Determining Social Constraints in Time Geography through Online Social Networking

Grant McKenzie[a]

[a] *Department of Geography, University of California, Santa Barbara*
(grant.mckenzie@geog.ucsb.edu)

Advisor: Prof. Dr. Martin Raubal
Institute of Cartography and Geoinformation, ETH Zurich, Switzerland

Keywords. Social Networking, Time Geography, Social Constraints

Introduction

In recent years, research in the area of online social networking (OSN) has grown dramatically. Online applications such as *facebook* have exceeded 500 million users [1] and the micro-blogging service *twitter* recently reported over 140 million tweets a day [2]. The stated purpose of these applications is to allow users to share their likes, interests and information pertaining to their day-to-day activities with friends and colleagues. Recent advances in mobile technology have permitted these social applications to push in to the realm of Location Based Services allowing an increasing number of OSN users to add spatial context to their social interaction. Online social networking is moving from location-static to location-dynamic social posts. On the data side, these social networking applications are amassing petabytes of user-contributed content, content from which spatio-temporal locations may be inferred.

The field of spatio-temporal geography can gain a lot from exploration and analysis of this rich new source of social communication. Analysis of OSN may provide social details ranging from gender and age to shopping trends or common modes of transportation. Further analysis may also offer insight into socially driven movement and travel behavior. For example, one's interest in water sports may lead them to choose a less direct path along the coast (as opposed to the faster, inland route) when traveling between cities.

This social dimension adds new constraints, providing additional context to activities currently only measured through spatial and temporal dimensions. In terms of applications, the results of analysis in this area may decrease the need for user input in a location-based service while providing more pertinent information to the end user. An example of this might involve suggesting local concerts based on the history of songs played on one's mobile device.

1. Background and Related Work

Research in the area of time geography has perpetuated the notion that our daily lives are made up of activities that require space and time [3][4]. The motivation behind these activities is driven by certain capabilities and affordances. These affordances exist in the physical, social-institutional, and mental realms [5]. Given knowledge of these affordances it is possible to predict a user's movement with a certain degree of probability. Physical constraints such as a road network decrease the uncertainty of locating a user in space and time [6][7]. With each additional constraint, location uncertainty diminishes. Mental affordances also play a role in restricting a user's space-time prism by weighing options and making decisions based on the physical and social-institutional limitations of the environment [5]. These social-institutional limitations may be anything from *hours of operation* restrictions to the social graces of wearing black to a funeral.

In fact, Hägerstrand's initial discussions on time geography placed social interaction in a pivotal role. Given the increase in Information and Communication Technologies (ICT), his concept of "coupling constraints," has expanded outside the range of "voice and eye" to encompass new forms of communication such as online social collaboration. The idea of *social capital* is of interest to this research as well. Social communication and the network that supports it, contain measures of value that are apparent in the relationships and information produced across the network. With the increase in ICT, the study of social capital has made its way online [8].

Related research is also being conducted in the area of transportation. Transportation and travel behavior has looked at real-world social network interaction and how social relationships influence travel behavior [9]. An individual's movement patterns and spatial behavior give insight into his/her intentions [10]. In contrast, the issue of social exclusion has also been explored, showing that movement in time and space is restricted to disadvantaged groups [11]. This proposed research also extends the Social Positioning Method. The method links individuals' mobile phone movements with their basic social data (provided via questionnaire) [12]. The finer (though varying) resolution of OSN data allows for more accurate prediction of a user's spatio-temporal movement. Each social interaction, post and status update provides increased accuracy.

Collaboration between social networking services and online search engines has already begun to focus on providing socially relevant search results (and advertisements). Search engines such as Bing® and Google® access information from a user's social graph to provide search results tailored specifically to her likes, socio-demographic stature and common interests with friends [13]. Research in the area of spatio-temporal geography can take this a step further by linking this socially relevant information to a user's movement in space and time.

2. Research Questions

The primary purpose of this research is to explore user contributed social data available through current OSN and develop a model that integrates social constraints with spatio-temporal information. This model will be used to further increase the probability of locating a specific user or group in time and space.

Which social factors lead individuals to make the decisions they do? How do these factors rank in terms of influence on location and travel behavior? Which types of activities in space and time are strongly constrained by social interaction? These questions pertain to the overall study of how existing social networking data can provide additional, context relevant information to a spatio-temporal probability model.

Other aspects of this research will involve visualization of spatio-temporal change within a social network and comparison between real and online social networking. What types of social information are users more likely to share online? To what degree (and in which aspects) does an online interaction reflect a physical interaction?

3. Data and Approach

Given the early stage of this research, the approach and methods are merely *proposed*. Further investigation into social data mining techniques must be explored as well as gaining a better understanding of the inherent limitations of accessing private data on OSN.

In order to leverage the relation-based structure of a social network, random as well as relationship-based sampling is presented as an appropriate method to gathering data [14][15]. A user's interests and likes are often mirrored by the people with whom they associate. In exploring an individual's motivation for making a decision, the motivations of their relations must also be examined.

Initial studies of social data mining techniques have looked at the interlinking structure and dynamics within this structure. Relationships (in the real world as well as online) are constantly changing, growing stronger and weaker. 'Crowds' can develop around a specific topic within seconds and disperse moments later [16]. Data mining within this area has been approached by a number of techniques including probabilistic latent semantic indexing [17][18], latent dirichlet allocation [18][20], unsupervised clustering [21] and other forms of probability estimation based on social content [22] have been examined.

One possible approach to integrating social data in to a spatio-temporal model would be to examine the frequency of topics mentioned (via social interactions) across a specified time frame (e.g. 24 hour period, one week). Temporal patterns may appear in the types of social interaction and the distribution of topics (e.g. Bars in the evenings, coffee shops in the mornings).

The importance of online relationships must not be undervalued. Weights should be applied to nodes (friends/followers) on the social graph indicating strength of relationship. Information provided by close friends should be valued higher than those of the mere acquaintances. These weights and distribution of topics over time would provide the input to a probability model.

4. Expected Outcomes

The intended outcome of this research is a model capable of ingesting socially relevant context data and produces a number of socially constrained probability values that can be used to infer a user's location. It is expected that analysis of an individual's online social networking data will produce significant probability estimates (predictions) of an individual's location in space and time. It is anticipated that these models will supply more accurate estimates than if OSN data was not exploited.

Figure 1 shows the probability of finding a user at a specific location in space and time through analysis of her social network. From her basic profile information we have access to her work and home locations. Her profile also lists *Starbucks®* (coffee shop) and *Soundgarden* (musical act) as two of her likes/interests. One might assume that she would travel from home to work sometime between 08:00 and 10:00, possibly stopping at *Starbucks* on the way. The graph shows a shift from a high probability at home (before 08:00) to a high probability at work (after 10:00). At 11:30 am, a coworker sends out a group lunch invitation, hence the small probability shift back to the coffee shop around noon. After work, analysis of her social networks show no results. The graph reflects this lack of information by showing equal probabilities at all known establishments at approximately 18:00. Finally, analysis of her OSN show ten close friends posting messages related to a performance by *Soundgarden* at a local venue tonight. The high number of postings from close friends is reflected by a very high probability of her attending the performance.

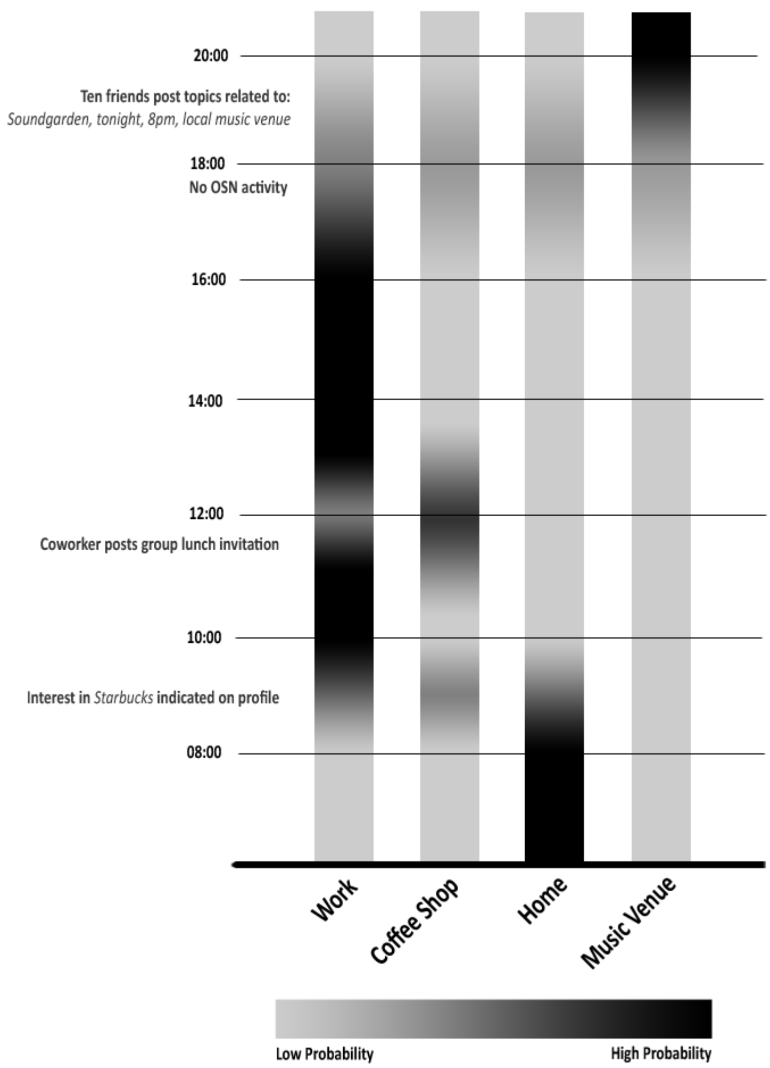

Figure 1. Spatio-temporal probability

Validation of this approach will be a key component of the research. Human participant studies will involve analysis of a participant's online social network data in order to create a predictive model of travel behavior and reduce uncertainty. The resulting model output will be compared against existing spatio-temporal models that do not include social data.

5. Summary and Concerns

The relevance of this research extends beyond theoretical constraints of time geography. Social constraints in spatio-temporal geography have real-world implications in the areas of location-based services, national security, and emergency response. This research also has its fair share of obstacles.

Privacy and data access are two major concerns when dealing with an individual's data [23]. Most social networking data available online is subject to strict privacy guidelines and is often inaccessible to academic researchers. Much of the data, though provided by an individual, is considered property of the OSN provider once it is published online. Though these are significant concerns, there are a number of sources for public social networking data and plenty of methods to access the data necessary to conduct this research [24][20].

References

[1] facebook Press Room http://www.facebook.com/press/info.php?statistics. (Accessed May 2011).

[2] twitter Blog #numbers. http://blog.twitter.com/2011/03/numbers.html (Accessed May 2011).

[3] Hägerstrand, T. What about people in regional science? *Papers in Regional Science*, *24*(1), (1970). 6–21.

[4] Miller, H. J. Activities in Space and Time. *Handbook of Transport 5: Transport Geography and Spatial Systems* (2004). 647-660.

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

[6] Kuijpers, B., Miller, H., Neutens, T., & Othman, W. Anchor uncertainty and space-time prisms on road networks. *International Journal of Geographical Information Science*, *24*(8), (2010). 1223-1248.

[7] Winter, S. Towards a Probabilistic Time Geography. *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '09*, 528. New York, New York, USA: ACM Press. (2009).

[8] Ellison, N. B., Steinfield, C., & Lampe, C. The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication*, *12*(4), (2007). 1143-1168.

[9] Carrasco, J. A., & Miller, E. J. Exploring the propensity to perform social activities: a social network approach. *Transportation*, *33*(5), (2006). 463-480.

[10] Kiefer, P., Bamberg, D.-, Raubal, M., & Schlieder, C. Time Geography Inverted : Recognizing Intentions in Space and Time. *ACM GIS*. (2010).

[11] Miller, H. J. A Measurement Theory for Time Geography. *Geographical Analysis*, *37*(1), (2005). 17-45.

[12] Ahas, R., & Mark, U. Location based services?new challenges for planning and public administration? *Futures*, *37*(6), (2005) 547-561.

[13] facebook Blog http://blog.facebook.com/blog.php?post=437112312130 (2010) (Accessed May 2011).

[14] Barabasi, A. *Linked: How Everything Is Connected to Everything Else and What It Means*. Plume, New York, NY. (2003).

[15] Caverlee, J., & Webb, S. A large-scale study of MySpace: Observations and implications for online social networks. *Proc. of ICWSM*, *8*. (2008).

[16] Goodchild, M., & Glennon, J. A. Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, *3*(3), (2010). 231-241.

[17] Hofmann, T. Probabilistic latent semantic indexing. *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '99*, (1999). 50-57. New York, New York, USA: ACM Press

[18] Mei, Q., Liu, C., Su, H., & Zhai, C. A probabilistic approach to spatiotemporal theme pattern mining on weblogs. *Proceedings of WWW. 2006* (2006). 533-542.

[19] Blei, D. M., Ng, A. Y., & Jordan, M. I. Latent Dirichlet Allocation. (J. Lafferty, Ed.)*Journal of Machine Learning Research*, *3*(4-5), (2003). 993-1022.

[20] Ramage, D., Dumais, S., & Liebling, D. Characterizing microblogs with topic models. *International AAAI Conference on Weblogs and Social Media*. (2010).

[21] Ritter, A., Cherry, C., & Dolan, B. Unsupervised modeling of Twitter conversations. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (2010). 172–180.
[22] Cheng, Z., Caverlee, J., & Lee, K. You are where you tweet: a content-based approach to geo-locating twitter users. *Proceedings of the 19th ACM international conference on Information and knowledge management* (2010). 759–768. Toronto: ACM.
[23] Boyd, D. M., & Ellison, N. B. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, *13*(1), (2008). 210-230.
[24] Pultar, E., & Raubal, M. A Case for Space : Physical and Virtual Location Requirements in the CouchSurfing Social Network. *Time*, (c), (2009). 88-91.