted by spatial@ucsb in Santa Barbara, California, in December 2010. Social networks, their influence on geography, and vice versa will continue to grow in the coming years hence making the research problems covered in this dissertation relevant in contemporary and future times.

1.1 Problem Statement

Information and Communication Technologies (ICT) such as wireless and phone networks are critical for the success of modern, Internet-based social networks. ICT

---

[5] http://www.groupon.com
[6] http://www.livingsocial.com
[7] http://www.oscars.org/awards/academyawards/83/nominees.html

allows for the seamless integration of audio and visual functions regardless of the device being used for data access. While travel behavior with ICT, global transportation, and social networks (Larsen et al. 2006a) are currently popular research subjects, a geographer's perspective on the interconnectedness of these topics is needed to shed new light as to how all of these impact spatial behavior.

Contemporary technologies allow for more and more opportunities for virtual presence in social settings such as social networks. Individuals create online personas that may or may not be accurate. While purely virtual interactions are possible, the actions are still rooted in the real, physical world. Hence there is an intrinsic connection between virtual and physical presence that is the theme central to this dissertation research.

Individuals are able to play different roles in an online social network depending on time. Social network members may contribute by providing a resource for other members at one time and at another time they may take and use resources provided by others. Thus the following document investigates how social capital is influenced by the performance of different social network roles and the geographical distribution of connections.

Demographics of individuals can also affect one's ability to take part in a social network. Specifically age, gender, and place of residence afford both enablers and barriers for social network contributions. These variables and their influence on social network participation are studied here using count-based models.

Collecting data is an important part of doing work with social networks. Data mining techniques allow for automated data gathering from sources with similar structures. For example, social network profiles for different individuals vary but still share elements such as a profile picture, age, and gender. Utilizing data mining with web data is examined here and the potential for web-based social network data is presented.

Specific groups of people utilizing an amalgamation of networks need to be studied in order to understand how virtual data networks and contemporary social networks link with items associated with physical presence such as transportation and cultural exchange. These individuals will also need to participate in a social network through distinct roles and come from a variety of backgrounds. With this in mind, the following work uses survey techniques to illustrate the influence of geography, biographies, and roles in how individuals involve themselves in a social network.

In this dissertation, the global CouchSurfing (CS) social network is used as a case study. Created in 2004, it now has more than 2.5 million members from over 200 different countries with more than 300 languages represented. This Internet-based social network provides hospitality exchange for travelers on any of the world's continents. Through the website members provide each other with a place to sleep (couch, floor, futon, bed, or other) while a reference system allows individuals to form their own levels of trust. This inherent amalgamation of social, transportation, and data networks makes CS an excellent source for studying how multiple networks merge together and influence human travel behavior.

In the modern, mobile society social networks have gained functionality and are easier to maintain with technology and the Internet. In particular, mobile devices with maps and GPS allow for people to travel quickly from their origin to their destination. While this may hold true for some people in all of their journeys, there is a growing number of people who desire more of a cultural travel experience and potentially a slower route between destinations (Priskin and Sprakel 2008). The CS network caters well to individuals who want to trek "off the beaten path" and see how the locals live, embedding themselves in a less typical and less touristic trip. This is out of the "comfort zone" for some but the thousands of people, ranging in age from 18 to 89, joining the CS social network every week in 2011 demonstrate a growing trend and desire for this novel type of travel. To many of these individuals a place is much more than physical characteristics and includes a key personal element. For some choosing this method of travel it is a matter of saving money but for other people that have sufficient funds for staying at private establishments such as hotels, there is a feeling of a more authentic travel experience that comes from interacting with the local population.

The work here also addresses the need for geographic visualization (geovisualization) of individual behavior that employs multiple networks. These techniques can be used with data from the past for historic purposes or for future planning purposes based on currently available status information. A conceptual model is needed in order to understand multi-network, hybrid (virtual and physical) spatial behavior and the various elements involved. Additionally, this will clarify the order and influence of distinct elements allowing for better travel planning.

1.2 Objectives

Specific objectives of this dissertation include:

1. A conceptual model of multi-network spatial behavior involving virtual and physical presence along with the choices, steps, and influences of each network involved.
2. Geographic visualization (geovisualization) methods for multi-network travel and related spatio-temporal constraints.
3. Geospatial data mining with the Web in order to add value and utility to discovered location information.
4. Examine the effects of demographic variables on how active a member is within a social network community.
5. Measures of social capital for individuals in social networks that have both physical and virtual location requirements.

1.3 Contributions

This dissertation contributes to the need for understanding travel behavior for long distance travel combined with the use of information from a social network. In order to conduct research on an amalgamation of social, data, and transport networks a variety of methods are necessary.

Within the following chapters progressive tourism is presented and a conceptual model is introduced. Time geography methods are employed in a social network setting, which brings about new constraints based upon one or more individuals and a multitude of networks. Geospatial data mining techniques are shown with the presentation of GEDMWA: Geospatial Exploratory Data Mining Web Agent. Data mining with social network data provides a useful method for collecting and analyzing a large repository of social network data.

A survey of over 200 participants provides insights as to how biographies affect travel behavior through the use of a social network. Count-based models are used to analyze these data and address hypotheses of how gender, age, and geography affect social network usage. Also examined is the notion of capital within a digital social network. This dissertation contributes measures explicitly taking into account geographical details in addition to the different roles an individual can play within a social network structure. These measures are based upon the results from a second survey of expert users who participate in the CS social network.

The audience for this dissertation is any individual with an interest in social networks, how humans travel and experience geography, and generally anyone interested in how technology affects humans and their behavior. Also covered in these chapters is how the topics of time geography, web crawling, data mining, and social capital play a role in activities that combine multiple networks in physical and virtual spaces. This dissertation aims to gain a better understanding of the challenges associated with these topics and shows how to conduct research in this area.

1.4 Dissertation Outline

This dissertation is outlined as follows: the next chapter presents the concept of progressive tourism with its associated methods. The chapter also contains background information on pertinent methods used throughout this thesis along with the conceptual model and sample scenarios. Chapter 3 discusses physical and virtual presence requirements in modern social networks, a geospatial data mining agent, and utilizes survey results to examine the effects of demographic variables on social network participation and activity. The fourth chapter pertains to measures of social capital and the particular influences of geography and an individual's role within a social network. The final chapter contains overall conclusions and opportunities for future work.

2. Progressive Tourism

2.1 Background

In this section the concept of a synergy of social, transportation, and data networks is divided into the following sub-questions:

- What behavior do travelers exhibit when utilizing a novel amalgamation of social, transportation, and data communication networks?
- How can these network levels be visualized efficiently to be utilized as a decision support tool for travelers and cutting-edge e-tourism systems?

These questions require individual in-depth research programs; nevertheless, for developing a more holistic understanding of this topic, the questions should be integrated. In order to tackle this work of geographic complexity the task begins at the conceptual level that first investigates each of the critical elements individually, and then analyzes their potential contribution to the system as a whole. Next, a data model will be created to test, explain, and represent different elements of the complete information. This model will be valuable for any researcher studying single or multi-level network architectures; and, on a broader scale, any individual experiencing geography first-hand, through tourism and travel.

Traveling can be a thrilling and stimulating experience; however, it can also be costly and emotionally draining. Presence in a new environment provokes questions, and demands answers to physical and geographical issues. This research focuses on the existence of hospitality and hosting combined with travel. When visiting a new place individuals are more likely to take on distant and exotic travels if there exists a connection with a local host. This can be beneficial for both the host and the visitor, as it can lead to exchange of culture and food, contributing to the development of companionship. Traveling for business as well as for leisure can fit into this category, and relate to these principles. Trips made in this fashion have been common for many years through family, friends, friends of friends, and beyond. However, expanding upon these concepts and using newly developed technological resources brings about a more recent form of travel, yet to be studied in detail.

2.1.1 Volunteered Geographic Information

Volunteered geographic information (VGI) is an area of research targeted towards enabling general audiences to author and submit information about their environment, complementing existing information sources and services with a user volunteered web of places. People should not only be enabled to access spatial information about their current location, but also to author and edit such data, and to interact with systems and friends that are physically separated. This encompasses a seamlessly integrated environment, where the real world is intertwined with the digital, and mobile devices serve as portals and handles to this digital world. The information flow is directed mainly from the user to a growing and distributed set of databases that integrate the volunteered information. The field of VGI has seen tremendous growth in the 21st century (Goodchild 2007). Previously, information has been volunteered by Internet users across the globe via structures such as Wikipedia. Now those elements are gaining a useful geographical element (Hecht & Raubal 2008).

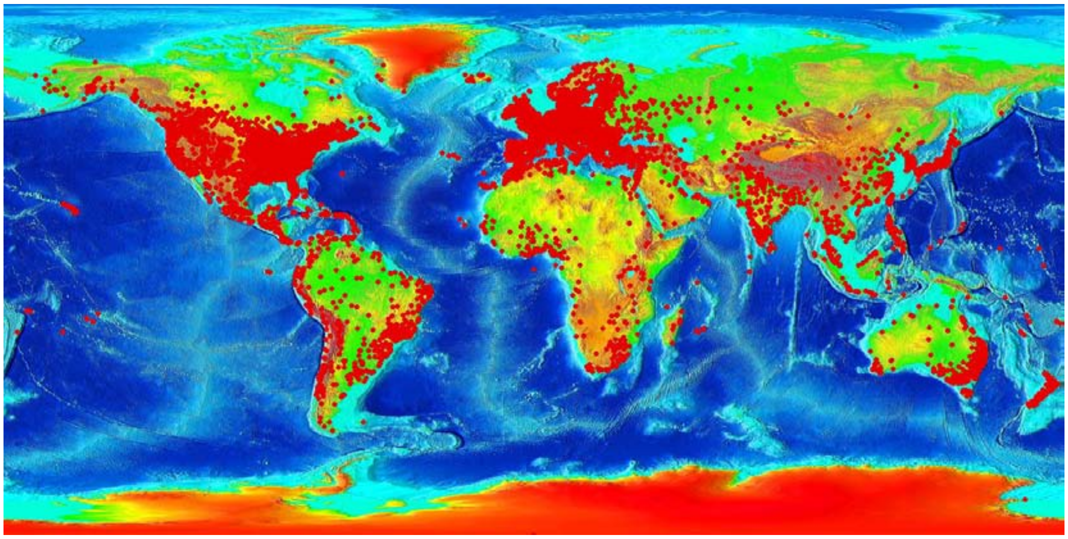
2.1.2 Social Networks and CouchSurfing

In order to embark upon this research a sound, existing collection of diverse individuals is needed to demonstrate the power of the network amalgamation: social, transportation, data and communication networks. This launches the effort to understand and characterize the unique spatio-temporal behavior of these travelers. In this research an existing online network is utilized as a case study. The CouchSurfing (CS, http://www.couchsurfing.org) project is a free, online social network non-profit consisting of people from all over the world who are willing to share time and their homes with travelers. Members allow guests to stay for free at their home on a couch, in an extra bedroom, or on the floor. Users may also offer guidance with tours or simply meet.

A guest first contacts a host via e-mail. The host resides in a desired location and a guest may also specify search criteria, such as age and gender, for potential hosts. There is additional information on each member's profile about the individual such as favorite activities, life philosophies, pictures, and references left by other members. Once guest and host meet in person they can share a variety of activities. After face-to-face meeting in a physical location the social network moves back to the virtual world as reviews are left for one another that will affect future usage of the service. Members often start using the network by either exclusively being a guest or hosting others but later engage in both as is necessary for the network to succeed (Lauterbach et al. 2009). While members initially interact through virtual messages, the core purpose of the network is an affordance of physical interaction, blending physical and virtual social networks.

A key element that holds all of these people together in one system is the use of a digital social network. Each member creates a profile containing VGI such as their current city of residence, previous travels, and future travel plans. A profile can also contain personal descriptions, hosting capabilities, languages spoken, and photographs. This CouchSurfing social network is increasingly important as the links created between nodes can be used to gather a level of safety for travelers as well as a global map with the geographic locations and places of connected friends. Members began signing up for the CS project in 2004 with hundreds joining per month. By 2005 thousands were joining each month and steady growth has continued over the years right up to 2011, by when over ten thousand people were joining during peak months. Now with over 2.8 million members (Figure 1) from all around the planet this social network continues to influence travel behavior around the earth.

Figure 1. Map of CouchSurfing members in 2011 from couchsurfing.org.



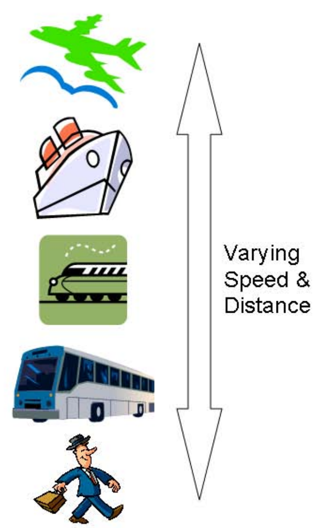
Users of CS now cover a vast portion of the globe with members on each continent. It is a contemporary, emerging form of connection. The world is quickly opening its doors more than some may suspect due to contemporary world news. Further expansion of this non-profit traveler system is expected to carry on well into the future; therefore any results discovered in this research afford a utility for prospective future developments.

2.1.3 Effects of Technology and Transportation on the Concept of Place

This research focuses on how the geographical spread of people in an Internet-based social network can influence the travel choices of each individual in the network. This innovative behavior revolves around the concept of place (Couclelis 1992) as the setting is continually changing on a wide variety of scales from city to country to continent. The role of place is unique in this system where people are searching for subjective places that are not merely space or location but highly personalized (Cresswell 1996). In this sense hosts can also point guests to places that are more than simple coordinates as the recommendations contain a degree of subjectivity. Hence place in this research is defined by all three network levels working together to provide travel opportunities, as well as link physical locations and subjective places. These places have emphases on any combination of a traveler's preferences for language, culture, guidance, and safety among others.

There is a wide variety of Information and Communication Technology (ICT) tools at a modern traveler's disposal, and these tools play a major role in their travel choices (Janelle 2004). Technology affords the participation of distributed individuals in the network to take part in the same project as is done with community modeling (Maechling et al. 2005), public-participation GIS (Keßler et al. 2005), or web-based communities (Pultar et al. 2008). Members of the CS system utilize ICT in combination with multi-modal transportation: different modes of transportation such as train, plane, bus, or boat. The use of ICT begins with communication through asynchronous email messages while guest and host are on separate continents. An email request can be viewed at a time convenient to the host, and does not interrupt their regular daily activities. Each member can evaluate the other using the online social network with biographies and references from other members. Further person-to-person communication can be carried out until both parties are satisfied. Later, when the members are on the same continent, more messages may be exchanged until both host and guest are in the same country or city. This is due to the fact that a guest may initially travel to a city in a country such as Berlin in Germany but have a trip planned a week later for Hamburg, Germany. While in Berlin the guest can call their future host in Hamburg to confirm and finalize trip details. After all of this, synchronous communication can occur via means such as a telephone to confirm meeting places, times, and personal descriptions. This process can be seen as taking place in a hierarchy of transportation scales starting at the highest level with cross-continent mobility granted through commercial planes or sea vessels. Below this are forms of travel such as national trains spanning across countries. Another scale lower are local trams and buses in cities until an individual reaches a final destination (Figure 2).

Figure 2. Various transit modes with different speed and distance capabilities.



This study concerning the synergy of multiple networks is built upon earlier research in travel behavior modeling, time geography, and visualization. Work in social activity-travel behavior points towards “the main individuals' driver to perform a trip is mostly with whom they interact rather than where they go.” (Carrasco et al. 2006, p.1) In other words the social aspects of whom a person interacts with in their travels can be a very important factor in determining where a person goes. However, this can be reversed in the CS system, as this electronic network allows people to select a location first and then decide with whom to interact. “A major part of social activity destinations are at homes of specific persons rather than at places that can be ‘chosen’ depending on attractors such as costs, environment, and proximity.” (Carrasco et al. 2006, p.2) This hints at the lack of choice available in many social activity destinations but this is different with the CS system because places can be chosen by CS members based on factors such as environment and location, in addition to other attributes of potential hosts.

For this dissertation individual space is the environment about which an individual can move and perform actions while activity space is the range of potential activities available for a person given their characteristics and attributes. Both individual and activity space are important in the CS system, while place plays a large role if a traveler has a strong desire to be at a specific place. This alters a guest’s preferences and threshold level in choosing a potential host. Place-based and individual-based decisions (Golledge & Stimson 1997) play a key role in this process and this work seeks to discover thresholds for people's transportation behavior. For example, a CouchSurfer primarily desiring cultural experiences can adjust her plans mid-trip and travel farther (perhaps at higher cost as well) in order to partake in a more desirable trip (e.g., free surfing lessons at a beach house on an island).

2.1.4 Spatial Behavior

Recent contributions combining social networks with travel have been described by Dugundji et al. (2008). They state that the contemporary changes in urban transportation show a need for policy integration and that the modern approaches to transportation analysis are pushing the capabilities of the underlying theories. Axhausen & Gärling (1992) have presented related work on conceptual frameworks of travel behavior. Golledge & Stimson (1997) provide a collection of spatial behavior knowledge and models that describe general spatial behavior. These models present a general method for investigating the steps and factors that contribute to travel behavior.

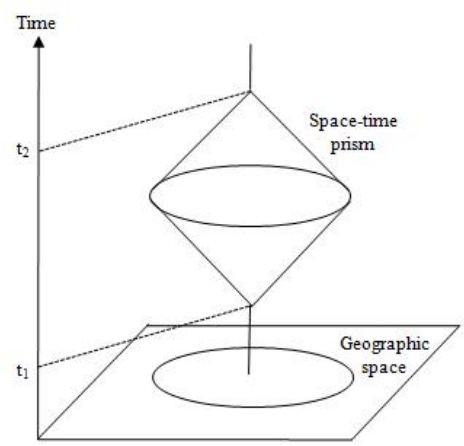
The STARCHILD conceptual framework (Root & Recker 1983) is an activity-based approach in which individuals are assumed to choose options with optimal utilities. The model characterizes travel behavior elements into pre-travel and travel stages. This classification is used in the conceptual model presented in this chapter, where an individual's pre-travel actions (as well as during, and post-travel actions) include extensive use of ICT in the form of data networks and digital social networks.

The SCHEDULER framework (Gärling et al. 1986) fits desired activities into a specified time interval. In scheduling, fitting the activities into the timeframe is a key objective of most travelers – apart from others, such as minimizing the travel time or distance. Nonetheless, some travelers may look for the maximum amount of travel for the available time, in order to see as many places as possible, and experience the maximum number of diverse events. Explicit influence of ICT systems on travel behaviors is a modern phenomenon, and is therefore covered in this research.

2.1.5 Time Geography

A pertinent field of research related to this dissertation is time geography. People and resources are available at a limited number of locations for a limited amount of time. Time geography defines the space-time mechanics of location presence by considering different constraints (Hägerstrand 1970). The possibility of being present at a specific location and time is determined by a person's ability to trade time for space, supported by transportation and communication services. Space-time paths depict the movement of individuals in space over time. Such paths are available at various spatial and temporal granularities, and can be represented through different dimensions. All space-time paths must lie within space-time prisms (STP). These are geometrical constructs of two intersecting cones (Lenntorp 1976). Their boundaries limit the possible locations a path can take (Figure 3). The time budget is defined by $\Delta_t = t_2 - t_1$ in which an agent can move away from the origin, limited only by the maximum travel velocity. From a network perspective, such as taken in this research, movement is limited by the network geometry and the maximum travel velocity, which can vary for different edges and times. The geometry of the STP in a network therefore forms an irregular shape. Algorithms for calculating the network time prism (NTP) can be found in (Miller 1991) and (Raubal et al. 2007).

Figure 3. Space-time prism as intersecting cones.



Time geography defines different constraints that limit a person's activities in space and time. Fundamental physical restrictions on abilities and resources are summarized as capability constraints. Not having access to a car in order to trade time for space efficiently is one example for this type of constraint. Coupling constraints refer to the requirement for a person to be at a specific location at a certain time or for a fixed time duration. For example, if two persons want to meet at a physical location, they have to be there at the same time. Certain domains in life are controlled by authority constraints: a person can only shop at a mall when the mall is open, such as between 9am and 8pm.

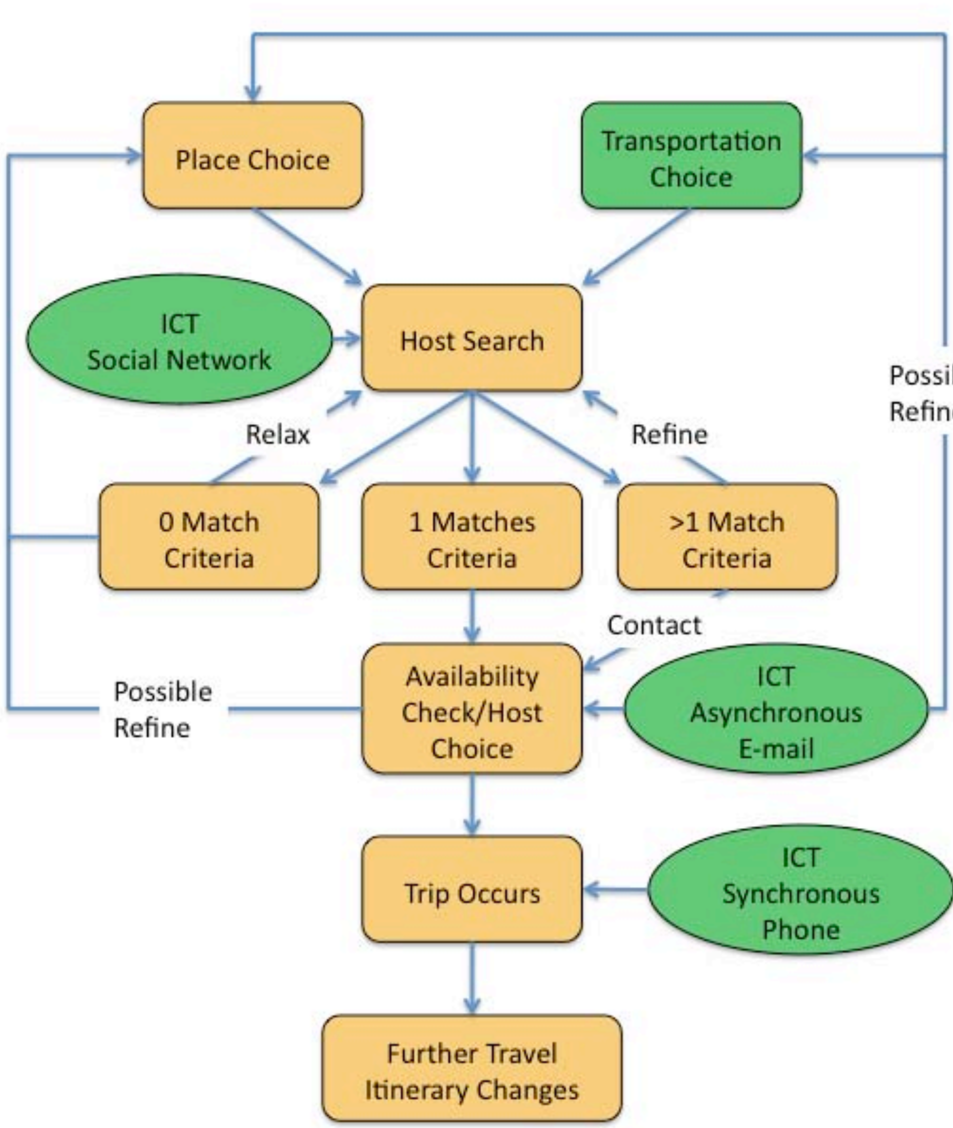
Time geography has been extended to the area of geographic information systems (GIS) regarding transportation networks to model and measure space-time accessibility (Miller, 1999; Wu & Miller, 2001), and for the analysis and theoretical understanding of disaggregate human spatial behavior (Kwan, 2000). It has also been advocated to integrate time geography with both GIS and location-based services (LBS) to achieve more user-centered systems (Miller 2005b; Raubal, Miller, & Bridwell 2004). Further applications in the geographic domain concern the structuring of dynamic wayfinding environments (Hendricks, Egenhofer, & Hornsby 2003) and the modeling of geospatial lifelines (Hariharan & Hornsby 2000). Analytical formulations of basic entities and relationships from time geography can be found in (Miller 2005a).

Visualization methodologies based on time geography can be utilized to display and analyze the integration of social, data, and transportation networks presented here. Several methods for computing and geographically visualizing human activity patterns based on time geography have been proposed and successfully employed (Kwan 2004; Ren & Kwan 2007). Space-time paths allow for depicting both the movement of individual travelers in space over time, and their utilization of the social and communication networks. Yu & Shaw (2008) have presented an adjustment of the STP and utilized a 3D-GIS representation to support visualization and analysis of human activity patterns in physical and virtual spaces. The paths are available at various spatial and temporal granularities, and can be represented through different dimensions. It is important to note that this behavior requires attention to the concept of scale with respect to both time and space (Montello, 2001).

2.2 Conceptual Model of Travel Behavior in a Cost-Free Lodging Network

Figure 4 shows a conceptual model of travel behavior, it highlights the critical elements and factors that affect a user's decisions. This flowchart helps in visualizing each step, in addition to presenting how the whole process fits together. Travel choices are made in a variety of locations and are heavily affected by the synergy of ICT tools, international transportation networks, and web-based social networks. In the absence of any one of these crucial networks the traveler’s behavior would be greatly altered. Therefore, this type of travel is quite unique and highly dynamic, and does not lend itself for analysis by traditional travel behavior models. The application of this system is further described in the rest of this section.

Figure 4. Conceptual model of dynamic spatial behavior utilizing three network levels. Green elements are explicit networks utilized in the framework: data, social, and transport networks.



Initially a user chooses a place to travel. The individual's choice may be for any purpose including vacation, cultural, and/or business. The concept of scale is important with the initial place decision as users may vary in their desire to visit a specific city, country, or continent. The results and further refinement of their decision behavior will move through various scales (both virtual and physical) before reaching street-level with corporeal presence between guest and host. At this point an initial transportation mode choice may be made or hypothesized by a user but further potential refinement is necessary in this highly dynamic model.

Utilizing any level of initial place and transportation choice the traveler will perform a host search based on their individual criteria. At this step user-based trip preferences are gathered including desired place along with any combination of language, age, verification level, and gender among others. Following criteria specification, possible hosts are generated that match the desired characteristics. This may return no results and require the user to relax their criteria and search again until at least one host match is found. If one host matches the specified criteria the user can choose that host and move to the next step of checking availability. If more than one host matches the criteria a user may further refine her search to narrow down the results or may choose to contact multiple potential hosts. Also with more than one potential host generated the user has the ability for further evaluation (before contacting via e-mail) by utilizing characteristics of a modern Internet-based social network. These include profile information, number of 'friend'-links, personal references left by other members, and photos. The weights a person applies to each of the criteria vary widely between users. For instance, one traveler may place more weight on staying with a host of a specific gender versus another that may be more concerned with finding a host within a particular age interval.

Once a traveler has chosen to contact a possible host about availability then the asynchronous ICT tool of e-mail is utilized for initial communication. At this step a guest contacts a host that is local to the desired area of travel. Here the highly dynamic nature of this travel behavior is demonstrated as advice may be given that alters a user's transportation mode choice and potentially their place choices at various resolutions. Next, availability is verified and if both parties are satisfied the trip occurs. At this step the model maintains its flexibility as users may check e-mail at an airport and choose different standby flights based upon the current situation potentially leading to a new place choice or other changes in itinerary. This also relates to the mid-trip itinerary changes described in the next section with sample travel scenarios.

Typically, the travel network infrastructures work as an enabler allowing mobility options with planes, trains, and buses. It is important in this system to also consider the potential barriers of the transportation network with employee strikes or a natural disaster event. With this system a user can quickly change her place choices and find a new host where there are no transit barriers. Upon arrival at a desired destination travelers may find synchronous ICT technologies such as telephones useful to verify initial meeting locations and acquire personal descriptions. Guest and host meet face-to-face in physical space at this later step making the transition from virtual to corporeal presence. After an individual completes her stay the framework allows for further travel along with itinerary changes. For example, hosts may have strong recommendations of places to see and/or connections with other potential hosts the guest had not originally planned. Here again we see the appropriateness of the highly dynamic and flexible model provided for this travel behavior. The individual may continue travel to any number of places for any time period starting at the top of the framework with a new place and transportation choice. The conceptual framework presented in Section 3.1 and Figure 4 provides a starting point for how this behavior is modeled and raises further research questions such as those of traveler safety and model evaluation. Additional real world accounts such as those given in the next section entitled "Sample Scenarios" will further justify the creation of this new conceptual model. In conclusion, this model serves as a basis for further inputs, datasets, and adjustments into the future.

2.3 Sample Scenarios

This section provides travel scenarios using the CouchSurfing system to further demonstrate the necessity for a novel conceptual travel behavior model. Two types of travelers are presented: the first has little to no constraints and the second has many constraints. Different levels of constraints are possible and further apparent as a result of the scenario descriptions, which exemplify the capabilities of this highly dynamic, multi-network system.

Sample North American traveler A desires a trip to Europe to experience different cultures. Using ICT and the dynamic system described here the individual is able to purchase minimal tickets, use a flexible itinerary, and minimize lodging fees. At the start of the trip the individual plans on spending one week in the Netherlands but after three days decides to visit the nearby city of Antwerp in Belgium. A suitable CS host is found the day before who welcomes the arrival, leading to unique food, music, and history knowledge. The traveler discovers discount airline tickets from a previous host and books an inexpensive flight to Brno, Czech Republic. Leading up to the departure of the flight multiple potential hosts have been contacted but final plans are not solidified. Checking e-mail at the airport less than an hour before departure instantaneously reveals solidified plans of what people and places will be visited. Castles, casemates (chambers in fortresses), and college campuses are all part of the tour given by a local but they are only able to host for two days. Traveler A examines a map and notices Vienna, Austria, is nearby. The tourist only has time to record some phone numbers of potential hosts to call once the train arrives in Vienna. The individual calls members of the modern social network and finds a place for the evening and from there is able to call others to stay with for the next nights. The hosts provide museum recommendations, sack lunches, and travel tips for the area increasing the traveler's mobility, efficiency, and overall cultural experience. Traveler A moves north and stops in a small Czech town. She wanders into a local festival and is seen by a friend of a previous host who provides accommodation. After this a member of the network provides accompaniment and knowledge of the buses and trains on an excursion to a national park for a few days. At the end of traveler A's tour a host in Poland gives a knowledgeable tour as the host is training to become a tour guide. After a thorough journey around town the tourist returns home with a multitude of unique experiences in diverse places.

Travel via the CS network can be done for a variety of purposes such as business, leisure, study, culture, and language. The following is an overview travel description of a CS trip for primarily business but provides other purposes as well. Sample South American professor (traveler B) wishes to attend a conference in Barcelona, Spain, where he will give a presentation on a paper he wrote. Since he has never been to this location before, the individual decides to spend some days before and after the conference in Barcelona to gain more experience of the new culture. The traveler discovers a CS host of the same age who has lived over half of his life in Barcelona. The guest and host send each other a series of e-mails to arrive at an initial plan of activities that is convenient for both parties. The host then takes the guest to local museums and eateries not in traveler B's travel guide, but he is pleasantly surprised. After the trip occurs both parties are satisfied with the exchange of culture and experiences and hope to reverse the roles of guest and host in the near future.

2.4 Progressive Tourism and Time Geography

The research presented here utilizes VGI as a key component for collecting geodata that are visualized as 2-dimensional static spatial maps in a GIS. With the existence and necessity of this multi-network architecture demonstrated, this section focuses on how all of this relates to the field of time geography. Table 1 demonstrates sample concepts from time geography and how they relate to this work.

Table 1. Time geography concepts linked with multi-network travel behavior.

Time Geography Concepts	**Multi-Network Traveler Examples**
Authority constraints	A host provides a guest with information about operating hours of sites in addition to a sequence in which to visit.
Capability constraints	Host shares local and international transportation capabilities as well as the current status.
Coupling constraints	Initial corporeal meeting between guest and host; also synchronicity between a traveler and any mode of transportation.
Fixed vs. flexible activities	Traveler wants to mountain climb meaning the possible locations and schedule are flexible vs. a traveler designating a fixed activity such as

	visiting the Louvre.
Potential path areas and prisms	These are created using a traveler's desired travel destinations in a specified duration of time along with available modes of transportation.
Space-time stations	Various tourist activities such as a historic church or a host's residence are space-time stations.

This investigation provides a unique application of user-centered time geography (Raubal et al. 2004) as the travelers have a wide array of journey preferences such as length of stay, gender of host, accessibility to public transport, and host's age among other factors.

Once data has been collected, the information can be further analyzed and visualized via static 2-dimensional individual network maps (Larsen et al. 2006b). However, travel is a dynamic activity varying in both space and time, leading to a necessity for more advanced methods of visualization made possible by a recent boom in geospatial technology. For maximum cross-platform utility, a visual portion of these research results is exhibited by means of 3-dimensional space-time maps in a virtual globe environment such as Google Earth (http://earth.google.com/). The extensive information repository of world data currently available in this software makes it an ideal platform to demonstrate and visualize results. This offers the possibility for interactive visualizations of space-time maps with tools affording the ability to control spatial and temporal variables. Of specific importance in this context is the Keyhole Markup Language[8] (KML), an eXtensible Markup Language (XML) geospatial data file format accepted by the internationally recognized Open Geospatial Consortium[9] (OGC). This information representation provides a highly flexible and efficient technique for working with modern spatio-temporal data. Once representations have been created, conversion between various geospatial data types can be automated with the Geospatial Data Abstraction Library[10] (GDAL). Exploring time geography in this novel way supplies useful visualization and analysis tools for geographers and spatiotemporal researchers as a whole.

The visualizations will also be useful for travelers planning a trip. These e-tourism tools will further allow users to see potential paths and therefore their many travel options given their individualized constraints of space and time. The results of these tools can also aid further analysis via utilization of spatio-temporal queries available in a dynamic Geographic Information System (GIS; Pultar et al. 2010). Another key feature of this approach to time geography is the wide accessibility to persons of all socio-economic status due to the no-cost software platform of Google Earth[11]. As an experiment of capabilities and proof-of-concept, the sample physical travel scenario described in the previous section is visualized in Figures 4 and 5. In this example each location has a temporal attribute that corresponds with a time

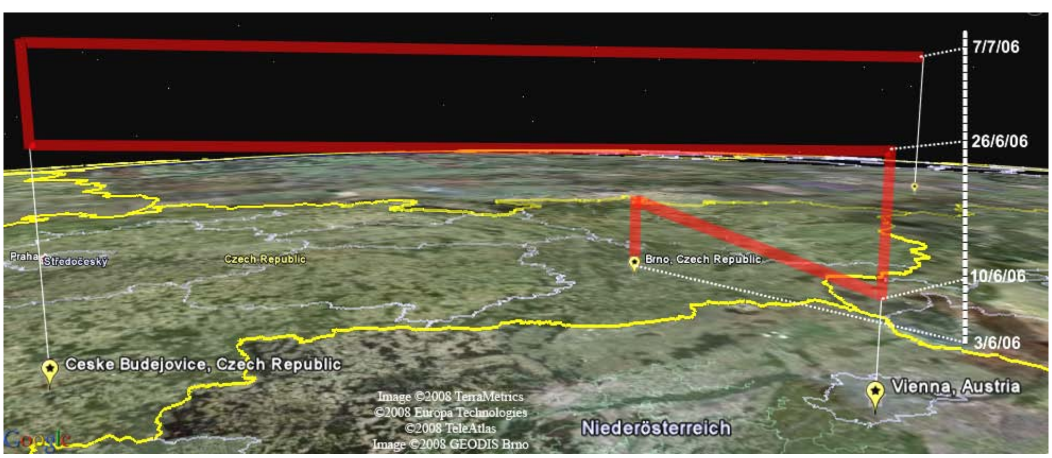
---

[8] http://www.opengeospatial.org/standards/kml
[9] http://www.opengeospatial.org/
[10] http://www.gdal.org/
[11] http://earth.google.com

slider tool in addition to the visible height variable. Utilizing an intuitive user interface, any traveler will be able to input their preferences, constraints, and capabilities to acquire a 3-dimensional time-geographic representation of their possibilities. For example, the figures also demonstrate a potential solution for a journey involving 6 stops. Given a starting and ending location a traveler recommender system (TRS) (Ricci 2002) could return this option for travel order based on a user's preferences. The traveler has a space-time prism with height boundaries decided by the traveler's available time. The lowest height represents the beginning of the trip and is bounded above by the ending time of the trip. With respect to the lodging system in this research, different optional sequences may be shown based on host availability. In taking a trip this visualization methodology aids a traveler in making choices and changes in a travel itinerary. These tools exemplify ways of how we see this research contributing to the development of future cutting-edge e-tourism systems (such as a TRS) and also providing high-utility tools for spatio-temporal researchers (Sharda 2009).

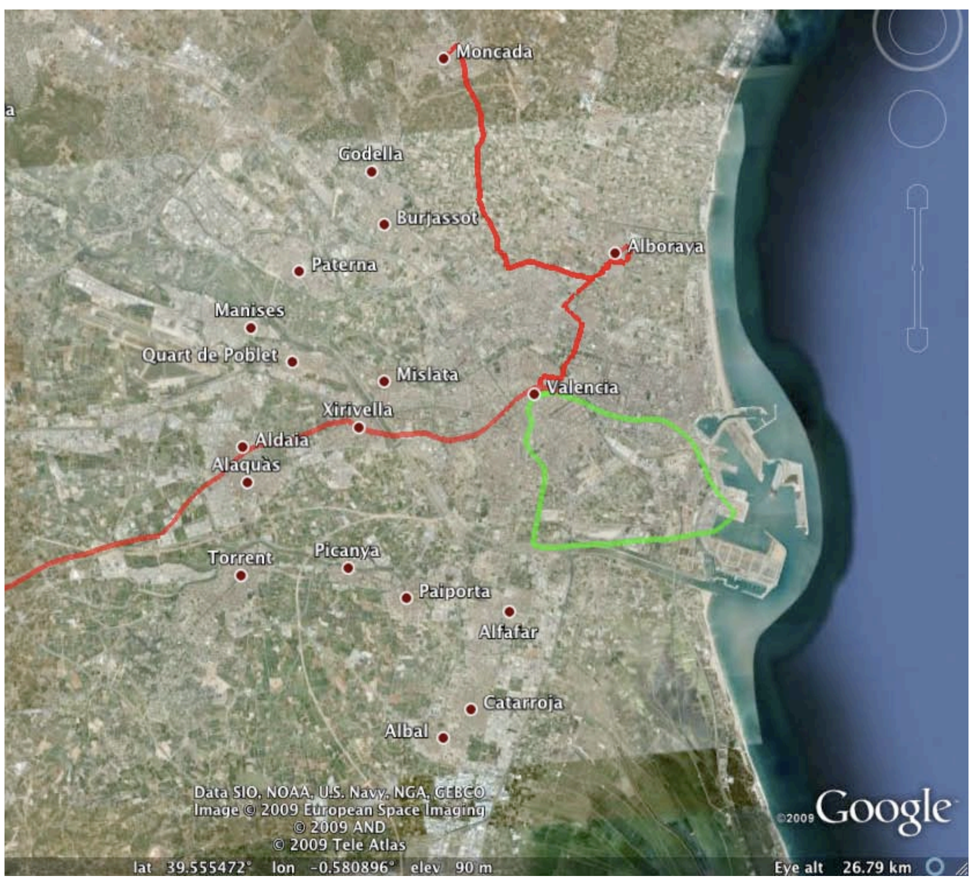
Figure 5. Space-time path of example traveler scenario.

Figure 6. Tilted view of a space-time path showing temporal visualization with height.



Furthermore, the methods of time geography applied to social networks can be utilized for more than one individual at a time. In Figure 7 space-time paths are shown for an example CouchSurfer guest and host. In this scenario the guest begins in Moncada, Spain, and travels to Alboraya, Spain, on 13 July, 2009. The traveler spends a day exploring Alboraya before meeting their CS host in Valencia. The host has stayed in the same location (for this geographic scale) doing work until the guest arrives. Once the guest arrives the two independent space-time paths merge as the two CS members eat meals and sleep in the house. The next day the host takes the guest on a tour around town and to the pier and back. After this the two individuals go their separate ways with the host going back to work and the traveler continuing to the north. Further contact is maintained in a virtual manner until the hospitality is reciprocated or a repeat stay occurs.

Figure 7. Space-time paths of host (blue), guest (red), and both host and guest (green).



This time geography approach can quickly visualize the overall travel behavior (location and temporality) of the social network members described here. Additionally indicated here is the switching of the spatial relationship type (virtual to physical) between the two individuals by their space-time paths merging into the green path around Valencia. This example also demonstrates the utility of the framework's constraints for these types of scenarios. The capability constraints affect the velocity and route a traveler takes. The authority constraints determine when an individual is able to dine out or when movement by bus or train is possible. Coupling constraints requiring people to be at a specific location at the same time are key for the social network members to move from virtual to physical space and have a face-to-face meeting.

This chapter has addressed the two initial objectives from Section 1.2 pertaining to a conceptual model and geovisualization methods for multi-network spatial behavior. Introduced here were the constraints relevant for an individual using the CouchSurfing network and visualization methods that are applicable to the rest of this dissertation as well as other multi-network spatial behavior requiring both physical and virtual presence. More work with additional constraints from all of these networks as well as explicitly integrating and visualizing all of the network levels is left open for future work.

3. Types of Presence and Data Sources for a Location-Based Social Network

3.1 Background

The Internet and social networks are bringing individuals around the globe closer together in a virtual world. Their popularity has increased in the past few years with networks such as Facebook[12] and Twitter[13] allowing users to interact with each other after social links are made (Wilson et al. 2009). These user interactions (messages, comments) are often digital and involve no corporeal or physical presence. Hence interaction among individuals in social networks has quickly become more virtual than physical and explicit location or presence is required in less scenarios. Sale transactions, photographs, videos, books, and more are exchanged via the Internet without the need for face-to-face communication. These contemporary social networks are based in a virtual world where biographies and pictures need to be maintained with a computer. However, physical location and space do still matter and are key elements in today's Location Based Social Networks (LBSN). The necessity of accounting for physical location in modern, Internet-based social networks is addressed in this work. Utilized here is a web-based social network for travelers called CouchSurfing (CS) as a case study. See Section 2.1.2 above for a

---

[12] http://www.facebook.com
[13] http://www.twitter.com

review of the CS network. Also presented are the different spatial and temporal components required for the functionality of such a network.

In this LBSN physical location is key as some larger cities are more popular places to travel. This may be due to accessibility with airports, large train stations, cultural activities, or other pursuits that affect the amount of tourism at a destination. Travelers may also crave a more authentic trip and thus stay with hosts in remote locations that do not experience much tourism.

The continuous growth of online social networks has instigated many research projects by people around the world. Social networks allow a user to create an abstract version of an individual that is presentable to the world. Online identities vary in different degrees when compared to who the person is and depending on what they wish to portray or gain from an online presence. This and other social network research has led to special sections of journals (Ellison and Boyd 2007) and conferences (IEEE Social Computing Conference). An amalgamation of skills in varying disciplines is helpful in studying social networks. This includes, but is not limited to: sociology, psychology, communication, computer science, geography, and mathematics. The scope of digital social networks is large and different disciplines are needed to contribute to social network studies concerning language and friendship (Herring et al. 2007), ethnicity (Gajjala 2007), religion (Nyland and Near 2007), and gender (Hjorth and Kim 2005).

Traveling purely in cyberspace (Seidel et al. 2009) is possible too. Hence a travel destination can be presented through the eyes of the programmer or virtual world creator. This provides an initial view of the environment in digital form. It is also an introductory look at the spatial layout and topology of a given place. Incorporating cultural value is difficult in the CS network, especially with travelers who have large amounts of social and cultural capital. While cultural value in the CS network may be a future research topic, Molz (2007) presents work on mobile hospitality with CS as one of the examples and Pultar and Raubal (2009) describe CS as progressive, transformative tourism. Additionally, Lauterbach et al. (2009) as well as Bialski and Batorski (2007) examine trust and reputation within the CS network.

3.2 Measures of Capital

Through a LBSN travelers collect and share spatio-temporal information or volunteered geographic information (VGI, Goodchild 2007). See Section 2.1.1 above for a review of VGI. Specifically in CS, references are left for people, places, and activities that are publicly viewable by other network members. For example, after a stay a guest leaves a reference such as:

"I stayed with X for 5 days. I planned on staying 3 days but the collection of temples and parks in his city is incredible. The town had a relaxed atmosphere with plenty of activities (such as hiking and ziplining) to do during the day and X was able to take me out for great Asian food at night."

Here a positive reference is left for the host, increasing the host's network capital and encouraging others in the network to either host X or be hosted by X.

Recommendations are also made for the location where this LBSN member lives. Last, there are suggestions of activities applicable in this space. Given the dynamic nature of data on the Internet, this review may be written one month before another traveler considers a trip to this place. Therefore this service provides a more up-to-date source of location information versus a travel guide that went to press one or more years previous. In this way travelers seeking a primary goal of practicing a language, e.g. Spanish in Argentina, can rank possible destination cities within a country based on the new people, places, and activities possible.

Activities such as skiing, foreign language conversation, and swimming necessitate real-world experiences that spawn from digital social networks. Hence to grow one's network capital social communication skills are applicable in both virtual and physical settings. Transit capital is also needed to facilitate mobility between locations. In order to discover and experience new places a LBSN member has to be able to transport oneself. This can happen by any number of means such as cars, airplanes, and buses but without this critical physical component the LBSN is non-functional.

In this network the overall capital measure can be broken down to the following components:

- Network capital – references of people, places, activities
- Transit capital – mobility
- Social capital – geographical distribution and quantity of existing connections within the LBSN (see Chapter 4 for more information)

Combining these measures affords a member differing levels of abilities based on their adeptness and resources. Network capital is described here as the information made available to members of the LBSN regardless of their location. This includes locations and activities that are highly recommended by other members. This information is not widely available in other sources such as books and magazines therefore giving an advantage to members of the network. Transit capital is quantifiable in terms of available money (even if it is minimal to none in the case of a hitchhiker) whereas social capital is based on the number and spatial distribution of social connections a person has and the strength of those connections (see Chapter 4 for more information). Communication skills can be taught and learned through experience and while they are needed for interaction they are more complicated to quantify than train costs or amount of positive references.

In the CouchSurfing LBSN there are three main components to gaining capital within the network.

- Initial search for CouchSurfers and e-mail communication [Virtual]
- Corporeal meeting, communication, lodging [Physical]
- After meeting a negative, neutral, or positive reference is left and a friendship link may be created [Virtual]

Location plays a role in all of these steps, but is most restrictive in the second step containing the face-to-face meeting. Search and initial communication can be performed at a network member's convenience from any device with access to the web. Similarly, the final step can be completed at a user's home or on the road.

3.3 Mobile Devices and Mining

Mobile devices can be used more and more in the CouchSurfing process as the LBSN can be initially accessed using a web browser on a mobile phone. User preferences for discovering new people, places, and activities can be combined with transportation mode choices to make recommendations (Pultar et al. 2009). Text messages and e-mails can be exchanged closer to the travel date and eventually voice calls. All of this is possible on one mobile device. Even further, mobile phone programs such as Latitude[14] and Loopt[15] provide real-time location information that can ease the process of synchronizing space and time between members of the LBSN. This can be especially helpful when a guest is in an unfamiliar environment. Using these services one can see each individual's location on a virtual, interactive map. These types of Location-Based Services (LBS) aid in making the transition from virtual to physical presence that is necessary to gain capital in the LBSN described in this thesis.

Another use for LBSNs such as CouchSurfing is to use mobile devices with location-based technology to report nearby network members that are interesting. At the basic level this includes friends and those with similar interests as is done with the programs mentioned above. However, suggestions can be made for friends of friends where two people in the same location can meet based on their pre-existing tie with a middle connection. This gives a common ground and trustworthiness

---

[14] http://www.google.com/latitude
[15] http://www.loopt.com

factor as both members know and trust the first-degree friend that links them. With pre-existing social networks where members have built up their connections this can be an effective way to build their network capital. Given the basis that CouchSurfing connects travelers around the world this is a useful, global application. Generally, the members of this LBSN are highly mobile travelers making it quite likely to meet a second-degree friend, or friend of a friend, in a passive manner through the common, shared friend.

The interface is similar to modern mapping applications on mobile devices where live location feeds of existing friend connections are broadcasted (Figure 8). In addition to existing friends new suggestions are given, based on link strength, of currently nearby individuals that share at least one common friend node. This provides an opportunity to meet face-to-face and establish a new, quality link.

Figure 8. Screenshot from a mobile device where green points are locations of suggested 2nd degree friends or friends of friends with information on the common friend link. The black point is the user's current location and red points are locations of 1st degree or directly connected friends.

Users of the CS network can search for a person that shares similar interests, e.g., juggling. Additionally, activities that a member considers interesting, such as paragliding, can be found by looking at other members' interests and by scraping text from references left for the user. Here is an example reference:

"X hosted me for 5 days and took me out hiking the second day. I also went out hiking my last two days since it was such an accessible activity. Enjoy your trip to China next month."

This suggests the availability of hiking near this host's residence, which may be a smaller town not known for this activity.

By restricting a search of the network by location (e.g., city and country) we can then use data mining to discover interesting activities. Similarly, we can restrict the search by activity and find locations where they are possible. Given an initial list of interesting activities (e.g., scuba, paragliding, climbing, swimming) the references left for network members in a specific location can be acquired through spatial data mining in order to get an idea of how frequently people engage in the activity at this location. Initially a raw number of references mentioning an activity keyword can be gathered. If desired the number may later be normalized by how many members currently reside in a location to get some idea of the density of the activity in this place.

This is a unique form of spatial data mining and knowledge discovery as it utilizes social network accounts involving both locals and travelers to describe geo-tagged locations around the world. In the future scripts and small programs can be written to extract opinions and reviews of spatial information from social networks as described here. These tools can be used to discover knowledge and as a starting point employing similar methods to those used by geospatial agents for general websites, see Pultar et al. (2008) and the presentation of GEDMWA in the next section.

3.4 GEDMWA: Geospatial Exploratory Data Mining Web Agent

3.4.1 Introduction

The amount of geospatial information available through the Internet is ever-increasing but discovering the data and converting it to a useful format is a challenge. This is because of the unique combination of techniques and the necessary communications between the components. An increasing number of users are volunteering geographic information via means such as social networks, Google Earth[16], and Wikimapia[17]. However, as anything published on the web is available to the public it is a repository of VGI.

Internet users often provide location information, e.g., latitude / longitude coordinates for a video, a picture, travel destination, or other social data. In a Location-Based Social Network (LBSN) such as CouchSurfing (CS) each individual provides his or her place of residence, locations traveled in the past, locations to travel in the future, and more information about themselves and their connections in the social network. These geographic facts come in a variety of forms, e.g., coordinates may be in other forms such as state plane and place names can be encountered such as the name of a country or region. Hence in this context a geographic fact is a location in the form of coordinates or a place name. Within a social network setting, data mining individual profiles to acquire the frequencies of

---

[16] http://earth.google.com
[17] http://www.wikimapia.org

place names yields information about the strength of ties between places. For instance, in examining a quantity of CS profiles from Australia, place names such as England and the United States may occur most frequently, suggesting a stronger relationship and travel exchange between these places than others. Also with social data mining methods the frequency of words in profiles concerning activities, such as surfing in the case of California, provide additional information about a place.

The rest of Section 3.4 discusses in more detail the search for geographic facts in the form of latitude / longitude coordinates on the web. Coordinates in this form might appear in the text of travel websites or personal blogs. For example, an individual may write about a waterfall or a rare bird they saw along with coordinate information of where they where that comes from a GPS or other mobile device. In this way one captures information about a place described by a human who has been to the location. By using open source software solutions encountered geoinformation is stored, analyzed, queried, and visualized as the agent creates a data repository of what it discovers.

This section focuses on the development, prototypical implementation, and testing of such a geospatial agent: GEDMWA (Geospatial Exploratory Data Mining Web Agent) reads webpage data, follows links, and acquires spatial information for use in a GIS. Data exploration using this agent can discover previously unknown properties of a phenomenon and add value to geographic locations through explicit location-based semantic data mining of the Web. The resulting information is visualized and analyzed after GEDMWA converts it into proper GIS and virtual globe (Grossner et al. 2008) formats such as the Keyhole Markup Language (KML). A wide variety of user communities can then use a diversity of international sources to discover additional knowledge about the field of interest.

The geospatial agent here functions using VGI put on the Web by the general public. This creates a large dataset with many potential data providers across the globe such as social network members and blog writers. It also brings into question who puts geographic data on the web and what is the associated quality of data (Shi et al. 2002).

The goals of this research include:

1. Create a geospatial agent to locate VGI on the Web.

2. Produce valid geospatial data (KML, shapefiles, others) from these discovered data using open source software solutions.

3.4.2 Use Cases

GEDMWA has utility for various user communities around the globe. Take for example a group of mountain climbers. A blunt way of using geospatial data to find new areas to climb would be examining topographic maps and looking for suitable changes in elevation. However, this may not always yield desirable results, as blank mountain faces with no holds can be impossible to climb. Members of this community post information on the web with geographic coordinates and descriptions of climbing routes. With the use of a GPS receiver the general public can collect location data in the form of latitude and longitude among others.

Gathering this VGI using Geographic Information Retrieval (GIR) is possible using the geospatial agent developed in this research.

The GEDMWA can be initially directed to pages on the web that contain the phrase “rock climbing”. A user can then specify a number of pages to consider in the creation of this spatial database in order to limit the search. The Java™-based agent searches for patterns of coordinates in the pages and creates the associated geospatial data files. All of these georeferences can be compiled into one source and easily imported into a virtual globe environment or open source GIS, such as Quantum GIS[18], uDig[19], or gvSIG[20].

Assembling these data together adds a utility value to the discovered spatial locations for each member of the community. Additional use cases are found in a variety of communities such as those concerned with geysers, historical locations, landmarks, or animal watching (Larson and Craig 2006). GEDMWA is able to locate and describe what exists in the physical world by using the virtual environment of the Internet. This aids in the decision-making capabilities of users as they gain a better knowledge about locations of interest.

3.4.3 Related Work

An agent can be regarded as anything that perceives its environment through sensors and acts upon that environment through effectors (Russell and Norvig 2002).

---

[18] http://www.qgis.org
[19] http://udig.refractions.net
[20] http://www.gvsig.org

More specifically, agents are considered computer systems that are situated in some environment and can act autonomously (Woolridge 1999). Agents have been mainly dealt with in Artificial Intelligence but have recently also gained popularity in other fields such as geography (Batty et al. 2003, Benenson and Torrens 2004). Sengupta et al. (2008) provides an extensive review of the current state of these geospatial agents.

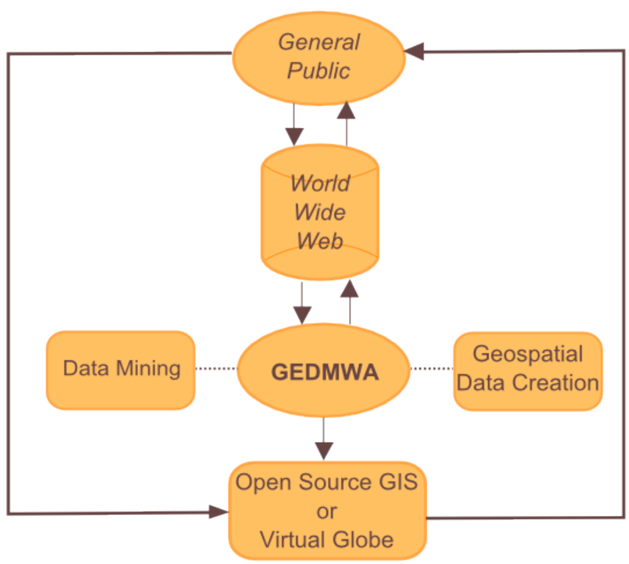
VGI harnesses the power of having geoinformation authors across the globe. The Internet aids in sharing this information publicly. The growth in VGI has quickly increased in recent years and continues into the future as an element of Web 2.0 (Goodchild 2007). The system presented in this section utilizes VGI as a key component for creating geodata that are visualized in a GIS or virtual globe environment. The web as a source of VGI has been realized for modeling vague places (Jones et al. 2008) in the UK as well. It is an excellent example of the potential the web has for GIScience as it generates boundaries and surfaces of regions based on a place name. For a further review of VGI see Section 2.1.1 of this dissertation.

Turning web data into usable information is a task that can utilize methods such as simple data retrieval or knowledge discovery. Data mining (Han and Kamber 2006) and its tools are critical parts of the geospatial agent. Geographic knowledge discovering and spatial data mining are tailored for extracting information and finding interesting patterns in databases with a spatial component. Miller (2007) describes this in further detail along with other key techniques. Specifically searching for information through data mining the web is explained in detail by Chang et al. (2001).

3.4.4 Methods

GEDMWA utilizes a number of tools and steps to complete its tasks. This includes VGI, the Internet, data mining and creation, and GIS. The agent's overall flow of data, methods, and tools is visualized in Figure 9 and described in the rest of this section.

Figure 9. GEDMWA Data Inputs and Outputs



Initially GEDMWA crawls the web from a starting URL searching for public geospatial data. This could be a topic search and the page returned by a web search engine query. It may be any valid URL and, in the interest of discovering geospatial data, one that is likely to have coordinates. The agent downloads the content of the page using Java™ methods for URLs:

```
URL searchURL = new URL("http://wikipedia.org");
URLConnection urlConn = url.openConnection();
InputStream urlStream = url.openStream();
```

After the connection is opened a read method is used to acquire the entire page text. Valid links are saved for further investigation and later followed recursively (abiding by the Robots Exclusion Standard[21]) until they reach the user-specified limit. A basic text database is used for testing the prototypical implementation where each URL's text is stored for later geographic data mining purposes. Future versions can support SQL queries utilizing more rigorous, open source database software such as HSQLDB[22].

Once the agent has collected the desired amount of pages it can begin the dissection and data mining process. The key part of this process involves the search for geographic coordinates. The Java™ programming language provides the powerful pattern-matching tool of regular expressions, which can be utilized to locate geographic coordinates within each page. Regular expressions allow for automatic retrieval of text strings containing sequences such as “dd.ddddddd” where

---

[21] http://www.robotstxt.org
[22] http://www.hsqldb.org

each d corresponds to a digit: 0 thru 9. E.g., the following code would find latitude decimal degrees coordinates in a form similar to “N 46.853441”:

```
[NS]\\s\\d{1,2}\\.\\d{3,10}
```

This searches text for an initial N or S (North or South) character, then a space before 1 or 2 digits followed by a dot (.) ending with 3 to 10 digits of precision. Similar expressions can be used to find longitude coordinates as well as coordinates in other formats. As another example, coordinates in the form of degrees, minutes, and seconds (e.g., N38° 35' 47") can be discovered by using the following code:

```
[NS]\\d\\d°\\s+\\d\\d'\\s+\\d\\d\"
```

This searches for either an initial N or S character followed by any two digits, the degree symbol, any amount of whitespace (space, tab, newline, other), any two digits, a single quote, any whitespace, any two digits, then finishing with a double quotation mark. The utility attained by harnessing the power of regular expressions quickly becomes apparent as many different forms of geoinformation can be discovered with little code.

Once regular expression code is written in proper syntax it provides an efficient way to search the large text databases the agent has retrieved from the web. Text surrounding any discovered coordinates can be stored and attached to the location as an attribute. The user defines the amount of text stored in the attribute with various combinations tested for the initial implementation of the agent. E.g., one may specify to use the 500 characters preceding a discovered coordinate or the 500 characters following the coordinate. Storing this information provides details on the context of the geospatial data including any available descriptions and metadata. The data can later be transformed into a standard geospatial format such as KML or a shapefile.

The transition from the discovered text-based geographic coordinates to proper geospatial data is aided by the use of the open source GIS toolkit GeoTools[23]. Written in Java™, GeoTools provides helpful methods for converting coordinates into geospatial data formats such as those officially accepted by the OGC. Code for KML output was custom-made for this research (although using GDAL can provide even more options for geodata formats) and can be imported into a virtual globe as discussed in the next section. The shapefiles created by this tool can be opened by an open source GIS such as QGIS or uDig.

3.4.5 Application

The application was designed using the Eclipse IDE version 3.3.1. Initial results were found using the agent with various starting web locations. Internet-based repositories of geospatial information were found for various user communities. The user community of climbers is explored in more detail using results from the Google search engine.

The agent created with this work retrieved resources found by searching for "latitude longitude climbing", "climb crag latitude longitude", and other phrases. The mass of data collected was then investigated for locations using the regular expressions methods described in the previous section. Coordinates were discovered

---

[23] http://www.geotools.org/

in both decimal degrees format (e.g., 38.596389, - 110.843611) and degrees, minutes, seconds format (e.g., N 36° 07'19", W 85° 17'07"). Whenever a location was found in a page the coordinates and text near the coordinates (e.g., 300 characters) were saved. This value is specified by the user allowing customization as desired. Once the coordinates and attributes were discovered the agent began the geodata creation process. In this example all of the discovered data were then used to create new GIS data (Figure 10) and virtual globe data (Figure 11).

Figure 10. Agent's shapefile results in Quantum GIS.

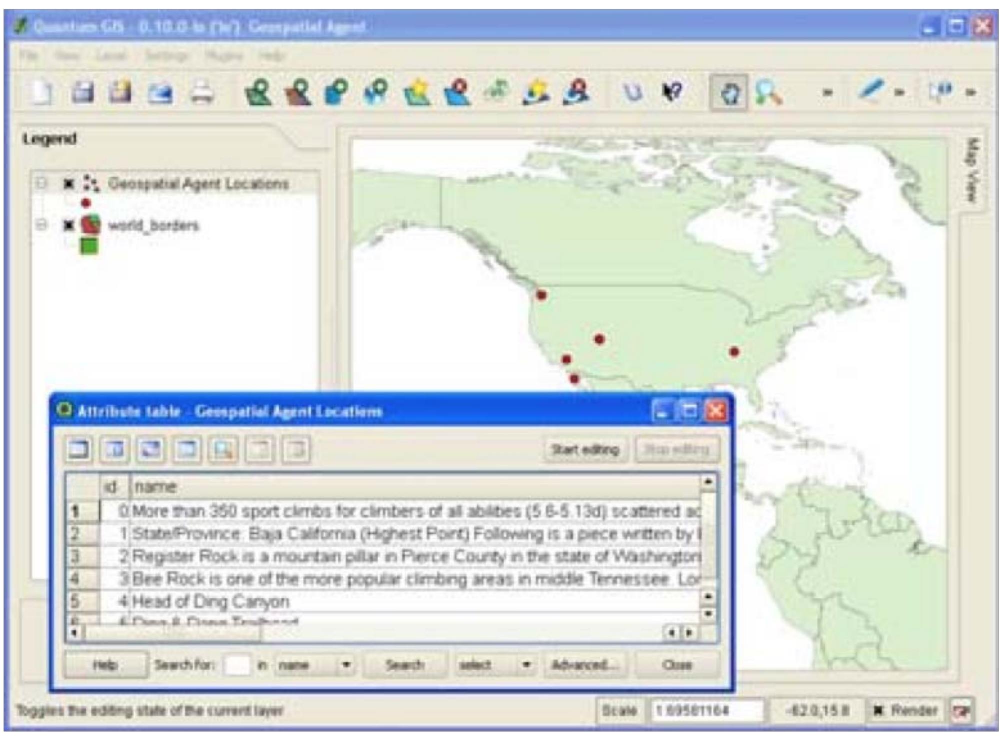
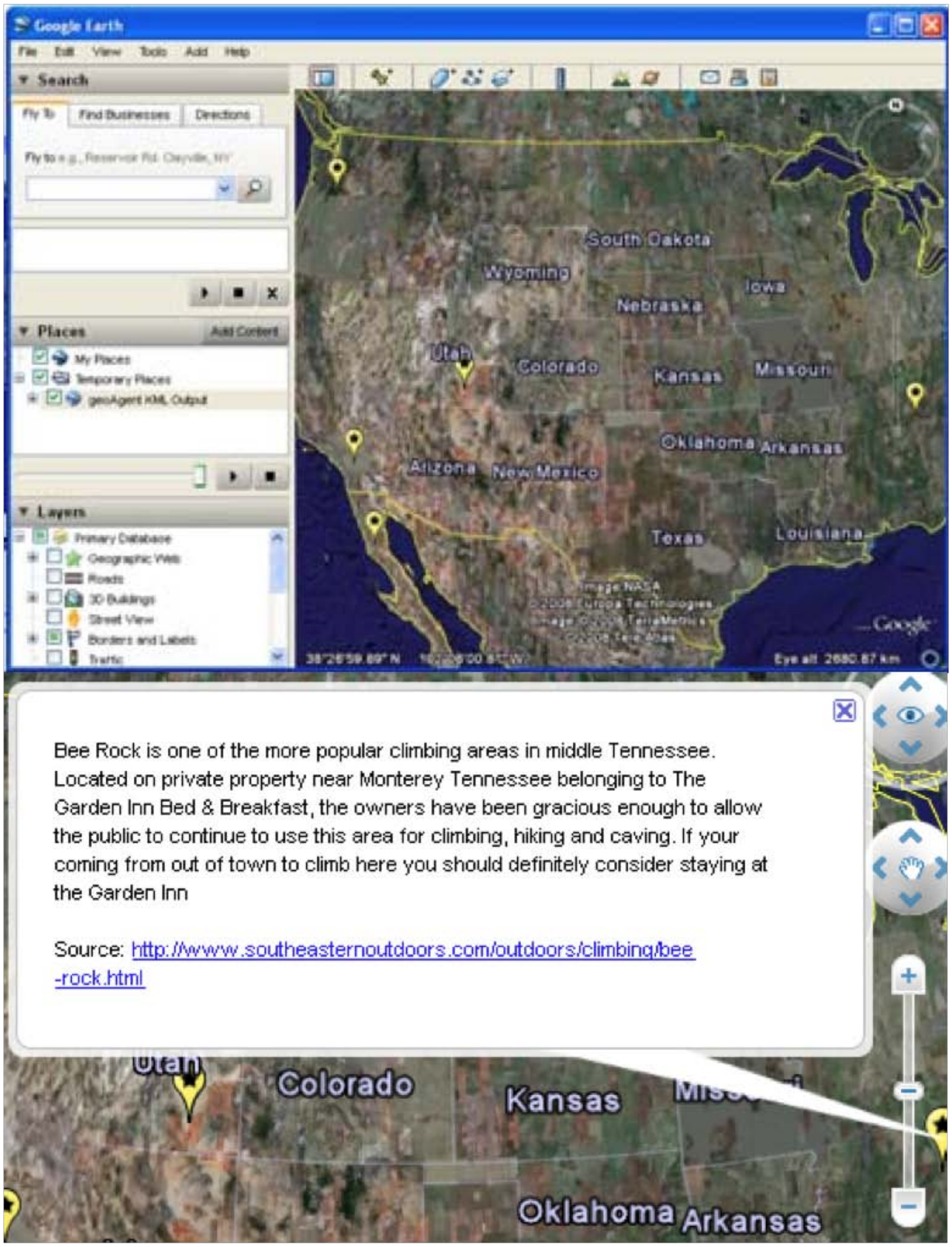


Figure 11. Agent's KML results in Google Earth



In the above figures a selection of resources discovered in the Americas is visualized. Additional information about each location is stored in an attribute table for the GIS data and in a description tag for the KML data. In creating the KML data a new Placemark Feature with associated Geometry is made for each location the agent discovers. Once the data files are created the user is able to perform any further operations desired using their tools of choice.

3.4.6 Discussion

The majority of the computation time for this agent depends on the speed of data retrieval from the web. Given that this agent can follow links from one country or continent to another, the amount of time an execution takes is highly variable. Times varied from a few seconds to many minutes in test runs with the prototype. The length of time it takes to retrieve data depends on the bandwidth of the system running the agent. The rest of the Java™ code utilizes regular expressions and basic input and output (I/O) techniques inheriting the efficiency of these methods as they are implemented in the language.

Quantification of data quality is a valid topic to be addressed in this context. This relates to quantifying the amount of value added to a discovered piece of geoinformation and assessing the data source. Identification of IP addresses can be used to determine general locations of persons providing VGI. These sources (e.g., Internet cafes, libraries) lend a starting point for establishing levels of data quality that pertain to particular user groups.

Determining the validity of discovered information for particular users is a task that can be approached from multiple directions. One initial method involves using cross page correlation where user communities specify a quantity that is a critical threshold for hinting at valid geospatial information. E.g., the user can specify that any discovered location and keyword pair be found at least 5 times before being stored in the database. Given issues of scale and resolution a buffer range can be specified as well, e.g., the 5 or more data entities must be within a circle of radius 100 meters to be considered appropriate for the user's dataset. These values vary between user groups and may be tweaked to a level that is most useful by the particular people searching for new knowledge.

With this agent implemented one begins to see how such a system adds value by synthesizing the collected data coordinates with useful information. An essential characteristic of deciding how much value the agent can add is whether or not the integrated data affects a decision (Frank 2003). Other variables to consider in quantifying the amount of value added to a location is the quantity of people in a specific user group in addition to the size and quantity of the database created for new locations. Currently for each location the coordinates, source page, and description information are gathered but adding additional attributes such as the amount of links to other pages can be used as a component of measuring and quantifying added value. Allowing for variability in the amount of text collected from a page provides a tool for analyzing how much data is needed to create useful information. For instance, one may collect 300, 500, and 700 words surrounding location texts and then examine what amount works best for the context.

3.4.7 Conclusions and Future Work

This section provides an implemented tool for searching the web for location data, amassing these data and mining them for coordinates, then creating new pieces of geodata. The amount of information gained varies between user groups. However, this agent does not retract value but rather adds value to any location data it discovers and creates a central repository available for further analysis. This research is also usable for exploration of the various types and quantities of geoinformation that users make available on the web. Finding common geospatial data formats published on the web will make this tool even more useful to additional user communities gaining utility from this research.

Adding a temporal extension to this agent would deliver both geospatial and time information pertaining to how up-to-date the created data is. This adds another factor for the relevancy and value of the added information. Utilizing HTTP headers retrieved via an HTTP request of a URL provides a value “Last-Modified” that tells how recently a page has been updated. The agent can then store this information with the geospatial data making it available for use in the analysis.

This work continues into the future with support for different types of geoinformation available on the web, such as Universal Transverse Mercator (UTM) and state plane coordinates. Here possibilities are discussed concerning the abilities attainable by combining geospatial agents with VGI, web crawling, and geospatial data mining and creation.

These methods are useful for a variety of reasons when combined with social network data. One is for analyzing the contents of profiles within social networks. For instance, data mining techniques used on web-based social network profiles provide details of what terms are most prominent. A prototype would do this for different geographies, e.g., separate datasets of profiles from the USA, England, and Australia, to demonstrate similarities and differences in word use, temporal descriptions, and presence of popular place names. Hence one would answer questions about what activities are taking place in a location and how a language is used differently in distinct regions.

Another utility of social network data mining is for creating a database of social network members, storing desired properties of each individual, e.g., languages spoken, number of pictures uploaded, to be used with analysis methods. It would answer questions about the distribution of languages within the network and which regions tend to have more detailed profiles. The next section uses an online survey in order to collect demographic data about CouchSurfing members and examine the influence of geography, age, and gender on social network participation.

3.5 Effects of Demographic Variables on CouchSurfing Activities

3.5.1 Introduction

This section uses survey methods to further examine how characteristics of CouchSurfers affect their use of the network. The work was done in order to examine real world members of the social network and their different levels of involvement. A survey was performed that gathered data from over 200 participants. Hence the sample data is drawn from the population of CouchSurfing network members. See the Appendix of this dissertation for the full survey. These data were then used with count-based models to see the effects of demographic variables on CouchSurfing activities.

Statistical methods are used in a large assortment of cases to identify, study, and solve a variety of problems in transportation (Goulias 2003). Data are collected and analyzed to aid in decision-making and understand complex processes. Statistical methods are not perfect hence errors and approximations may not always lead to appropriate actions. However, we are always searching for additional knowledge to advance our models, as model predictions are often better than nothing. Furthermore, there are future opportunities for social network researchers to utilize statistical methods for analyzing social network issues such as network measures and structures (Knoke and Yang 2008).

There are multiple steps in turning data into useful information. After data collection an appropriate model needs to be chosen based on characteristics such as the type of data (nominal, ordinal, interval, ratio, binary) and the phenomena that are being studied. While a complete coverage of all estimation models (such as those for discrete choices, logit, probit) is outside the scope of this writing, the appropriate family of models for working with counts of trips, CouchSurfing guest activities and CS hosting activities, are those known as count-based models.

The notion of count data models stems from the need to model processes that involve counts of activities such as trips or changes. Hence these counts are positive integers describing something such as the number of times an individual has participated in an activity. Therefore, using standard regression techniques is inappropriate, as it would yield incorrect results by predicting negative values for scenarios involving positive counts (Washington et al. 2003).

The Poisson regression model from Washington et al. (2003) is defined as:

$$P(y_i) = \frac{EXP(-\lambda)\lambda_i^{y_i}}{y_i!} \qquad \text{(Equation 1)}$$

with $P(y_i)$ being the probability of having y_i events. For example, if one were researching grocery store robberies $P(y_i)$ would be the probability that grocery store *i* would have y_i robberies per year. Poisson parameter λ_i is the expected amount of robberies per year at grocery store *i* where

$$\lambda_i = EXP(\boldsymbol{\beta}\mathbf{X}_i) \qquad \text{(Equation 2)}$$

and

$$LN(\lambda_i) = \boldsymbol{\beta}\mathbf{X}_i \qquad \text{(Equation 3)}$$

with $\mathbf{\beta}$ being the vector of coefficients and $\mathbf{X}_i$ the explanatory parameters. To estimate this model standard maximum likelihood methods are used with likelihood function:

$$L(\boldsymbol{\beta}) = \prod_i \frac{EXP\left[-EXP(\boldsymbol{\beta}\mathbf{X}_i)\right]\left[EXP(\boldsymbol{\beta}\mathbf{X}_i)\right]^{y_i}}{y_i!}$$

(Equation 4)

and log likelihood function:

$$LL(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left[-EXP(\boldsymbol{\beta}\mathbf{X}_i) + y_i\boldsymbol{\beta}\mathbf{X}_i - LN(y_i!)\right]$$

(Equation 5)

Ideally with the Poisson model the mean and variance are to be equal: E[yi] = VAR[yi]. However, if the mean is greater than the variance the data is underdispersed and if the mean is less than the variance the data is overdispersed. Examples and discussion of dispersion can be found in Karlaftis and Tarko (1998). The negative binomial model addresses this issue and is formed by changing Equation 2 to look as follows:

$\lambda_i = EXP(\boldsymbol{\beta}\mathbf{X}_i + \varepsilon_i)$ (Equation 6)

with the new term being gamma-distributed with mean 1 and variance α^2. This allows the mean and variance to differ resulting in:

$VAR[y_i] = E[y_i][1 + \alpha E[y_i]] = E[y_i] + \alpha E[y_i]^2$ (Equation 7)

Hence there are different methods available within the family of count-based models and the researcher must make an informed decision as to which is most appropriate, for an example see Ma and Goulias (1999). This decision is based on issues such as the knowledge of the activity behavior, the dataset, influences of each variable, and feedback that results from testing a variety of methods. Therefore a variety of methods will be tested below and decisions will be based upon multiple factors.

3.5.2 Data Collection

In this work 228 members of the CouchSurfing project from over 35 countries (see map in Figure 12 and distribution in Table 3) completed a survey to address the following hypotheses:

- Members of the CouchSurfing social network with more years of age tend to host more than younger members.
- Male users tend to play the role of guest more often than females.

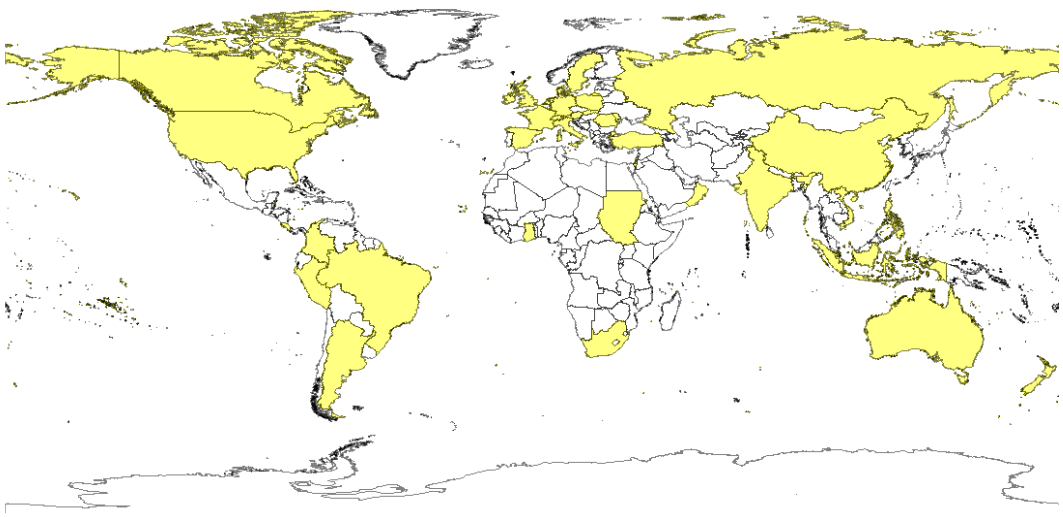
Hence the hypotheses state that hosts are more commonly older and female while guests are more likely male and have less years of age. This survey's questions pertained to the number of trips a member had taken in the past year using CouchSurfing, i.e., the count of how many times they had been a guest. The participants also provided the count of how many times they had been a host in the past year. Demographic data were also collected:

- Age
- Gender
- Current country of residence

Additional explanatory variables collected for each individual include:

- Whether they travel alone or with others
- If they use CS for any of the following purposes: tourism, business, study, cultural, language

Figure 12. Map showing survey participant countries as shaded, distribution in Table 3.



After acquiring human subjects approval from UCSB and before launching the survey publicly, a collection of local Santa Barbara, California, CouchSurfers was selected as a research focus group. These ten individuals were then invited to appear in person for the focus group gathering where they received a free meal and survey purpose and clarity was discussed. After this meeting, the survey was posted in message forums on the CouchSurfing.org website with no incentives resulting in 228 participants.

The 228 observations in this dataset had an average age of 32.3 and 38% of the participants were male meaning 62% were female. Females in this collected dataset were more willing to give their time since they represent 43% of the 2.5+ million members of CouchSurfing on 1 March, 2011. In future surveys, techniques may be used to ensure the dataset has more males than females. The ages of individuals that completed the survey varied between 19 and 65 whereas the ages of the 2.5+ million members of CS are between 18 and 89. Last, the average age in this dataset, 32, was slightly higher than 28, which is the average age of the 2.5+ million CouchSurfers in March 2011. While the collected dataset is not perfect, it represents a variety of the geographies and backgrounds of individuals that form the CS social network. The full descriptive statistics are in the figure below.

Figure 13. Descriptive statistics for sample containing 228 observations.

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases	Missing
All observations in current sample						
GUEST	7.87281	10.3765	.000000	60.0000	228	0
HOST	12.2939	18.3792	.000000	100.000	228	0
AGE	32.3026	11.4868	19.0000	65.0000	228	0
MALE	.381579	.486843	.000000	1.00000	228	0
PURPBIZ	.149123	.356993	.000000	1.00000	228	0
PURPTOUR	.842105	.365445	.000000	1.00000	228	0
PURPSTUD	.157895	.365445	.000000	1.00000	228	0
PURPCLTR	.675439	.469241	.000000	1.00000	228	0
PURPLANG	.390351	.488902	.000000	1.00000	228	0
TRAVALON	.697368	.460408	.000000	1.00000	228	0

These variables are described below:

- GUEST: Number of times the individual has been a guest in previous 365 days
- HOST: Number of times the individual has been a host in the previous 365 days
- AGE: Age of the individual
- MALE: Variable for gender: 1 for male, 0 for female
- PURPBIZ: Binary variable stating whether or not the member uses CS for business purposes.
- PURPTOUR: Binary variable stating whether or not the member uses CS for tourism purposes.
- PURPSTUD: Binary variable stating whether or not the member uses CS for study purposes.
- PURPCLTR: Binary variable stating whether or not the member uses CS for cultural purposes.
- PURPLANG: Binary variable stating whether or not the member uses CS for language purposes.
- TRAVALON: Binary variable stating whether or not the member travels alone when they use CS as a guest.

3.5.3 Count-based models

The count-based models used in this analysis were implemented in the LIMDEP software package NLOGIT 4.0.1. Initially all of the variables concerning demographic information, purposes, and traveling party composition were used to describe guest and host activities. This was done using Poisson and Negative Binomial models. Additionally this brings about the concept of Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB) models. These Zero-Inflated models handle situations where zero values of counts (guest and host activities here) can occur from qualitatively different situations (Washington et al. 2003). One situation for a zero value in the host count is that the individual has not hosted during the observation period (previous 365 days). A different situation for a zero value is the lack of an ability to host due to conditions such as unwelcoming roommates. Hence the zero-inflated models will also be examined with the collected dataset.

The histograms in Figures 14 and 15 show the distributions of the data for guest and host events. Here it is apparent there is inflation in zero values, e.g., 51 of the 228 individuals in the study had not participated in a guest event in the past year. This may be for qualitatively different reasons as explained in the previous paragraph but these graphs of activity frequencies further reiterate the potential use of zero inflated methods of analysis.

Figure 14. Histogram of guest counts for dataset of 228 participants.

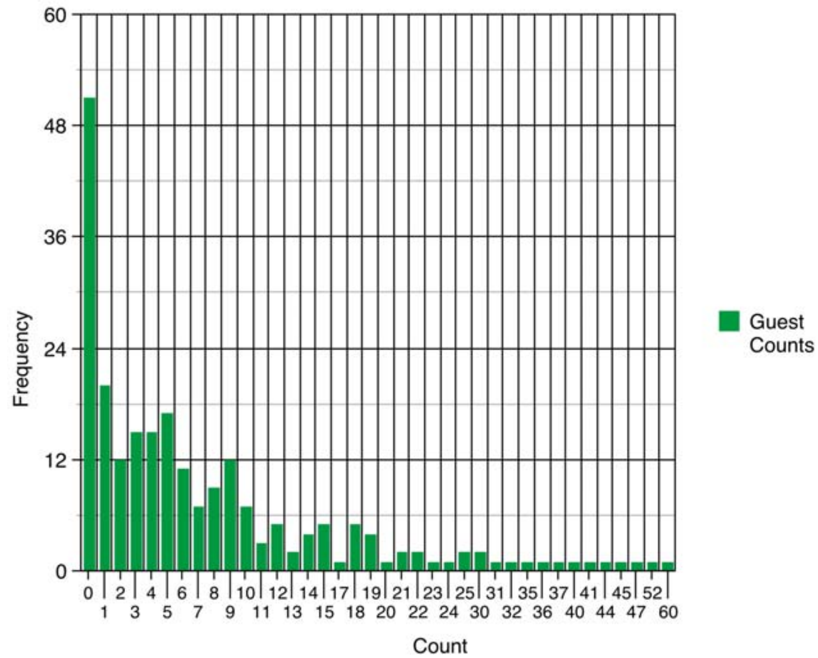
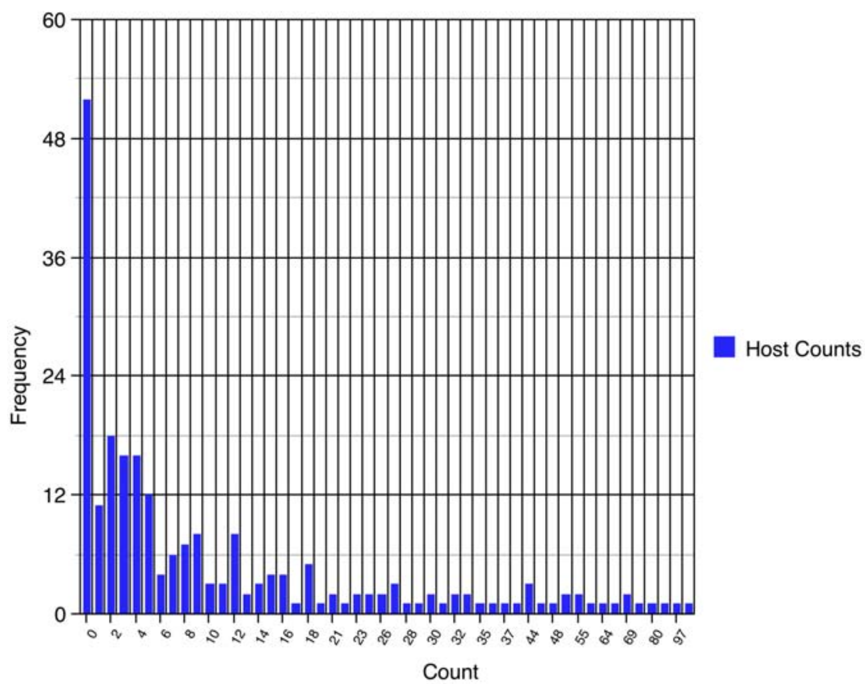


Figure 15. Histogram of host counts for dataset of 228 participants.



The initial runs in LIMDEP of the four models (Poisson, Negative Binomial, ZIP, ZINB) with GUEST as the dependent variable returned the following results:

Figure 16. Poisson model results.

```
+---------------------------------------------+
| Poisson Regression                          |
| Maximum Likelihood Estimates                |
| Model estimated: Nov 28, 2010 at 10:20:18PM.|
| Dependent variable                GUEST     |
| Weighting variable                 None     |
| Number of observations              228     |
| Iterations completed                  7     |
| Log likelihood function       -1407.954     |
| Restricted log likelihood     -1562.093     |
| Chi squared                    308.2783     |
| Degrees of freedom                    8     |
| Overdispersion tests: g=mu(i)  :  5.546     |
| Overdispersion tests: g=mu(i)^2:  5.724     |
+---------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	.79502842	.13164078	6.039	.0000	
MALE	.55390136	.04850253	11.420	.0000	.38157895
AGE	-.00315700	.00225174	-1.402	.1609	32.3026316
PURPBIZ	.45639633	.06141292	7.432	.0000	.14912281
PURPTOUR	.66477757	.08294114	8.015	.0000	.84210526
PURPSTUD	-.13123636	.06704090	-1.958	.0503	.15789474
PURPCLTR	.27913395	.05942352	4.697	.0000	.67543860
PURPLANG	.30084976	.05420990	5.550	.0000	.39035088
TRAVALON	.22496853	.05582595	4.030	.0001	.69736842

Figure 17. Negative Binomial model results.

Negative Binomial Regression	
Maximum Likelihood Estimates	
Model estimated: Dec 03, 2010 at 11:47:18AM.	
Dependent variable	GUEST
Weighting variable	None
Number of observations	228
Iterations completed	15
Log likelihood function	-686.9816
Number of parameters	10
Restricted log likelihood	-1407.954
Chi squared	1441.945
Degrees of freedom	1

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	.58072857	.43902909	1.323	.1859	
MALE	.56323268	.19365161	2.908	.0036	.38157895
AGE	-.00155573	.00792479	-.196	.8444	32.3026316
PURPBIZ	.52956148	.27563969	1.921	.0547	.14912281
PURPTOUR	.75915440	.23025352	3.297	.0010	.84210526
PURPSTUD	-.13547367	.28246995	-.480	.6315	.15789474
PURPCLTR	.31165691	.22787232	1.368	.1714	.67543860
PURPLANG	.28047790	.24500061	1.145	.2523	.39035088
TRAVALON	.29276767	.20923391	1.399	.1617	.69736842

Figure 18. Zero-Inflated Poisson (ZIP) Model Results

```
+----------------------------------------------------------------------+
| Zero Altered Poisson      Regression Model                           |
| Comparison of estimated models                                       |
|             Pr[0|means]       Number of zeros        Log-likelihood |
| Poisson          .00074   Act.=    51 Prd.=      .2     -1407.95419 |
| Z.I.Poisson      .21430   Act.=    51 Prd.=    48.9     -1140.90036 |
| Vuong statistic for testing ZIP vs. unaltered model is      5.2702  |
| Distributed as standard normal. A value greater than                 |
| +1.96 favors the zero altered Z.I.Poisson model.                     |
| A value less than -1.96 rejects the ZIP model.                       |
+----------------------------------------------------------------------+
+---------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+---------+-------------+---------------+--------+--------+----------+
---------+Poisson/NB/Gamma regression model
 Constant|   1.25441745       .04883337    25.688   .0000
 MALE    |    .48325881       .01798395    26.872   .0000    .38157895
 AGE     |    .00341937       .00078405     4.361   .0000   32.3026316
 PURPBIZ |    .36259708       .02175711    16.666   .0000    .14912281
 PURPTOUR|    .32540672       .02773042    11.735   .0000    .84210526
 PURPSTUD|   -.02327507       .02379236     -.978   .3279    .15789474
 PURPCLTR|    .17267843       .02104757     8.204   .0000    .67543860
 PURPLANG|    .34314623       .01987080    17.269   .0000    .39035088
 TRAVALON|    .15649375       .01963337     7.971   .0000    .69736842
```

Figure 19. Zero-Inflated Negative Binomial (ZINB) Model Results

```
+----------------------------------------------------------------------+
| Zero Altered Neg.Binomial Regression Model                           |
| Comparison of estimated models                                       |
|                 Pr[0|means]      Number of zeros        Log-likelihood |
| Poisson            .00074    Act.=    51 Prd.=      .2     -1407.95419 |
| Neg. Bin.          .07484    Act.=    51 Prd.=    17.1      -686.98155 |
| Z.I.Neg_Bin        .20641    Act.=    51 Prd.=    47.1      -683.73472 |
| Vuong statistic for testing ZIP vs. unaltered model is      1.8622   |
| Distributed as standard normal. A value greater than                 |
| +1.96 favors the zero altered Z.I.Neg_Bin model.                     |
| A value less than -1.96 rejects the ZIP model.                       |
+----------------------------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
---------+Poisson/NB/Gamma regression model
 Constant|     .83265224       .35610953     2.338   .0194
 MALE    |     .51192461       .15641588     3.273   .0011    .38157895
 AGE     |   -.00218358        .00642479     -.340   .7340   32.3026316
 PURPBIZ |     .49238543       .21904673     2.248   .0246    .14912281
 PURPTOUR|     .69682725       .18849004     3.697   .0002    .84210526
 PURPSTUD|   -.13548823        .22365784     -.606   .5447    .15789474
 PURPCLTR|     .29538680       .18514668     1.595   .1106    .67543860
 PURPLANG|     .24289198       .19402209     1.252   .2106    .39035088
 TRAVALON|     .25618315       .16827016     1.522   .1279    .69736842
```

Given the discussion of the phenomenon and its zero-inflated property, one would suspect the ZIP or ZINB to be appropriate in this scenario and in the results above this is shown to be the case with ZIP. As described in the figures 17 and 18 the Vuong statistic is used to test whether or not a zero-inflated model is favored. In the case of ZINB, the value 1.8622 does not strongly favor or reject ZINB since it is neither greater than 1.96 nor less than -1.96. However, the Vuong statistic for the ZIP, 5.2702, favors using ZIP. With the ZIP all of the independent variables are significant at the 5% level with the exception of PURPSTUD (.3279). As a result this variable was removed from further analysis. The lack of significance of this variable (individuals using the CS network for the purpose of study) may be a result of a small dataset or that people really are not using CS for study purposes. The latter is plausible as other purposes (such as business, tourism, and language) may be a much more common use of this social network. From these results the influence of these variables is shown to be positive using ZIP.

An additional test was also done to examine the fit of the four models: Poisson, Negative Binomial, ZIP, and ZINB. This involved taking the predicted value for each model and calculating the difference between the predicted value and the actual value for each of the 228 individuals in the survey dataset. As an example, imagine that the ZIP model has predicted that a particular individual would have a guest count of 15 according to their values for the variables used in this model: MALE, AGE, PURPBIZ, PURPTOUR, and so on. There is also the actual count of guest activities that the individual reported in the survey results; for the example this particular individual reported 17 guest activities. Hence the absolute value of the difference between the observed (17) and predicted (15) is 2. This was done for all of the 228 individuals and each of the four models. The 228 absolute values of the difference were summed for each of the four models resulting in a value that describes how well one model predicted vs. another. These calculated values are below in Table 2:

Table 2. Differences between models for observed and predicted values.

Poisson	Negative Binomial	ZIP	ZINB
1536.13	1544.58	1435.38	1412.80

These results highlight that the Zero-Inflated versions of the models were better predictors: 1536.13 vs. 1435.38 and 1544.58 vs. 1412.80. This further suggests the appropriateness of the Zero-Inflated models for this work as was discussed earlier and shown in the histograms. While the ZINB and ZIP values above are close, given my extensive experience with the social network the independent variables used in the model (MALE, AGE, PURPBIZ, PURPLANG, TRAVALON, etc.) do have an effect on a CouchSurfer's count of activities in the network. Given that those variables are influential and significant in the ZIP model, it remains appropriate for this research.

To examine the hypotheses stated above, concerning age and gender, the ZIP model is used for guest and host counts with the PURPSTUD variable removed. The results are shown in Figures 20 and 21.

Figure 20. ZIP for Guest Counts

```
+----------------------------------------------------------------------+
| Zero Altered Poisson      Regression Model                           |
| Comparison of estimated models                                       |
|             Pr[0|means]       Number of zeros        Log-likelihood  |
| Poisson          .00074   Act.=    51 Prd.=       .2     -1409.91185 |
| Z.I.Poisson      .21445   Act.=    51 Prd.=     48.9     -1140.95819 |
| Vuong statistic for testing ZIP vs. unaltered model is      5.2730   |
| Distributed as standard normal. A value greater than                 |
| +1.96 favors the zero altered Z.I.Poisson model.                     |
| A value less than -1.96 rejects the ZIP model.                       |
+----------------------------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
---------+Poisson/NB/Gamma regression model
 Constant|    1.25495864       .04830780    25.978   .0000
 MALE    |     .48355726       .01794329    26.949   .0000     .38157895
 AGE     |     .00348617       .00077971     4.471   .0000   32.3026316
 PURPBIZ |     .35663924       .02068740    17.239   .0000     .14912281
 PURPTOUR|     .32173574       .02604496    12.353   .0000     .84210526
 PURPCLTR|     .17152231       .02095573     8.185   .0000     .67543860
 PURPLANG|     .33844557       .01880021    18.002   .0000     .39035088
 TRAVALON|     .15739183       .01961303     8.025   .0000     .69736842
```

Figure 21. ZIP for Host Counts

```
+----------------------------------------------------------------------+
| Zero Altered Poisson      Regression Model                           |
| Comparison of estimated models                                       |
|             Pr[0|means]       Number of zeros        Log-likelihood |
| Poisson           .00001   Act.=     52 Prd.=      .0      -2515.69081 |
| Z.I.Poisson       .22738   Act.=     52 Prd.=    51.8      -1941.81022 |
| Vuong statistic for testing ZIP vs. unaltered model is      5.9337   |
| Distributed as standard normal. A value greater than                 |
| +1.96 favors the zero altered Z.I.Poisson model.                     |
| A value less than -1.96 rejects the ZIP model.                       |
+----------------------------------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
---------+Poisson/NB/Gamma regression model
 Constant|    1.78814559       .02367362    75.533   .0000
 MALE    |     .25634347       .00855411    29.967   .0000     .38157895
 AGE     |     .01542153       .00042918    35.932   .0000   32.3026316
 PURPBIZ |     .21197847       .01139673    18.600   .0000     .14912281
 PURPTOUR|     .22887094       .01416424    16.158   .0000     .84210526
 PURPCLTR|     .37335935       .01141655    32.703   .0000     .67543860
 PURPLANG|    -.10818015       .00921431   -11.740   .0000     .39035088
 TRAVALON|    -.14195989       .00951460   -14.920   .0000     .69736842
```

Again we see that the Vuong statistic (5.2730 and 5.9337) favors the use of ZIP. Between these two results one sees the influence of different demographic variables on an individual's activity as a host or as a guest. With respect to gender, the coefficient for MALE in the ZIP for guest counts, .4835, is positive as it is in the ZIP for host counts, .2563. Thus in this dataset male CouchSurfers are more likely to be guests and hosts. Hence the males in this sample are traveling and hosting more with CouchSurfing, which may be related to issues of safety for women travelers.

With respect to age variable the coefficient in the guest model is smaller, .0034 vs. .0154 in the host. Thus meaning that older CouchSurfers here have higher counts of hosting activities. This may be due to older people being more settled and generally having more space in their home for hospitality. Older individuals may also have an “empty nest” once their children have grown up and moved away, leaving space for guests such as CouchSurfers.

Concerning the variable of using CS for business purposes (PURPBIZ) the coefficient for being a guest is larger than that for being a host, .3566 vs. .21197. This suggests using the social network for business has a stronger influence on the number of guest activities than the number of host activities. Those that are guests more often than hosts have the mobility to achieve their business goals on a global scale. On the other hand, those who host more often have to wait for someone to arrive in order for them to share any business purposes. One example is an individual attending a conference or meeting and using CS to find a place to stay vs. other lodging such as a hotel. In this example the individual is a guest and has the mobility to stay with any member on any continent. If a less mobile individual, who participates in more hosting activities, also has a restaurant business, they are more constrained by other factors that limit the amount of guests they receive such as the appeal of their city for travelers.

The PURPTOUR variable also has a larger coefficient in the guest counts (.3217 vs. .2288) for this sample hence influencing an individual’s count of guest activities more than their hosting. Those that use the social network more for traveling and being a guest can be tourists more so than individuals that are more stationary and spend most of their time being a host. The opposite happens in the case of the

PURPCLTR variable (.1715 in guest counts and .3733 in host counts) suggesting hosts are welcoming people into their home for culture sharing purposes. This may also have an influence on a host's preference to have guests from far-away places or countries other than their own.

With the variable for language purposes (PURPLANG) there is a flip from positive to negative in the sign of the coefficient: .3384 for guest counts and -.1081 for host counts. Thus using the social network for language purposes in this sample is more associated with guest activities than host activities. An individual's setting plays a role in one's ability to learn a foreign language where immersion in the foreign country is suggested as a good strategy for learning the appropriate language. This is the role a guest can play more easily than a host since they are more mobile.

The last variable, TRAVALON, describes "with whom" (Goulias and Kim 2005) an individual travels or if they travel alone when they are a guest. Here there is another switch in the sign of the coefficient from positive to negative: .1573 for the guest counts and -.1419 for the host counts. This suggests that those who travel alone are likely to participate in more guest activities than host activities. An independent, solo traveler may feel more confident in being in a foreign place and thus use the network more for touring purposes. Whereas an individual that does not prefer to travel alone can feel more comfortable in their home location and participate in more hosting activities.

In the previous models geographic elements are implicit such as where someone has traveled and where their guests have traveled. The following uses more explicit geographic details derived from the country of residence variable collected for each of the 228 survey participants. The map in Figure 12 highlights the countries of residence for the participants and these countries are listed in Table 3 along with a count of how many are from each place. Table 4 shows the counts of participants per continent.

Table 3. Counts of survey participants with country of residence.

Country	Count	Percent of Sample
Argentina	4	1.7
Australia	1	.4
Austria	19	8.3
Brazil	20	8.7
Canada	28	12.2
Cape Verde	1	.4
China	2	.8
Colombia	4	1.7
Costa Rica	1	.4
Denmark	1	.4
France	10	4.3
Germany	17	7.4
Ghana	1	.4
India	4	1.7
Indonesia	9	3.9
Ireland	2	.8
Israel	16	7.0
Italy	12	5.2
Moldova	1	.4
Morocco	1	.4
Netherlands	3	1.3
New Zealand	1	.4
Oman	1	.4
Peru	1	.4
Philippines	1	.4
Poland	2	.8
Romania	1	.4
Russia	1	.4

Serbia	1	.4
Singapore	7	3.0
Slovakia	2	.8
South Africa	5	2.1
Spain	1	.4
Sudan	1	.4
Sweden	1	.4
Switzerland	1	.4
Turkey	4	1.7
United Kingdom	7	3.0
USA	30	13.1
Vietnam	3	1.3

Table 4. Counts of survey participants with continent of residence.

Continent	Count	Percent of Sample
Africa	9	3.9
Asia	48	21.0
Europe	81	35.5
Latin America	30	13.1
North America	58	25.4
Oceania	2	.8

The following six countries were chosen for further analysis due to their counts being 15 or greater: Austria, Brazil, Canada, Germany, Israel, USA. Binary dummy variables were created for each of the respective countries and count-based models are used to examine whether or not these countries of residence play a role in guest and host activity counts. Here again the Zero-Inflated Poisson model is used as it is appropriate given the Vuong statistic, significance of variables, and predicted number of zero counts. Below are the ZIP results for guest and host counts using the country variables along with those of age and gender.

Figure 22. ZIP for Guest Counts with Country of Residence

```
+----------------------------------------------------------------------+
| Zero Altered Poisson      Regression Model                            |
| Comparison of estimated models                                        |
|               Pr[0|means]       Number of zeros        Log-likelihood |
| Poisson            .00050   Act.=     51 Prd.=      .1     -1497.33300 |
| Z.I.Poisson        .22331   Act.=     51 Prd.=    50.9     -1188.87288 |
| Vuong statistic for testing ZIP vs. unaltered model is       5.5985   |
| Distributed as standard normal. A value greater than                  |
| +1.96 favors the zero altered Z.I.Poisson model.                      |
| A value less than -1.96 rejects the ZIP model.                        |
+----------------------------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
---------+Poisson/NB/Gamma regression model
 Constant|    2.15233785       .02782288    77.359   .0000
 MALE    |     .33875530       .01613910    20.990   .0000    .38157895
 AGE     |    -.00018462       .00072923     -.253   .8001   32.3026316
 AUSTRIA |     .23819644       .02097353    11.357   .0000    .08333333
 BRAZIL  |     .00084112       .03421334      .025   .9804    .08771930
 CANADA  |    -.01019051       .02750658     -.370   .7110    .12280702
 GERMANY |     .29749644       .02445258    12.166   .0000    .07456140
 ISRAEL  |    -.05648484       .04073498    -1.387   .1656    .07017544
 USA     |    -.17304321       .03276374    -5.282   .0000    .12719298
```

Figure 23. ZIP for Host Counts with Country of Residence

```
+----------------------------------------------------------------------+
| Zero Altered Poisson      Regression Model                           |
| Comparison of estimated models                                       |
|             Pr[0|means]       Number of zeros        Log-likelihood |
| Poisson          .00001    Act.=    52 Prd.=      .0     -2508.46025 |
| Z.I.Poisson      .22840    Act.=    52 Prd.=    52.1     -1928.78898 |
| Vuong statistic for testing ZIP vs. unaltered model is      5.9233   |
| Distributed as standard normal. A value greater than                 |
| +1.96 favors the zero altered Z.I.Poisson model.                     |
| A value less than -1.96 rejects the ZIP model.                       |
+----------------------------------------------------------------------+
+---------+--------------+----------------+---------+---------+----------+
|Variable| Coefficient   | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+---------+--------------+----------------+---------+---------+----------+
---------+Poisson/NB/Gamma regression model
 Constant|    2.23255929       .01696605   131.590   .0000
 MALE    |     .19376814       .00844369    22.948   .0000    .38157895
 AGE     |     .01442078       .00046522    30.998   .0000   32.3026316
 AUSTRIA |   -.08754755        .01874922    -4.669   .0000    .08333333
 BRAZIL  |     .07848315       .01382095     5.679   .0000    .08771930
 CANADA  |   -.20259169        .01603738   -12.632   .0000    .12280702
 GERMANY |     .43856182       .01171203    37.445   .0000    .07456140
 ISRAEL  |   -.47582557        .03889416   -12.234   .0000    .07017544
 USA     |   -.21349984        .01573005   -13.573   .0000    .12719298
```

After introducing this country of residence data the influence of gender on an individual's role in the network remains similar. The guest model has a positive gender coefficient as does the host model meaning guests and hosts are more likely to be males. The age variable is not significant at the 5% level in the guest model here hence no conclusions are made with respect to age. The country variables significant in both guest and host models are Austria, Germany, and USA. Between these three countries the coefficient for the USA is the lowest, suggesting that in this sample the social network members from the USA are participating in less guest and host activities compared to those in Austria and Germany as well as the countries not included in the model which have a coefficient of zero. This may be due to contemporary travel trends or the recent status of economies. Also of note here is how much German members stand out with respect to hosting activities, a much higher coefficient than Austria and the USA, suggesting Germany is quite a hospitable region. With the sample size of 228 the counts of individuals in these countries of residence were all 30 or less (Table 3) and this resulted in a number of variables being insignificant. The next models use the continent of residence data shown in Table 4, which have larger counts of individuals.

These four continents from Table 4 were chosen for further investigation in a count-based model scenario: Asia, Europe, Latin American, and North America. Each of the four continents has its corresponding binary dummy variable that is used in the ZIP model along with age and gender variables. The predicted number of zero counts, Vuong statistic, and significance of variables suggest a fit for the ZIP model as shown below in the following figures.

Figure 24. ZIP for Guest Counts with Continent of Residence

```
+----------------------------------------------------------------------+
| Zero Altered Poisson      Regression Model                           |
| Comparison of estimated models                                       |
|             Pr[0|means]       Number of zeros        Log-likelihood |
| Poisson          .00061   Act.=     51 Prd.=      .1     -1460.34071 |
| Z.I.Poisson      .21849   Act.=     51 Prd.=    49.8     -1167.29178 |
| Vuong statistic for testing ZIP vs. unaltered model is      5.8609   |
| Distributed as standard normal. A value greater than                 |
| +1.96 favors the zero altered Z.I.Poisson model.                     |
| A value less than -1.96 rejects the ZIP model.                       |
+----------------------------------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
---------+Poisson/NB/Gamma regression model
 Constant|    1.46724463       .09671523    15.171   .0000
 MALE    |     .34245345       .01502094    22.798   .0000     .38157895
 AGE     |    -.00238960       .00078601    -3.040   .0024   32.3026316
 ASIA    |     .67114309       .09366456     7.165   .0000     .21491228
 EURO    |     .97006052       .09495543    10.216   .0000     .35526316
 LATAMER |     .67501941       .09743792     6.928   .0000     .13157895
 NTHAMER |     .63273610       .09338925     6.775   .0000     .25438596
```

Figure 25. ZIP for Host Counts with Continent of Residence

```
+----------------------------------------------------------------------+
| Zero Altered Poisson      Regression Model                           |
| Comparison of estimated models                                       |
|             Pr[0|means]       Number of zeros        Log-likelihood |
| Poisson          .00001   Act.=     52 Prd.=      .0     -2467.59536 |
| Z.I.Poisson      .22797   Act.=     52 Prd.=    52.0     -1903.45714 |
| Vuong statistic for testing ZIP vs. unaltered model is      6.1205   |
| Distributed as standard normal. A value greater than                 |
| +1.96 favors the zero altered Z.I.Poisson model.                     |
| A value less than -1.96 rejects the ZIP model.                       |
+----------------------------------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
---------+Poisson/NB/Gamma regression model
 Constant|    1.12871002       .03010209    37.496   .0000
 MALE    |     .19500823       .00841811    23.165   .0000    .38157895
 AGE     |     .01542867       .00038781    39.784   .0000  32.3026316
 ASIA    |     .96757352       .02336717    41.407   .0000    .21491228
 EURO    |    1.24691408       .02319328    53.762   .0000    .35526316
 LATAMER |    1.05762951       .02581196    40.974   .0000    .13157895
 NTHAMER |     .81303205       .02287345    35.545   .0000    .25438596
```

This model agrees with the previous models that in this sample guests and hosts tend to be male more than female. The age variable is significant in both ZIP models using continent data and the change of sign (-.0023 and .0154) between guest and host indicates that the older a person gets the less they will participate in guest activities but they will participate in more hosting activities. As mentioned previously some explanations for this may be due to mobility at different ages and/or stages of life and an older person having a higher likelihood of having accumulated more resources to share with others. The continent data shows Europeans being the most active in this sample for both guest and host activities whereas the least active here are those in North America. Another interesting point is where Latin America stands in its performing of guest activities vs. host activities. In the guest model the coefficients for the continents of Asia (.6711), Latin America (.675), and North America (.6327) do not vary greatly but in the host model Latin America separates itself more from North America and Asia (1.0576 for Latin American vs. .813 for North America and .9675 for Asia) suggesting that Latin America engages in more hosting activities than the other two continents. This hints at members of the CouchSurfing social network in Latin America being more hospitable but also may be a result of economic restraints. For example, countries in Latin America are not members of the USA visa waiver program. Hence if members in these locations wish to participate in this social network it may be easier for them to play the role of host vs. guest. Last, this model suggest that Asia, Europe, Latin America, and North America with their positive coefficients have higher frequencies of guest and host counts than Africa and Oceania.

3.5.4 Conclusions and Future Work

This section has used data from a survey of 228 individuals from the CouchSurfing social network. Count-based models have been used with this sample to examine characteristics that affect how much an individual performs host and guest activities. Future research will be done to work with larger datasets that contain members from more locations. Additionally, more variables (such as positive, negative experiences, strength of connection with friends) will be collected to further determine the influences for an individual to perform one role in a social network vs. another role.

Regarding the initial hypotheses, the count-based models have suggested that in this sample older members are hosting more and male users are being guests and hosts more often than females. Future work may also be done to examine as to why males are traveling and hosting more than females. It is also of note that the estimates the models provide are indeed estimates and more work with random, larger samples will be helpful in further investigating the hypotheses.

Some ideas for why these differences in age and gender are happening are given above but a deeper explanation is valuable to further fine-tune the social network as a whole such that it continues to be successful. Even though CouchSurfing members are the greater population studied here these results may be of use to other travel and social networks. The results can help individuals associated in all facets of tourism (such as transport, lodging, alimentation, tours) to better understand how to serve their clients and specifically what are the characteristics of the clients they are serving. Future samples can be gathered from other populations in order to explore the differences and similarities in various populations.

In this sample whether or not the participants travel alone was used as an independent variable. An individual may state they are a solo traveler because they are the only one staying with a host. However, this person could be traveling with others that are staying with different hosts. Does this count as truly traveling alone? The entire traveling party size and its composition is an intriguing topic for future work. For instance, what differences and similarities are there between traveling families (e.g., mother, father, and children) versus a group of friends? Hence there is interaction and schedule coordination between multiple guest travelers that may be collected in future strategic surveys.

The use of geographic information in the form of countries and continents further reiterates the influence of age and gender on travel while suggesting how the location where a person lives can affect their role and activity in a global, Internet-based social network. All of this comes together and highlights the intrinsic connection between virtual and physical presence in modern, Internet-based social networks.

The third chapter tackled the third and fourth objectives in Section 1.2 which concern geospatial data mining and the effects of demographic variables on social network activity. Mining social network data to find and discover activities or connections between locations is a useful element to add to the conceptual model from the previous chapter such that individuals can plan their travels using activity information. This would allow for an individual to first choose an activity in which they wish to participate and then the corresponding potential locations are presented. Likewise the demographic variables relate to the CS network as a whole such that one is aware of what demographics are using the network and to what extent. Thus if the CS network desires to increase activity levels of certain roles with existing members these results are a starting point for who to target for what.

4. Location Based Social Network Capital

4.1 Social networks and measures of social capital

Currently, there is limited knowledge about the relationships connecting social and spatial interactions (Carrasco et al. 2008, Axhausen 2005, Winter and Robins 2007). Yet Internet-based social networks are pervasive in the present day and this research adds to our knowledge of them by explicitly combining spatial behavior with social networks. “The Internet leads to new forms of social capital that cannot be easily captured with existing forms of measurement” (p. 124, Quan-Haase and Wellman 2004). Measurement approaches have been suggested for Facebook™ (Ellison et. al 2007) and mental health (Congdon 2010) but for social networks with physical and virtual location requirements (Pultar and Raubal 2009a) techniques for assessing social capital have not been explored. Social capital in social networks with both physical and virtual location requirements is addressed in this chapter. The proposed technique makes use of the geographic distribution of an individual’s connections. Also, the various roles a user performs in the network are drawn upon in these methods for social capital. The main goal of this chapter is to develop a formal measure for social network capital considering physical and virtual location and test it in a real-world scenario.

Modern digital social networks are commonly embedded in the virtual world. Recent technologies have made it easier for individuals to share a location with increasing mobility using wireless and handheld devices (Raubal 2011). This affords a return of social networks to the real, physical world. Scenarios utilizing these dimensions are referred to as Location Based Social Networks (LBSN), which are "social network services where people can track and share location related information with each other, via either mobile or desktop computers" as defined by the 2009 International Workshop on LBSN (Zhou and Xie 2009). See Section 3.1 of this dissertation for more about LBSN. Given the steady importance of an individual's location these LBSN have an increasing utility along with unique forms of social capital associated with them.

In the CouchSurfing (CS) social network members perform both virtual and physical events, making it a good fit for this research. The reader is directed to Section 2.1.2 above for a review of the CS network.

Time geography (Hägerstrand 1970) has been used to represent paths of individual movements through physical space for various periods of time (Frändberg 2008). Though in this network both travel movements as well as stationary hosting activities are key components for an individual using the network. The stationary hosting activities are virtual travel instances where culture, language, and other information are exchanged. Hence time geography principles will be used here for this combination of physical and virtual travel. However, in time geography an agent is at one place at one time in the physical travel sense. Yet here we are suggesting a host can perform virtual travels by hosting someone from another place as they experience language, culture, food, and so on. Additionally, a person can host more than one person at a time in a sense being at multiple locations at one time, e.g., hosting a person from Germany at the same time as hosting a person from Japan. Hence this one-to-one relationship is extended in the given scenario.

A 3-D spatiotemporal visualization technique utilizing these concepts is implemented. This demonstrates a user's social network capital growth with respect to their participation level and travel movements. In addition, equations are presented as a complementary quantitative approach for measuring social network capital combining virtual and physical spaces. This chapter specifically addresses the following research questions:

- How is an individual's social network capital represented in a setting combining physical and virtual spaces?
- How do geographical distances influence measures of social capital?
- How can different roles of a member in a social network affect social capital?

Social capital combines experience, activity, and respect among other concepts. Regarding a digital social network, interactions occur between members that affect their ability to use network resources in the future. Measures of an individual's capabilities are influenced by their previous actions. A higher number of established connections signifies a person's involvement within the network, which means higher levels of trust from others in the network. In addition to the amount of links an individual has, the distance (geographic, cultural, international borders, or other) of each friend plays a part in the overall measurement of one's commitment to the network. Last, members of a social network perform different functions or tasks for the group, e.g., administrative vs. publicity. The role (e.g., accountant, host, president) of each individual affects their measure of social capital as network members see some positions as more critical than others.

An experiment is performed as part of this research to test the appropriateness of the components given in the above paragraph. Data for this chapter is acquired from the Internet-based CouchSurfing network. Individual travel activity profiles are created from data describing where network members have stayed. Using this data combined with time geography results in 3-D geovisualization of social network capital. In addition, the proposed equations provide a quantitative approach for measuring social capital in the network. The survey is performed to determine whether this capital measure reflects the view of experienced users. These results answer questions of travel behavior using physical and virtual spaces concerning:

- Representation of social capital, both visual and computational.
- Utility of a geovisualization resource showing an individual's social network activity: spatiotemporal variation, breadth, and intensity.
- Influence of geography and location on an individual's social capital.
- Change in social capital worth between different roles of network members.

The purpose of this chapter is to describe social capital measures in social networks with a particular focus on the spatial distribution of connections and different member roles. The next section discusses related work in the fields of location based services and social capital. Section 4.3 presents methods for measuring social capital in social networks. The design of the experiment developed to evaluate these measures is introduced in Section 4.4. Section 4.5 reports the results of the experiment and Section 4.6 discusses these results. The chapter concludes with a final discussion and areas for future research.

4.2 Related Work

Time geography (Hägerstrand 1970) focuses on the individual's movement through space and time. See Section 2.1.5 for a review of time geography. The time scale pertinent in this work is selective (Frändberg 2008) and on the order of months rather than hours or one day.

With increasing availability and coverage of Information and Communication Technologies (ICT), time geography and its integration with ICT is an active topic among researchers. Raubal et al. (2004) use location based services (LBS) and an extended theory of affordances with time geography to focus on an individual's specific preferences. An overview of the interaction between human activity-travel behavior and ICT is given by Kwan et al. (2007). Couclelis (2009) presents a new conceptual framework for time geography with a focus on ICT.

Social capital has become a more and more popular term in the recent decades. The precise definition has been debated and that evolution has been well analyzed (Fulkerson and Thompson 2008, Schafft and Brown 2003). Theoretical frameworks have been developed (Coleman 1988) but some present social capital as a process (Putnam, Leonard, and Nanetti 1993) and others as a resource (Bourdieu 1983, Loury 1977). There are also those that examine this concept given the long ancestry of the human species. From the *Journal of Economic Psychology* there is the evolutionary psychologist's view that "social relationships that do not even indirectly or remotely contribute toward individuals' reproductive success do not count as social capital" (p. 877, Kanazawa and Savage 2009). The final, lasting definition of social capital may use ideas from multiple camps, but for now the term remains heavily disputed.

Putnam (1995) defines social capital "by analogy with notions of physical capital and human capital - tools and training that enhance individual productivity - "social capital" refers to features of social organization such as networks, norms, and social trust that facilitate coordination and cooperation for mutual benefit." (p. 67) There are issues of scale with social capital as it can pertain to an individual, a group of people, a city, or a government. The ways these analysis units gain and lose social capital vary widely and this makes their quantification complex.

Social capital can offer a different perspective to economic methods: "Current interest in the concept of social capital in the field of national development stems from the limitations of an exclusively economic approach toward the achievement of the basic developmental goals…" (p. 529, Portes and Landolt 2000). Socialization methods have changed over the decades with more virtual relationships becoming available. Putnam (2000) discusses the increasing disconnect individuals have within their homes and communities. He also discusses bridging and bonding social capital, which are sometimes referred to as weak and strong ties (Grannovetter 1973). Bridging social capital or weak ties are acquaintances that are not very close but where a lot of information is shared. Strong ties are more likely family members and best friends that can provide emotional support whereas a new job offer is more probable to come from a weak tie. In the network discussed here most of the connections will be of the bridging capital form but some develop into bonding capital.

4.3 Capital Measures

As discussed in the related work section, social capital means different things to different people. Generally social capital has an effect on a person or group's abilities and a higher amount of social capital increases their productivity. With an egocentric approach the raw number of connections for an individual is an indicator for their amount of social capital. Characterizing these connection links as either weak or strong ties affects the influence of each tie. In a modern digital social network such as LinkedIn™ or Facebook™ a raw count of "friends-links" is a measure of social capital. This is represented as:

$$C_t = N_l \qquad \text{(Equation 8)}$$

where C_t is the total measure of an individual's social capital as defined by N_l, the number of links or connections an individual has.

Here a person that maintains a larger social network has greater access to more resources, e.g., employment opportunities in the case of LinkedIn™. This is a basic approach that leads to questions such as:

- How much more capital do strong ties provide vs. weak ties?
- How does real-world distance between individuals shape their capital?

The methods proposed in this section address the questions of how different roles in a social network affect social capital combined with how geographic location influences a person's social capital.

We will use the CouchSurfing (CS) network (Pultar and Raubal 2009b) described in Sections 4.1 and 2.1.2 as a case study. In this social network there are distinct roles that individuals perform in the network, e.g., being a guest vs. being a host. Additionally, there is the breadth of international network members with individuals on each continent (See Section 3.5). This provides a geographically spread network with which to study the influence of location on social capital with the following methods.

Thus, for an individual in the CS social network, the total measure of social capital (C_t) can be calculated according to the following equation:

$$C_t = \alpha C_g + \beta C_h \qquad \text{(Equation 9)}$$

where C_g is an individual's network capital measured by how many guests an individual has had and the proximity of each guest. The number of hosts an individual has had and the proximity of those hosts determine the value for C_h. The weights, α and β, are used to put more influence on having guests or having hosts. The sum $\alpha + \beta = 1$ and an equal weighting would signify both weights to be 0.5 whereas other values for α and β would stress the significance of one activity over another. The initial magnitudes for α and β are empirically derived from a survey of experienced network members. The components C_g and C_h are defined by the following equations:

$$C_g = \sum_{i=1}^{i=N_g} [D_g(i)]^{\frac{1}{2}} \qquad \text{(Equation 10)}$$

where

N_g = Number of guests an individual host has had for the given time interval

and

D_g = Distance function to compute distance of guest i from the individual host

$$C_h = \sum_{j=1}^{j=N_h} [D_h(j)]^{\frac{1}{2}} \qquad \text{(Equation 11)}$$

where

N_h = Number of hosts for which the individual has been a guest within the given time interval and

D_h = Distance function to compute distance of host i from the individual guest

The reasoning behind these equations is that distance has an influence on social capital. Thus social capital is also composed of a measure of how near or far a social network connection is. The measure is more than how many ties an individual has; the proximity of each link is also taken into account. Raw distances in any units, international border crossings, or any combination of these and other human or physical geographical measures define the distance function so long as it is applicable to the social network of interest. The exponent for the distance value (square root for Equations 10 and 11) allows for capital to grow with distance but avoids a doubling in distance producing a doubling of capital. Hence adjusting this exponent affects the impact of distance and may be altered for other networks where distance has either more or less influence on capital. For example in an online gaming network such as Second Life[24] or World of Warcraft[25] physical distance will have little to no influence. Whereas in a network such as a book club, a member's social capital may decrease as a member that lives farther away can be inhibited by

[24] http://www.secondlife.com
[25] http://www.worldofwarcraft.com

physical distance and attend the social network's gatherings less often. For more on this concept of inter-temporal time zone-based accessibility at the global scale see Harvey and Macnab (2000).

There is also the notion of negative social capital in this specific network as users are able to leave negative references as a way to help others make an informed decision. The existence of a negative reference does not simply signify a CouchSurfer to avoid but can have an effect on an individual's overall social capital. Hence one negative reference among three positive is more prominent than one negative among two hundred positive. For this system we propose a value N to subtract from one's overall capital such that

$$N = RV \qquad \text{(Equation 12)}$$

where

R = Number of negative references

V = Penalty per negative reference, maximum value of a connection in the dataset

This value varies between social networks and is to be refined over time. However, a penalty value V as defined here satisfies the goal of a negative reference having an influence on a member with relatively few (< 10) references but not hindering an experienced user with many (> 50) references from future use of the system for what may have been a cultural misunderstanding. This results in the following:

$$C_t = \alpha C_g + \beta C_h - N \qquad \text{(Equation 13)}$$

where

C_t = Total social capital for an individual

α = Weighting factor for having a guest, 1 - β

β = Weighting factor for having a host, 1 - α

C_g = Individual host's social capital based on the quantity of guests an individual has had and the proximity of each guest.

C_h = Individual guest's social capital based on the quantity of hosts an individual has had and the proximity of each host.

N = Value of penalty for negative references, which is a constant based on the quantity of negative references.

Other techniques include a percentage loss system (e.g., 25% penalty for each negative reference) or an increasing value (e.g., 100 for the first, 200 for the second, and so on). This topic of negative social capital remains an item for future research as other large networks such as eBay™ and Amazon™ also feature a system capable of negative feedback.

These formulas provide an objective measure of social capital explicitly taking into consideration an individual's role in the network and the distance between members. This allows for better matching of social network members and an objective measure of experience and trustworthiness based upon previous actions. Additional work including an experiment with expert users is described in the next section. This experiment will examine whether or not CS members agree distance has an influence on social capital within the social network. Also these users will provide insight concerning how playing different roles in the network affect one's ability for social capital.

4.4 Design of Experiment

An experiment was conducted to answer questions pertaining to:

1) How social capital is affected by the distance between network members.
2) The effect of different network roles on social capital.
3) The utility of geovisualizations of location based social network capital.

The survey was designed using the LimeSurvey (http://www.limesurvey.org) open source software installed on the SpaCE (Spatial Cognitive Engineering Lab at UCSB, http://geocog.geog.ucsb.edu/) server. Participants were directed to the website with the questionnaire. See the Appendix of this dissertation for the full survey. The 20 participants have all been members of the CouchSurfing network for at least a year. The survey was designed to be short (less than five questions) and did not include demographic data. The participants were selected based on their experience and their participation in a research group associated with the

CouchSurfing social network. This group has its own messaging forum and its members are those interested in research related to the CS network.

Initially survey participants were given a background and description of the geographic visualization methodology:

> *In these geographical visualizations (geovisualizations) the solid black line is the CouchSurfer's home location. For these space-time paths, time is represented in the z-axis (e.g., April, May, June 2009) and geographic space in the plane containing the world map. The red lines represent a traveling activity where the CouchSurfer individual was a guest at that location and green means the individual hosted a person from the shown location.*

This was followed by two images depicting the experiences of two CouchSurfers: A (Figure 26) and B (Figure 27).

Figure 26. Spatiotemporal geovisualization of CouchSurfer A's hosting and traveling.

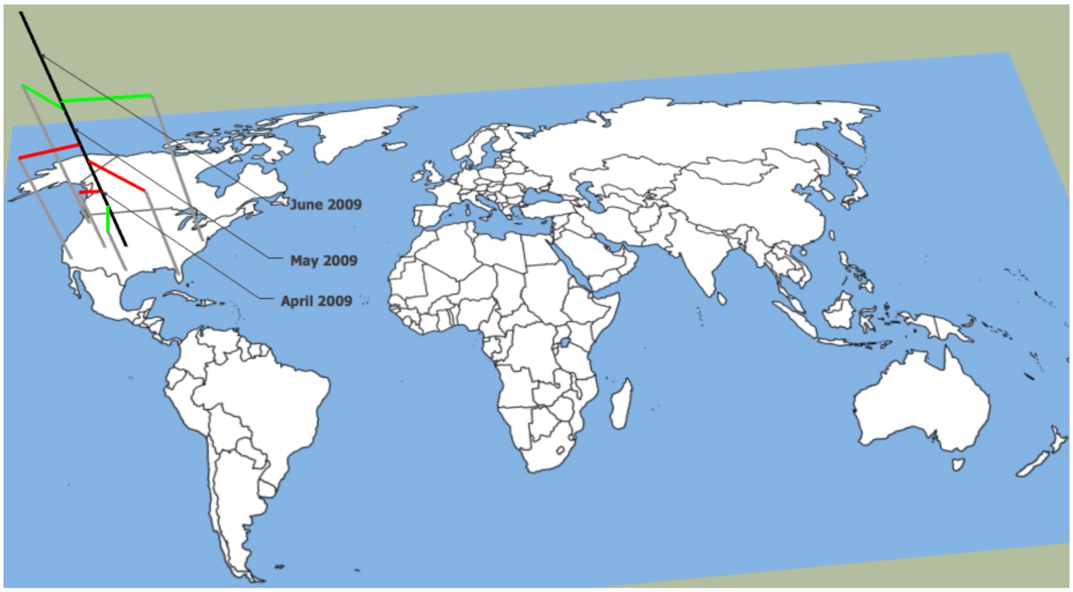
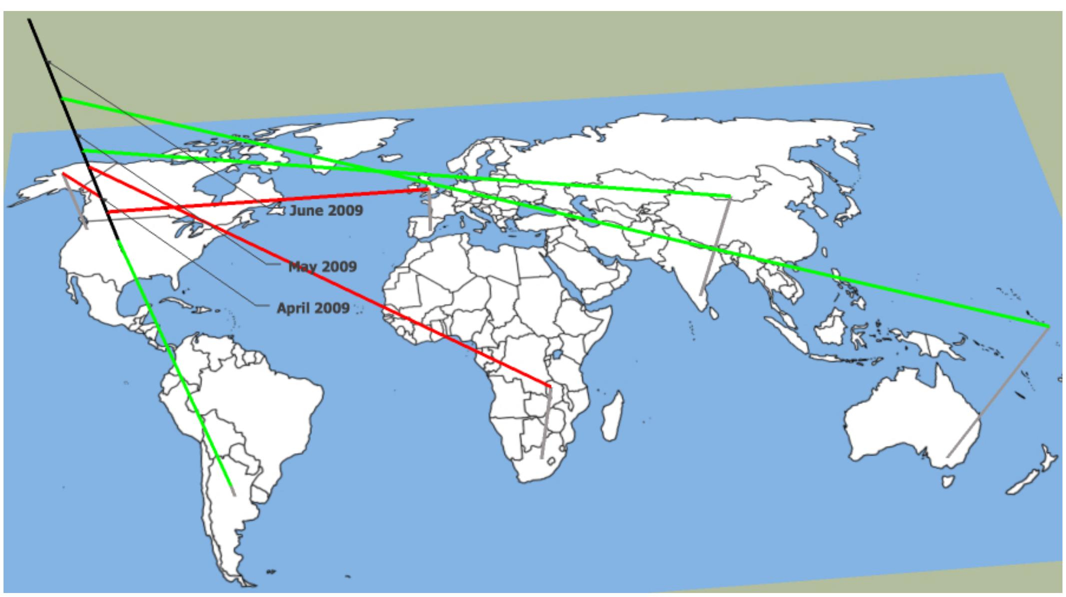


Figure 27. Spatiotemporal geovisualization of CouchSurfer B's hosting and traveling.



Participants were first asked to comment on the utility of the geovisualizations. The next question was stated as such: "Social capital is a measure of concepts such as experience, activity, knowledge, respect, understanding within the social network. Given that they have similarly positive experiences and connections, do you feel CouchSurfer A or B has more social capital based on their activities?" Given that the travels of both CouchSurfer A and B were equally positive, the key difference is in the distance between the locations of events performing as a guest or host. Here we address the following question: Is geographical distance an essential component of social capital or is a raw number of connections a better gauge?

Last, the survey contributors were asked about the significance of being a host vs. being a guest. Specifically they were asked: "Is being a guest or host more important (or are they equally important, i.e., 50/50) for gaining social capital and using the system successfully?" This tackles the issue of whether or not there is a need to stress different network roles and activities by assigning different weights.

4.5 Results

This section presents and discusses the data collected using the experiment described in the previous section. There were 20 responses from experienced (members for more than two years) users that are discussed here. Next participants were asked to comment as to the utility of the figures and stated:

"It's helpful to have visualizations of how members utilize the system to facilitate movement from location to location, for themselves or for others."

"I don't think knowing where they have traveled will really make a difference in whether I would host someone or not."

"Breadth of experience can be a factor in deciding whether to host a surfer. It might also help surfers choose a host."

"If an individual is very active and has a lot of connections, these lines will quickly get muddled and be hard to see."

"They provide a visual representation of places I have not been to before and give me immediate visual feedback about places in the world I have visited. Because of the immediate visual feedback, I can easily scan the world map and look for regional areas I have yet to visit and make an attempt to go there."

Next respondents were asked whether the individuals depicted in Figures 21 and 22 had the same social capital within this network or if one possessed more social capital. No response deemed individual A in Figure 22 to have higher social capital while five (25%) claimed no difference between the two and fifteen (75%) claiming individual B in Figure 23 has more social capital within the network. Those that stated there is no difference highlight other factors in gaining capital such as language and local connections:

"B has moved around more, as A has only surfed & hosted within the USA, so there's a good possibility B would have more experience, knowledge & understanding. B might only be surfing in cities with people who speak English fluently & belong to a certain ethnicity."

"Interacting on a local level, meeting the same people over and over and developing relationships versus going abroad and meeting lots of different people from all over. Everyone will utilize CS differently and that is the beauty of it."

The others demonstrate an influence of geographical breadth on an individual's social capital with short and long responses:

"B: due to breadth of travels to (seemingly) more diverse/different parts of the world."

"B"

"Because we live a globally-connected world, whether in commerce or as a result of traveling for leisure, I would value more the information that CSr B over that of CSr A. When I host CSrs I look for people who are experienced travelers because they have dealt with a variety of people across the world, and as a result, I feel, have a better perspective about it. Likewise, when I am hosted, I like being hosted by individuals who have travelled a lot. Experienced surfers and hosts have valuable information about how to navigate places and dealing with people."

The last question asked individuals to specify the importance of acting under different roles in a social network. Respondents were asked: is being a guest or host more important (or are they equally important) for gaining social capital? The outcome (Figure 24) was some individuals specifying that being in one role is more valuable than the other while some users stated both roles are of equal importance:

"Being a host definitely gains more social capital, as heavy-guests with little hosting experience can sometimes be seen as freeloaders. Thus a visualization in the above style that is clearly more green would indicate a very active host."

"Both are equally important. It is best if a CSer is both a good host and a good guest, though many times a person is not able to participate in both activities. Long-time CSers sometimes frown upon those who would utilize the system, for "freeloading" and only staying with others while not contributing to the larger CS community. There are, of course, alternate ways of reciprocating if one is unable to host."

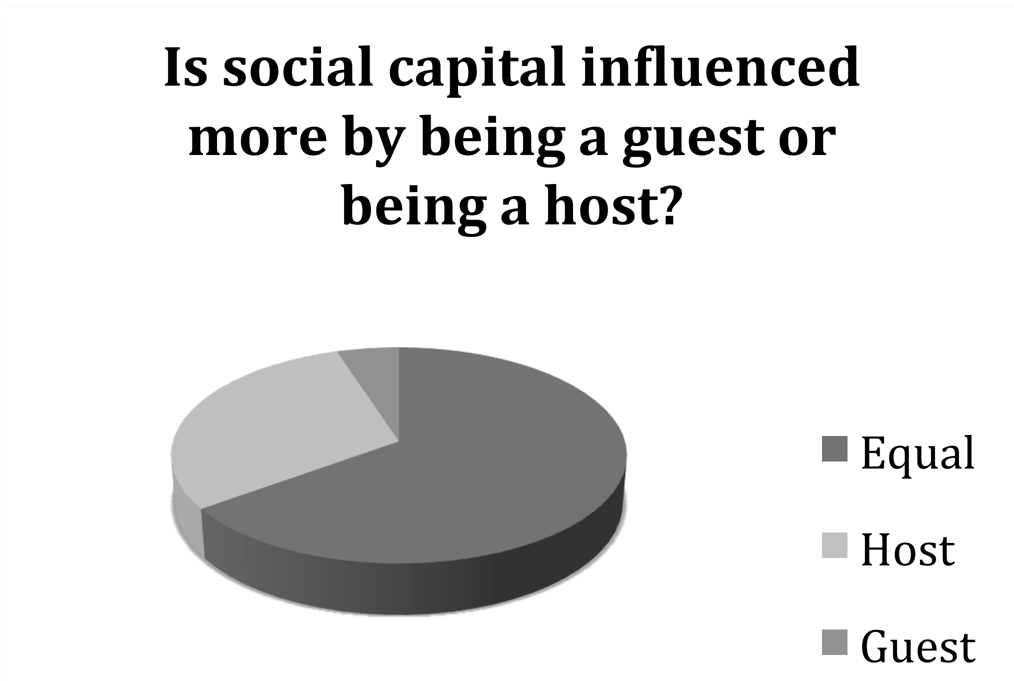
"Both are important, but I think it is slightly more important to have experience being a host. In my experiences people that have been a guest as well as a host will have the most insight on how to be a great guest as well as a host. When first getting started it is more important to gain experience by hosting people."

"Both. It's very good for people to know both sides. I hosted someone who's mostly only been a guest and he was awful. I think if he hosted more he might understand what a gift it is

to host people. Of course, that may just be his personality. There are always things you forget or don't think about when you only do one thing - maybe to make sure you put all your stuff back in your backpack & keep it in a corner when you're out & about (guest) or to make yourself available for questions or recommendations (host). I understand not everyone can do both, but I think it enriches the experience for everyone if you can."

"I learned the most about being a good guest by being a host and knowing what to expect. I definitely prefer to host those who have hosted themselves. I think hosting is a vital part of Couchsurfing, otherwise you aren't paying back your couch karma!"

Figure 28. Results for influence of different roles on social capital.



4.6 Discussion and Development

The results from the first question pertain to the utility of the geovisualizations. The responses point to the utility of the visual methods but also the need for refinement in addition to the importance that temporal and geographical scales have with respect to comprehensibility. For various time scales and durations the methods may need to be altered for optimal utility. One potential solution is to portray a smaller data range, such as the 10 most recent activities, in order to avoid a noisy output. Choice of color can also be revisited in order to ensure proper use of hues. This means assuring that participants are not colorblind and that appropriate color combinations are used so as to not suggest anything by using a certain color. Additionally, an auto-scaling technique may be implemented to change the amount of data displayed based on the zoom level.

The effect of geography on social capital is addressed in the second question in the above data results. While some depicted that a member's spatial distribution of travel was not particularly influential, the majority placed a higher value of social capital with the user in Figure 27. Although there are other factors influencing social capital such as language and local interactions, the prominent influence of geography and space is shown in the results from the previous section. Specifically, with 75% of the respondents stating that the individual depicted in Figure 27 has more social capital this suggests geography has an effect. It also is important to note the different ways distance can be measured as large changes can take place over short distances, e.g., crossing a border. Hence 100km between two populations in the same country with the same language is different from 100km between two populations in different countries that speak different languages. While the methods presented in this chapter do not capture this cultural change it is a topic for future development.

In the CouchSurfing network an individual can act as a guest or host. They may be one role at one time and the other at a different time, but typically an individual is either traveling and being a guest or staying at their residence being a host. Members perform both of these distinct roles, but is one more influential on an individual's social capital? Do those that host more often have a better chance of being hosted when they travel? What about those that act as guest and host equal amounts? This relates to the weights for α and β as given in Equation 9. Are the two roles equally important meaning weights of 0.5 for α and 0.5 for β?

The final question of the survey is directed at this issue concerning the effect of a person's role on social capital. We present the case for different roles in a network having different amounts of influence on social capital. Over half of the respondents made a case that hosting is more valuable for gaining social capital vs. being a guest. Others stated both roles are equally important while one claimed being a guest is more important for social capital. Both roles are necessary for the system to function but these results suggest the weight of α in Equation 9 be greater than 0.5 and the weight of β be less than 0.5.

In order to address this issue of values for the weights of different roles a program was developed in the Python language (http://www.python.org). This code was written such that real-world values of social network members could be used to see the feasibility of implementation. Additionally, different loading factors for different roles can be tested in an object-oriented programming environment. The Python language was chosen due to its speed, ability to run on multiple platforms, and integration with the libkml project. Libkml (http://code.google.com/p/libkml/) allows for creation of KML content in a Python environment.

For the initial version a CouchSurfer class was created with attributes for location and distances from their guests and hosts. An associated method was written to compute social capital values where α and β (Equation 9) are variables that can be easily changed between different executions of the program. The Unified Modeling Language (UML) diagram of the class is given in Figure 29 to show attributes and methods.

Figure 29. UML diagram of class for individual in social network CouchSurfing.

This program was implemented resulting in the core Python code provided in Figure 30. Real-world data from four members in South America were chosen for testing in this chapter. All members had between 11 and 23 connections in the social network and varied in their activity as a guest or host. For example, one individual had been a host for 10 out of their 11 connections (91%) while another had hosted for 3 of their 18 connections (17%). Thus this sample exhibits the changes in overall social capital based on the values for the α and β components in Equation 9.

Figure 30. Python code sample that utilizes libkml for social capital geovisualization.

```
factory = kmldom.KmlFactory_GetFactory()
document = factory.CreateDocument()

for couchsurfer in COUCHSURFERS:
  social_capital_normalized = couchsurfer.social_capital()/normalization_val
  placemark = factory.CreatePlacemark()
  placemark.set_name(couchsurfer.id)
  coordinates = factory.CreateCoordinates()
  coordinates.add_latlngalt(couchsurfer.lat, couchsurfer.lng, 0)
  coordinates.add_latlngalt(couchsurfer.lat, couchsurfer.lng,
                            social_capital_normalized)
  linestring = factory.CreateLineString()
  linestring.set_altitudemode(1)  # Above sea level (absolute)
  linestring.set_tessellate(True)
  linestring.set_coordinates(coordinates)
  placemark.set_geometry(linestring)
  if social_capital_normalized < .333:
    placemark.set_styleurl('#Style1')
  elif social_capital_normalized < .666:
    placemark.set_styleurl('#Style2')
  else:
    placemark.set_styleurl('#Style3')
  document.add_feature(placemark)
kml = factory.CreateKml()
kml.set_feature(document)
print kmldom.SerializePretty(kml)
```

Utilizing Python and libkml, Placemarks were created in KML to show the results of the social capital measures in the Google Earth™ virtual globe environment. The code design allows for quick variation of the weights that influence the importance of being a guest or a host. In Figure 31 the four members are shown in their home location with a line height that represents social capital as presented in this chapter's equations. Using a raw distance measure in kilometers these social capital measures are done for each individual using α and β weights of 0.5. However, the influence of different roles is demonstrated in Figure 32 as the same individuals are used but hosting is given a larger influence on social capital (α = 0.8, β = 0.2) as suggested by the results from Section 4.5. This gives different social capital measures and in particular a higher value to the individual in Uruguay (.73 instead of .37 when $\alpha = \beta = .5$) that has the least total connections but has had the largest amount of guests: 10. Normalizing each individual's value provides a measure between 0 and 1 and makes comparing values more meaningful. Since the maximum possible value for this measure is infinity, a finite value is used for normalization. In this dataset each value is divided by the maximum value of the current sample. This also provides standardization for the visualizations in Figures 31 and 32, for example a value of 0.63 will be the same height in each result. Additionally of note in the figures below is the change in the member with highest value from Buenos Aires to Rosario depending on the weights of α and β.

Figure 31. Implementation of social capital measures with α and β weights at 0.5.

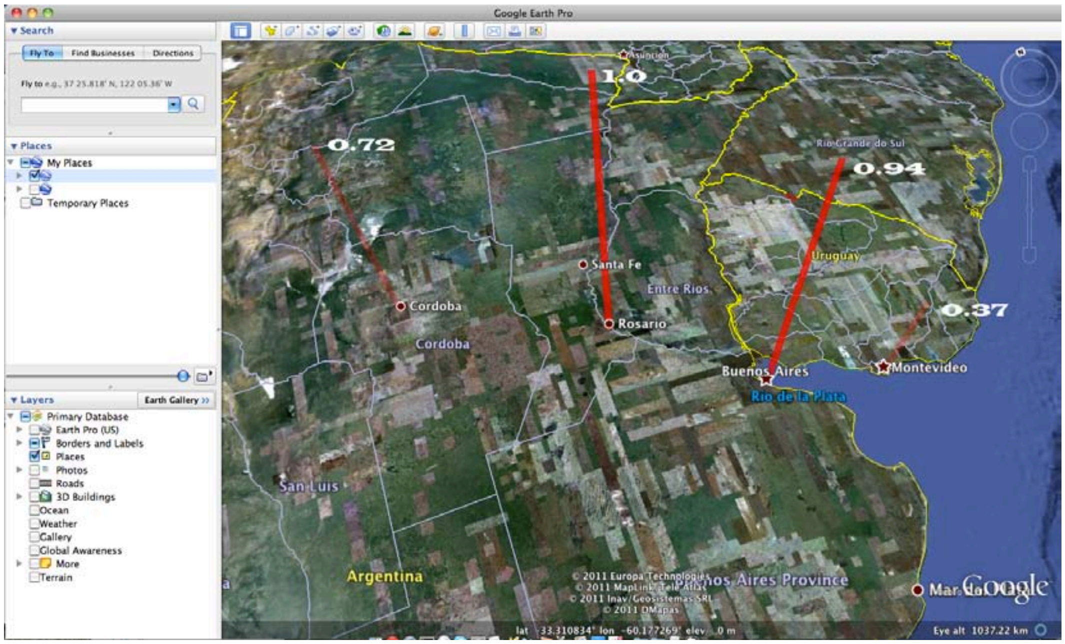
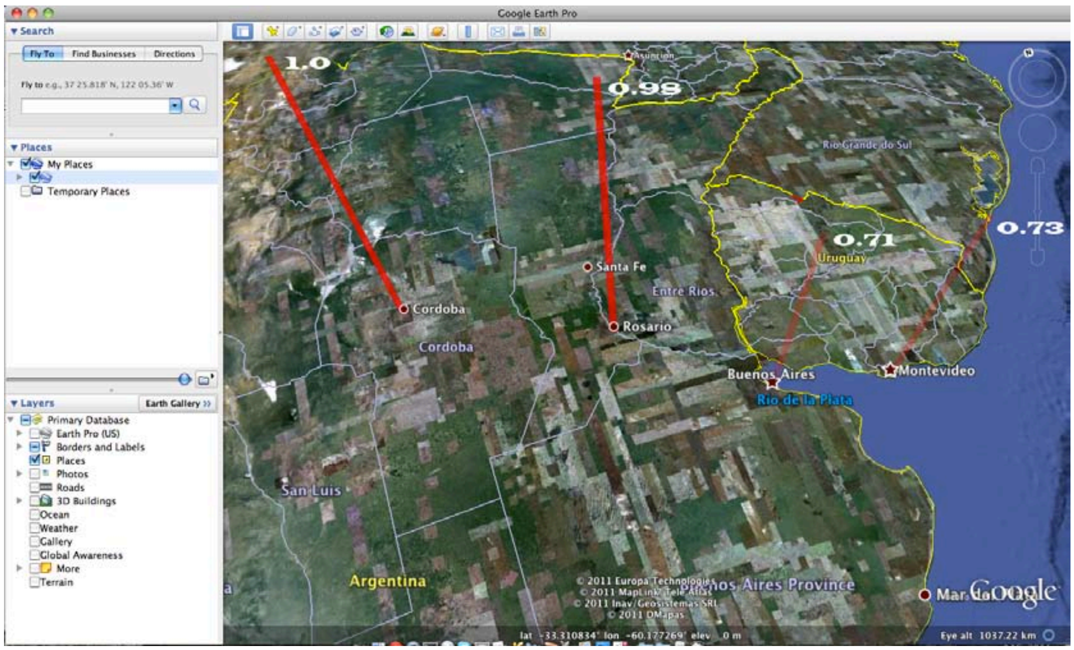


Figure 32. Implementation of social capital measures with α of 0.8 and β of 0.2.



This software will remain useful in further work for this and other social networks as the quantifications of social capital are modified. It provides a solid ground for testing the effectiveness of these social capital measures. These results can then be given to panels of expert users to verify their utility in the social network and revised as needed.

4.7 Conclusions and Future Work

This chapter has proposed new methods for quantifying social capital that explicitly take into account: 1) geographical distance, and 2) individual roles, within a social network. Time-geographic methods were used to represent social capital within a social network that combines physical and virtual spaces. By combining this with a panel of expert users we have confirmed the influence of distance and role in the CouchSurfing network. Specifically, the survey results suggest that ties with individuals at longer distances have more of an influence on overall social capital than connections at shorter distances. We have also addressed the issue of different roles in a social network and their influence on social capital. This was done through the equations presented in Section 4.3 where an individual's social capital was divided into separate components based upon their role.

As a result we have developed software to calculate a member's social capital and implemented it within a virtual globe environment. Individuals in the social network can then use this result in order to determine whether or not they would like to connect with another individual, e.g., be a guest or host. The concept of normalizing or standardizing these resulting values remains to be examined more in future work with surveys of greater magnitude. While standardizing by area or population of an individual's country of residence may not be appropriate, standardizing an individual's total social capital value by an individual's number of connections yields a measure of efficiency. This metric may then be used to demonstrate efficiency and how much capital a member gains on average per connection.

The maps presented to emphasize the effect of distance on social capital are misleading in terms of distances. For instance, the line between the USA and Australia traverses Europe, Africa, and Asia but the shortest distance that an airplane would fly would be over the Pacific Ocean. However, the primary purpose of the maps is to show an individual with connections only in their home country vs. an individual with connections to people outside of their home country. For this reason the maps still present the main purpose to the reader.

Other social networks may be studied in a similar way. This is due to the fact that many contemporary social networks such as YouTube™ and Facebook™ have both physical and virtual components. Additionally, this measure and visualization can also be used for different kinds of social (travel) networks and those that have rating systems such as eBay™ and Amazon™.

This final chapter undertook the fifth and final objective from Section 1.2 involving social capital in social networks where members interact in both virtual and physical settings and perform different roles. The results relate to the rest of this dissertation concerning geovisualization methods and future refinements of the equations and weights are left for future work. Last, the social capital measures can be tied in to the conceptual model where an individual selects a host by allowing a person to specify their preferred level of social capital in a potential host.

5. Overall Conclusions and Future Work

In this dissertation the chapters covered a wide variety of topics in geography. Within human geography this dissertation has discussed spatial behavior and how humans move and interact with others in both physical and virtual environments. The concepts of place and social standing have been presented in the context of individuals using the global CouchSurfing social network. Time geography has been used to examine spatio-temporal constraints as well as present geovisualizations for the multi-network traveler.

The dynamic, multi-network spatial behavior presented in this dissertation leads to new geographies. By making the world seem smaller and more connected, there are advantages and disadvantages. Positive outcomes include awareness of individual differences in addition to sharing and conservation of resources. However, contemporary social networks can lead to a normalization of humans where culture and language differences become smaller and smaller. Connections contribute to an individual's communication needs and thus the most useful languages will survive while others may die at a more rapid rate than if these networks did not exist.

Internet-based social networks and their spatio-temporal nature is what ties together the ideas of physical and virtual presence while bringing about new ideas and future continuations of this work. For instance, the time-geographic representations in the previous chapters can get cluttered with too much information but they are a start for representing spatio-temporally what occurs between two individuals with distinctly different roles in a social network (guest and host in CS). Individuals initially interact and move independently in virtual space then later move together in physical space (e.g., touring around town, staying at the same house). In this dissertation the physical position of individuals was primarily used but the representation of other networks (e.g., data and communication, multiple modes of transportation) is left for future work. Additionally, further investigation can be done with space-time prisms of potential paths an individual can take based on the current status of all the relevant networks. Virtual globe environments afford a readily available platform for combining all of this multi-network information to aid in providing user-centered travel information.

Trip planning with CS has some differences versus other ways of traveling but also shares some similarities. Travelers use the web when trip planning whether or not they are using CS. Sites that offer discount airline tickets are an example of a resource useful to those traveling with or without CS. When traveling with CS it is a social experience where the guest and host spend time together. The guest may be given a tour around town by their local host and given recommendations on places to see and where to eat. With a hotel and other forms of travel it is a more hands-off experience that some may prefer, but with CS there is interaction with a local inhabitant. Hence the social capital measures for CS discussed in this dissertation important for one's ability to use the system. Similarly, references and rankings for hotels and bed & breakfast establishments incorporate a social element where an individual may give a recommendation for where to stay. The difference is with CS a physical recommendation is made in person by a host to a guest versus a virtual review and ranking for a hotel on the web. Last, CS is free and without costs highlighting another difference. A traveler's expectations vary based on how much money they are paying and what they feel they deserve. A traveler using CS has to understand the variability in the residences of their hosts.

CouchSurfing shares a connection with other social networks that have physical and virtual presence requirements such as LinkedIn where a virtual profile may lead to a physical profile. Also, other networks that depend on trust measures for success such as eBay share a connection with CS since positive and negative references affect an individual’s ability to use the network in the future.

Geospatial data mining was presented in Section 3.4 and these general methods given are useful for mining social network data present in all of the sections of this dissertation. For example, one may discover new activities possible at a location due to items present in a member’s profile. In a social network profile an individual may include data about where they have lived or traveled, philanthropy, identity, education, preferences for food or drink, and/or favorite hobbies. By using these data geographic areas can be described based upon real world, up-to-date social information. This aids in confirming what may be common sense knowledge for some individuals, such as the potential to go diving or surfing in Costa Rica, but at the same time it may present new ideas and show growing trends in unexpected locations for activities such as bird, whale, or dolphin watching.

One may also use geospatial data mining to investigate connections between geographic areas by utilizing VGI provided in social network profiles. For instance, in a travel network such as CouchSurfing references are stored which link a guest's location with a host's location. Hence by examining members in a specific geographic area such as New Zealand one will discover places where trips to New Zealand are most popular as well as what destinations are prominent for New Zealanders. This also has implications for travel companies and planners since being aware of current trends in travel behavior help fulfill desires of a trip.

Mining social network data also allows for time geography's space-time paths to be constructed from a social network member's previous travels information. This can alert a social network member of other members who have performed similar travels and may be suitable future travel companions. It also gives insights into sequencing of travel and hosting events in addition to any indications of repetition or regularity in behavior. Overall, data mining with social data provides a great amount of spatio-temporal information. Moreover, with mobile devices and frequent social updates this data source is regularly added to and updated over an extended period of time.

It is possible that the individuals surveyed about social capital in Chapter 4 supported the idea of distance being important because the data was presented in a map format. If other methods were used, such as the presentation of a list with languages spoken or the quantity of photographs the individual has uploaded, those surveyed may have said language or pictures are important. Hence in a future survey one may simply ask for a text description of what is most important in analyzing another person's profile. This would help to narrow down the critical components that are most valuable with regards to social capital for a specific social network. These measures could then be combined with the count-based models in Chapter 3 to examine the characteristics of individuals that have different levels of social capital and also investigate whether or not those with more social capital are indeed most active.

With the samples of data collected for this work, ideas were presented for how people travel with the use of social networks and also what characteristics these individuals have. Sampling error is present as the analysis was done with a sample of CouchSurfers and not the entire population. Sampling error cannot be completely avoided but using a larger sample size can help (Richardson et al. 1995). One cannot generalize for all humans with these results but we can form hypotheses for future testing and also suggest ideas about CouchSurfers for further investigation.

The surveys were posted in public forums where any member of the network was allowed to participate. This means that the individuals who completed the surveys had to make a choice to do so and this introduces self-selection bias and non-response error. Therefore the data are not truly random meaning the results are estimates that should be taken as estimates. Additionally there may be measurement error since the surveys here measured something that is only a proxy to the variable of interest. Some of these problems may be addressed in future projects by collecting larger sample sizes and also samples that have more characteristics more similar to the overall population. Also, in person face-to-face interviews can be used as a different data collection strategy in addition to picking random members and sending them a message asking for participation. More work with different survey techniques and analysis methods will aid in making more conclusions about this travel behavior in the future.

In the dataset collected for this dissertation the count-based model results suggest that the hosting role in this sample is performed more by older individuals. Also in this sample dataset the guest activities were performed more by younger travelers. With respect to gender males were found to be performing more guest and host activities. This could be only for this collected sample where there may have been a lot of males that are very active or perhaps females have more difficulty engaging in the activities due to issues of trust or safety. Hence more surveys in the future with greater depth and containing a larger number of individuals will aid in creating conclusions about who uses social networks for travel, for what purpose(s), and with whom they travel.

Future work may also confirm or question the results of these studies and address in more detail the reasons why different genders and age groups travel different with social networks. For instance, a hosting role is commonly played by females in Western societies but by males in other regions such as the Middle East. CouchSurfing allows for easier interaction between people from different origins helping to promote tolerance for others. Hence global hospitality networks are encouraging a changing of customs in various societies and shaping the future world.

The concept of age affecting a person's use of a social network can be examined in an even broader context as an individual's age can affect their ability to use Information and Communication Technologies (ICT). While there are older individuals using social networks (some over 80 years old in the case of CS), are they impeded by the fact that these technologies can be harder to learn in later years of life? Additionally, at what stages of life are certain types of social networks more useful to specific age groups? Likewise this has consequences for the social capital measures presented in Chapter 4, e.g., how much does age affect a person's capability for meeting other members in an online social network or for learning new languages?

The travel behavior discoveries from this work emerge from using Internet-based techniques. This makes sense for travel with digital social networks since members of the network are familiar with technology. In the future other data collection techniques could be used with hospitality networks such as in-person interviews where a guest or host is interviewed face-to-face. In the second chapter a conceptual model of dynamic travel behavior was presented where each step of the virtual and physical travel process was explicitly linked to any existing ICT layers. Future models should continue to pay attention to ICT influences as they become more ubiquitous every year allowing humans to live their lives in novel ways. The variables studied in section 3.5.2 relate a traveler's guest and host activity levels to their purposes for using the social network. While business and tourism purposes are common reasons for traveling, travel behavior as it relates to language, culture, and study purposes has potential for further investigation. Utilizing social networks such as CS while traveling helps lighten economic and cultural barriers encouraging an increase in long distance travel. To what effect these barriers are being removed makes for exciting future research along with studies examining whether or not hospitality networks make long distance travels more accessible to individuals that would otherwise not travel. Last, the continuing growth of hospitality exchange networks allows for new ways of planning and executing travel plans meaning predictive models will need to adapt in order to increase their utility.

There are additional implications for the general model presented in Section 2.2 such as tweaking the model for different age groups. Similarly, as gender was seen to play a role in the CS activities in this work, whether someone is male or female affects their ability to use ICT and socialize. On another level the citizenship and country that distributes the passport of an individual affects their ability to cross borders. Thus some connections can be more costly (e.g., purchased visas may be required for some countries but not for others) and trip lengths can be adjusted based on how long citizens of one country may stay in another specific country. As mentioned in Section 3.5.4 these types of international restrictions affect a social network member's ability to participate in a specific role such as a guest or host. Hence it is important to note that the general model is a starting point describing only one standard way of planning, but other ways are possible such as planning with the intention of staying with a specific host. Also of note is that at the end of the flowchart an individual may revise their travel plan and restart or end their travel after they arrive at their destination. Therefore these items too will need to be taken into account in future revisions of the general model.

The time geography methods used in this dissertation are useful for more than the historic trips shown in Chapter 2. Using time geography methods for travel planning with hospitality social networks is also an area to examine in the future. In addition, using web data mining in order to search for desired activities has an effect on an individual's constraints and where they are willing to go. Also relevant are the social capital measures that allow a person to do an informed search for who they wish to stay with on their trip. Thus all of these themes tie together and there is potential for an integrated travel planning and support system.

Social networks have been a growing research topic in recent years and this dissertation has contributed to the body of knowledge concerning social networks by investigating how geography, age, and gender affect social network usage. Also presented in this dissertation is a case where both physical and virtual spaces are critical components for successful utilization of a social network. Hence interactions in multiple environments and various network levels are important in the study of social networks. This dissertation also investigated the effects of explicit, distinct roles on a person's reputation in a social network.

The methods utilized throughout this dissertation are diverse yet common for a contemporary geographer. Additional research relating to the impacts of Internet-enhanced social networks on human travel behavior will continue long into the future as technology and mobile devices continue to play a larger and larger role in human lives on a daily basis.

References

1. AXHAUSEN, K. W. 2005. Social networks and travel: Some hypotheses. In: DONAGHY, K. (ed.) Social Aspects of Sustainable Transport: Transatlantic Perspectives. Aldershot: Ashgate.

2. AXHAUSEN, K. W. & GARLING, T. 1992. Activity-based approaches to travel analysis - conceptual frameworks, models, and research problems. Transport Reviews, 12, 323-341.

3. BATTY, M., DESYLLAS, J. & DUXBURY, E. 2003. The discrete dynamics of small-scale spatial events: agent-based models of mobility in carnivals and street parades. International Journal of Geographical Information Sciences, 17, 673-697.

4. BENENSON, I. & TORRENS, P. 2004. Geosimulation - Automata-based modeling of urban phenomena., Chicester, England, Wiley.

5. BIALSKI, P. & BATORSKI, D. 2007. Trust Networks: Analyzing the Structure and Function of Trust. International Network of Social Network Analysis SUNBELT Conference. Corfu, Greece.

6. BORDIEU, P. 1983. The Forms of Capital. In: RICHARDSON, J. G. (ed.) The Handbook for Theory: Research for the Sociology of Education. New York: Greenwood Press.

7. CARRASCO, J. A., HOGAN, B., WELLMAN, B. & MILLER, E. J. 2008. Collecting social network data to study social activity-travel behavior: an egocentric approach. Environment and Planning B: Planning and Design, 35, 961-980.

8. CARRASCO, J. A., MILLER, E. J. & WELLMAN, B. 2006. Spatial and Social Networks: The Case of Travel for Social Activities. 11th International Conference on Travel Behaviour Research. Kyoto, Japan.

9. CHANG, G., HEALEY, M. J., MCHUGH, J. A. M. & WANG, J. T. L. 2001. Mining the World Wide Web An Information Search Approach, Norwell, MA, Kluwer.

10. COLEMAN, J. S. 1988. Social Capital in the Creation of Human Capital. American Journal of Sociology, 94, 95-120.

11. CONGDON, P. 2010. A multiple indicatory, multiple cause method for representing social capital with an application to psychological distress. Journal of Geographical Systems, 12, 1-12.

12. COUCLELIS, H. 1992. Location, Place, Region, and Space. In: ABLER, R., MARCUS, M. & OLSON, J. (eds.) Geography's Inner Worlds. New Brunswick, New Jersey: Rutgers University Press.

13. COUCLELIS, H. 2009. Rethinking time geography in the information age. Environment and Planning A, 41, 1556-1575.

14. CRESSWELL, T. 1996. In Place/Out of Place, Minneapolis, University of Minnesota Press.

15. DIJST, M. 2006. ICT and Social Networks: Towards a situational perspective on the interaction between corporeal and connected presence. 11th International Conference on Travel Behaviour Research (IATBR). Kyoto, Japan.

16. DUGUNDJI, E., PAEZ, A. & ARENTZE, T. 2008. Social Networks, Choices, Mobility and Travel. Environment and Planning B.

17. ELLISON, N. & BOYD, D. 2007. Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication, 13, article 11.

18. ELLISON, N. B., STEINFELD, C. & LAMPE, C. 2007. The benefits of Facebook "Friends:" Social capital and college students' use of online social network sites. Journal of Computer-Mediated Communication, 12, article 1.

19. FINE, B. 2001. Social Capital versus Social Theory: Political Economy and Social Science at the Turn of the Millenium,London, Routledge.

20. FRANK, A., BITTNER, S. & RAUBAL, M. 2001. Spatial and Cognitive Simulation with Multi-agent Systems. Spatial Information Theory - Foundations of Geographic Information Science, COSIT 2001. Morro Bay, CA: Springer.

21. FRANK, A. U. 2003. Pragmatic Information Content - How to Measure the Information in a Route Description. In:GOODCHILD, M. F., DUCKHAM, M. & WORBOYS, M. (eds.) Perspectives on Geographic Information Science. Lond: Taylor & Francis.

22. FRÄNDBERG, L. 2008. Paths in transnational time-space: representing mobility biographies of young Swedes. Geografiska Annaler: Series B, Human Geography, 90, 17-28.

23. FULKERSON, G. M. & THOMPSON, G. H. 2008. The Evolution of a Contested Concept: A Meta-Analysis of Social Capital Definitions and Trends (1988-2006). Sociological Inquiry, 78, 536-577.

24. GAJJALA, R. 2007. Shifting frames: Race, ethnicity, and intercultural communication in online social networking and virtual work. In: HINNER, M. B. (ed.) The Role of Communication in Business Transactions and Relationships. New York: Peter Lang.

25. GOLLEDGE, R. G. & STIMSON, R. J. 1997. Spatial behavior : a geographic perspective, New York, Guilford Press.

26. GOODCHILD, M. F. 2007a. Citizens as sensors: the world of volunteered geography. GeoJournal, 69, 211-221.

27. GOODCHILD, M. F. 2007b. Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. International Journal of Spatial Data Infrastructures Research, 2, 24-32.

28. GOULIAS, K. G. 2003. Transportation systems planning: methods and applications, Boca Raton, Fla., CRC Press.

29. GOULIAS, K. G. & KIM, T. G. 2005. An Analysis of Activity-Type Classification and Issues Related to the With Whom and For Whom Questions of an Activity Diary". In TIMERMANS, H. J. P. Progress in Activity-Based Analysis, Elsevier, Oxford, England.

30. GRANOVETTER, M. S. 1973. The strength of weak ties. American Journal of Sociology, 78, 1360-1380.

31. GROSSNER, K., CLARKE, K. C. & GOODCHILD, M. F. 2008. Defining a Digital Earth System. Transactions in GIS, 12, 145-160.

32. GÄRLING, T., SÄISÄ, J., BOOK, A. & LINDBERG, E. 1986. The Spatiotemporal Sequencing of Everyday Activities in the Large-Scale Environment. Journal of Environmental Psychology, 6, 261 - 280.

33. HAN, J. & KAMBER, M. 2006. Data Mining: Concepts and Techniques, Second Edition, San Francisco, Morgan Kaufmann.

34. HARIHARAN, R. & HORNSBY, K. 2000. Modeling Intersections of Geospatial Lifelines. First International Conference on Geographic Information Science, GIScience 2000. Savannah, Georgia.

35. HARVEY, D. 1989. The condition of postmodernity : an enquiry into the origins of cultural change, Oxford, England ; New York, NY, Blackwell.

36. HECHT, B. & RAUBAL, M. 2008. GeoSR: Geographically Explore Semantic Relations in World Knowledge. In: BERNARD, L., FRIJS-CHRISTENSEN, C. & PUNDT, H., eds. The European Information Society - Taking Geoinformation One Step Further (Proceedings of the 11th AGILE International Conference on GIScience 2008), Girona, Spain. Springer, Berlin, 95-113.

37. HENDRICKS, M., EGENHOFER, M. & HORNSBY, K. 2003. Structuring a Wayfinder's Dynamic Space-Time Environment. In:KUHN, W., WORBOYS, M. & TIMPF, S. (eds.) Spatial Information Theory - Foundations of Geographic Information Science, International Conference, COSIT 2003. Kartause Ittingen, Switzerland: Springer.

38. HERRING, S. C., PAOLILLO, J. C., RAMOS-VIELBA, I., KOUPER, I., WRIGHT, E., STOERGER, S., SCHEIDT, L. A. & CLARK, B. 2007. Language networks on LiveJournal. Fortieth Hawai'i International Conference on System Sciences. Los Alamitos, CA.

39. HJORTH, L. & KIM, H. 2005. Being there and being here: Gendered customising of mobile 3G practices through a case study in Seoul. Convergence, 11, 49-55.

40. HÄGERSTRAND, T. 1970. What about people in regional science? Papers of the Regional Science Association, 24, 7-21.

41. HARVEY, A. & MACNAB, P. A. 2000. Who's up? Global interpersonal temporal accessibility. In: JANELLE, D. G. & HODGE, D. C. (eds.) Information, place and cyberspace: Issues in accessibility. Berlin: Springer.

42. JANELLE, D. 2004. Impact of Information Technologies. In: HANSON, S. & G, G. (eds.) The Geography of Urban Transportation. 3rd ed. New York: Guilford Press.

43. JONES, C. B., PURVES, R. S., CLOUGH, P. D. & JOHO, H. 2008. Modelling Vague Places with Knowledge from the Web.International Journal of Geographical Information Sciences, in press.

44. KANAZAWA, S. & SAVAGE, J. 2009. An evolutionary psychological perspective on social capital. Journal of Economic Psychology, 30, 873-883.

45. KARLAFTIS, M. G. & TARKO, A. P. 1998. Heterogeneity considerations in accident modeling. Accident Analysis and Prevention, 30, 425-433.

46. KEßLER, C., RINNER, C. & RAUBAL, M. 2005. An Argumentation Map Prototype to Support Decision-Making in Spatial Planning. In: TOPPEN, F. & PAINHO, M. (eds.) AGILE 2005 - 8th Conference on Geographic Information Science. Lisboa, Portugal.

47. KNOKE, D. & YANG, S. 2008. Social Network Analysis, California, Sage Publications, Inc.

48. KWAN, M.-P. 2000. Analysis of human spatial behavior in a GIS environment: Recent developments and future prospects.Journal of Geographical Systems, 2, 85-90.

49. KWAN, M.-P. 2004. GIS Methods in Time-Geographic Research: Geocomputation and Geovisualization of Human Activity Patterns. Geografiska Annaler B, 86, 267-280.

50. KWAN, M.-P., DIJST, M. & SCHWANEN, T. 2007. The interaction between ICT and human activity-travel behavior.Transportation Research A: Policy and Practice, 41, 121-124.

51. LARSEN, J., AXHAUSEN, K. W. & URRY, J. 2006a. Geographies of Social Networks: Meetings, Travel and Communications. Mobilities, 1, 261 - 283.

52. LARSEN, J., URRY, J. & AXHAUSEN, K. W. 2006b. Mobilities, networks, geographies, Aldershot, England ; Burlington, VT, Ashgate.

53. LAUTERBACH, D., TRUONG, H., SHAH, T. & ADAMIC, L. 2009. Surfing a web of trust: Reputation and Reciprocity on CouchSurfing.com. IEEE Social Computing 2009. Vancouver, British Columbia, Canada.

54. LENNTORP, B. 1976. Paths in Space-Time Environments: A Time-Geographic Study of the Movement Possibilities of Individuals. Lund Studies in Geography, B.

55. LOURY, G. 1977. A Dynamic Theory of Racial Income Differences. In: WALLACE, P. A. & LAMOND, A. M. (eds.) Women, Minorities, and Employment Discrimination. Lexington, MA: Lexington Books.

56. MA, J. & GOULIAS, K. G. 1999. Application of poisson regression models to activity frequency analysis and prediction.Transportation Research Record 1676, 86-94.

57. MAECHLING, P., CHALUPSKY, H., DOUGHERTY, M., DEELMAN, E., GIL, Y., GULLAPALLI, S., GUPTA, V., KESSELMAN, C., KIM, J., MEHTA, G., MENDENHALL, B., RUSS, T., SINGH, G., SPRARAGEN, M., STAPLES, G. & VAHI, K. 2005. Simplifying construction of complex workflows for non-expert users of the Southern California Earthquake Center Community Modeling Environment. SIGMOD 2005.

58. MILLER, H. J. 1991. Modeling accessibility using space-time prism concepts within geographical information systems.International Journal of Geographical Information Systems, 5, 287-301.

59. MILLER, H. J. 1999. Measuring space-time accessibility benefits within transportation networks: Basic theory and computational methods. Geographical Analysis, 31, 187-212.

60. MILLER, H. J. 2005a. A measurement theory for time geography. Geographical Analysis, 37, 17-45.

61. MILLER, H. J. 2005b. What about people in geographic information science? In: FISHER, P. & UNWIN, D. (eds.) Re-Presenting Geographical Information Systems. John Wiley.

62. MILLER, H. J. 2007. Geographic data mining and knowledge discovery. In: WILSON, J. P. & FOTHERINGHAM, A. S. (eds.)Handbook of Geographic Information Science. Blackwell.

63. MOLZ, J. G. 2007. Cosmopolitans on the Couch: Mobile Hospitality and the Internet. In: MOLZ, J. G. & GIBSON, S. (eds.)Mobilizing Hospitality: The Ethics of Social Relations in a Mobile World. Aldershot: Ashgate.

64. MONTELLO, D. 2001. Scale, in geography. In: SMELSER, N. & BALTES, P. (eds.) International Encyclopedia of the Social & Behavioral Sciences. Oxford: Pergamon Press.

65. NYLAND, R. & NEAR, C. 2007. Jesus is my friend: Religiosity as a mediating factor in Internet social networking use.AEJMC Midwinter Conference. Reno, NV.

66. PORTES, A. & LANDOLT, P. 2000. Social Capital: Promise and Pitfalls of Its Role in Development. Journal of Latin American Studies, 32, 529-47.

67. PRISKIN, J. & SPRAKEL, J. 2008. CouchSurfing: The question for an authentic tourist experience. Téoros, Revue de Recherche en Tourisme, 27, 68-71.

68. PULTAR, E., COVA, T. J., YUAN, M. & GOODCHILD, M. F. 2010. EDGIS: a dynamic GIS based on space time points.International Journal of Geographical Information Science, 24, 329-346.

69. PULTAR, E. & RAUBAL, M. 2009a. A Case for Space: Physical and Virtual Location Requirements in the CouchSurfing Social Network. In: XIE, X. Z. A. X., ed. International Workshop on Location Based Social Networks (LBSN'09) at 17th ACM International Symposium on Geographic Information Systems, SIGSPATIAL ACM GIS 2009, Seattle, Washington, USA. 88-91.

70. PULTAR, E. & RAUBAL, M. 2009b. Progressive Tourism: Integrating Social, Transportation, and Data Networks. In:SHARDA, N. (ed.) Tourism Informatics: Visual Travel Recommender Systems, Social Communities, and User Interface Design. Hershey, PA: IGI Global.

71. PULTAR, E., RAUBAL, M., COVA, T. & GOODCHILD, M. 2009. Dynamic GIS Case Studies: Wildfire Evacuation and Volunteered Geographic Information. Transactions in GIS 13(s1), 85-104.

72. PULTAR, E., RAUBAL, M. & GOODCHILD, M. 2008. GEDMWA: Geospatial Exploratory Data Mining Web Agent. In: SAMET, H., SHAHABI, C. & WOLFSON, O. (eds.) 16th ACM International Symposium on Geographic Information Systems, SIGSPATIAL ACM GIS 2008. Irvine, California.

73. PUTNAM, R. 1995. Bowling Alone: America's Declining Social Capital. Journal of Democracy, 6, 65-78.

74. PUTNAM, R. 2000. Bowling alone: The collapse and revival of American community, New York, Simon & Schuster.

75. PUTNAM, R., LEONARDI, R. & NANETTI, R. 1993. Making Democracy Work: Civic Traditions in Modern Italy, Princeton, NJ, Princeton University Press.

76. QUAN-HAASE, A. & WELLMAN, B. 2004. How does Internet affect social capital? In: HUYSMAN, M. & WULF, V. (eds.)Social Capital and Information Technology. Cambridge, MA: MIT Press.

77. RAUBAL, M. 2011. Cogito ergo mobilis sum: The impact of Location-Based Services on our mobile lives. In: NYERGES, T., COUCLELIS, H. & MCMASTER, R. (eds.) Handbook of GIS and Society Research. Los Angeles, London: Sage Publications.

78. RAUBAL, M., MILLER, H. J. & BRIDWELL, S. 2004. User-Centred Time Geography for Location-Based Services.Geografiska Annaler: Series B, Human Geography, 86, 245-265.

79. RAUBAL, M., WINTER, S., TESSMANN, S. & GAISBAUER, C. 2007. Time geography for ad-hoc shared-ride trip planning in mobile geosensor networks. ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING, 62, 366-381.

80. REN, F. & KWAN, M.-P. 2007. Geovisualization of Human Hybrid Activity-Travel Patterns. Transactions in GIS, 11, 721-744.

81. RICHARDSON, A., AMPT, E. & MEYBURG, A. 1995. Survey Methods for Transport Planning, Melbourne, Eucalyptus Press.

82. ROOT, G. S. & RECKER, W. W. 1983. Towards a Dynamic Model of Individual Activity Pattern Formulation. In:CARPENTER, S. & JONES, P. M. (eds.) Recent Advances in Travel Demand Analysis. Aldershot: Gower.

83. RUSSELL, S. J. & NORVIG, P. 2002. Artificial Intelligence: A Modern Approach, Prentice Hall.

84. SCHAFFT, K. A. & BROWN, D. L. 2004. Social Capital, Social Networks, and Social Power. Social Epistemology, 7, 329-42.

85. SEIDEL, I., GÄRTNER, M., PÖTTLER, M., BERGER, H., DITTENBACH, M. & MERKL, D. 2009. Itchy Feet: A 3D e-Tourism Environment. In: SHARDA, N. (ed.) Tourism Informatics: Visual Travel Recommender Systems, Social Communities, and User Interface Design. Hershey, PA: IGI Global.

86. SENGUPTA, R. & SIEBER, R. 2007. Geospatial Agents, Agents Everywhere... Transactions in GIS, 11, 483-506.

87. SHARDA, N. 2009. Tourism Informatics: Visual Traveler Recommender Systems, Social Communities, and User Interface Design. Hershey, PA: IGI Global.

88. SHI, W., FISHER, P. F. & GOODCHILD, M. F. E. 2002. Spatial Data Quality, New York, Taylor & Francis.

89. WASHINGTON, S., KARLAFTIS, M. & MANNERING, F. 2003. Statistical and Econometric Methods for Transportation Data Analysis, Washington, DC, Chapman & Hall.

90. WILSON, C., BOE, B., SALA, A., PUTTASWAMY, K. P. N. & ZHAO, B. Y. 2009. User Interactions in Social Networks and their Implications. ACM EuroSys 2009.

91. WINTER, S. & ROBINS, G. 2007. International Workshop on Social Space and Geographic Space. COSIT 2007, Melbourne Business School, Mt Eliza, Victoria, Australia.

92. WOOLDRIDGE, M. 1999. Intelligent Agents. In: WEISS, G. (ed.) Multiagent Systems - A Modern Approach to Distributed Artificial Intelligence. Cambridge: MIT Press.

93. WU, Y.-H. & MILLER, H. 2001. Computational tools for measuring space-time accessibility within dynamic flow transportation networks. Journal of Transportation and Statistics, 4, 1-14.

94. YU, H. & SHAW, S. L. 2008. Exploring potential human activities in physical and virtual spaces: a spatio-temporal GIS approach. International Journal of Geographical Information Science, 22, 409-430.

95. ZHOU, X. & XING, X. 2009. Proceedings of the 2009 International Workshop on Location Based Social Networks. LBSN 2009 at ACM GIS, Seattle, Washington.

Appendix

CouchSurfing Behavior Survey used in Chapter 3.

1 In which city and country do you live?

Please write your answer here:

2 How many CouchSurfing hosts have you stayed with in the last/previous 0-90 days?

Please write your answer here:

You may log in to http://www.couchsurfing.org and look at your references and friends list to recount how many people and when.

3 How many CouchSurfing hosts have you stayed with in the last/previous 91-180 days? (Enter 99999 if you were not a CouchSurfing member at this time)

Please write your answer here:

You may log in to http://www.couchsurfing.org and look at your references and friends list to recount how many people and when.

4 How many CouchSurfing hosts have you stayed with in the last/previous 181-270 days? (Enter 99999 if you were not a CouchSurfing member at this time)

Please write your answer here:

You may log in to http://www.couchsurfing.org and look at your references and friends list to recount how many people and when.

5 How many CouchSurfing hosts have you stayed with in the last/previous 271-360 days? (Enter 99999 if you were not a CouchSurfing member at this time)

Please write your answer here:

You may log in to http://www.couchsurfing.org and look at your references and friends list to recount how many people and when.

6 What inhibits you from traveling more often with CouchSurfing?

Please write your answer here:

7 How many people have you hosted through CouchSurfing in the last/previous 0-90 days?

Please write your answer here:

You may log in to http://www.couchsurfing.org and look at your references and friends list to recount how many people and when.

8 How many people have you hosted through CouchSurfing in the last/previous 91-180 days? (Enter 99999 if you were not a CouchSurfing member at this time)

Please write your answer here:

You may log in to http://www.couchsurfing.org and look at your references and friends list to recount how many people and when.

9 How many people have you hosted through CouchSurfing in the last/previous 181-270 days? (Enter 99999 if you were not a CouchSurfing member at this time)

Please write your answer here:

You may log in to http://www.couchsurfing.org and look at your references and friends list to recount how many people and when.

10 How many people have you hosted through CouchSurfing in the last/previous 271-360 days? (Enter 99999 if you were not a CouchSurfing member at this time)

Please write your answer here:

You may log in to http://www.couchsurfing.org and look at your references and friends list to recount how many people and when.

11 What inhibits you from hosting more often with CouchSurfing?

Please write your answer here:

12 What is your age?

Please write your answer here:

13 What is your gender?

Please choose **only one** of the following:

- ○ Female
- ○ Male

14 Describe your travel behavior before joining the CouchSurfing project.

Please write your answer here:

15 Describe your travel behavior after joining the CouchSurfing project.

Please write your answer here:

16 In which city and country did you grow up?

Please write your answer here:

17 Specify the purpose(s) of your trips with CS.

Please choose all that apply and provide a comment:

- ☐ Business
- ☐ Tourism
- ☐ Study
- ☐ Culture
- ☐ Language
- ☐ Other

18 What is the primary, #1 purpose of your CouchSurfing trips?

Please write your answer here:

19 What is the composition of your traveling party (who do you travel with?) for the majority of your CouchSurfing trips?

Please write your answer here:

20 What is your occupation?

Please write your answer here:

21 Write any comments you have.

Please write your answer here:

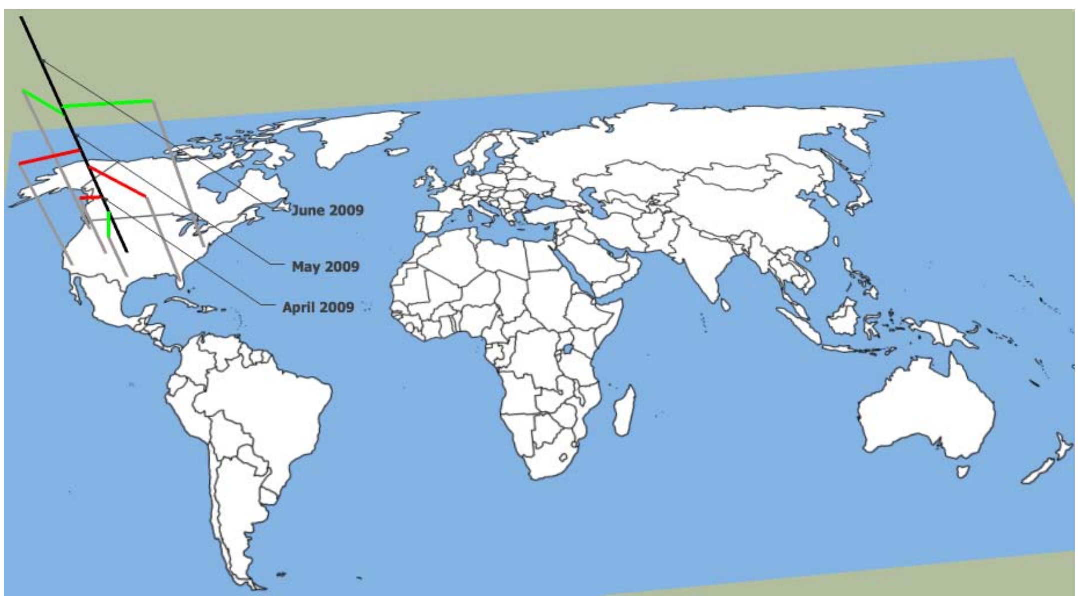
22 If it is ok to contact you for further details in the future please enter your e-mail address.

Please write your answer here:

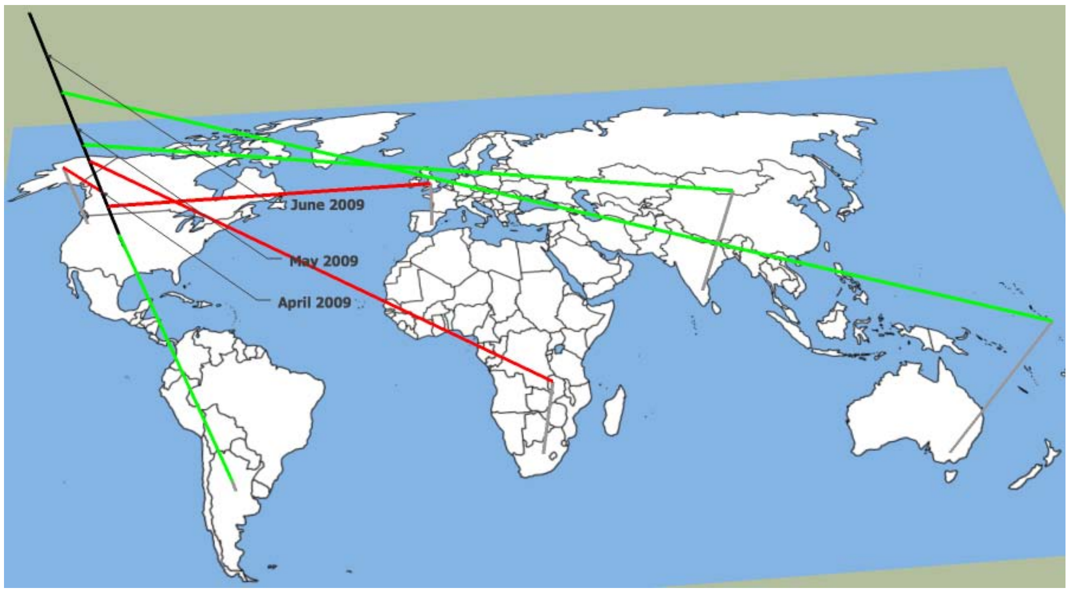
CouchSurfing Social Capital Survey used in Chapter 4.

1 *In these geographical visualizations (geovisualizations) the solid black line is the CouchSurfer's home location. For these space-time paths, time is represented in the z-axis (e.g., April, May, June 2009) and geographic space in the plane containing the world map. The red lines represent a traveling activity where the CouchSurfer individual was a guest at that location and green means the individual hosted a person from the shown location.*

CouchSurfer A:



CouchSurfer B:



Are these geovisualizations of a CouchSurfer's activity helpful?

Please choose only one of the following:

- ◯ **Yes**
- ◯ **No**

2 If so, in what ways could they enhance the CouchSurfing system?

Please write your answer here:

3 Social capital is a measure of concepts such as experience, activity, knowledge, respect, understanding within the social network. Given that they have similarly positive experiences and connections, do you feel CouchSurfer A or B has more social capital based on their activities? *

Please write your answer here:

4 Is being a guest or host more important (or are they equally important, i.e., 50/50) for gaining social capital and using the system successfully?

Please write your answer here:

5 For what amount of time or for how long have you been a member of CouchSurfing? Thanks for your time, feel free to leave any comments and your e-mail address if you may be contacted regarding this survey in the future.

Please write your answer here: