

Adding Social Constraints to Location Based Services

Grant McKenzie, Martin Raubal

Institute of Cartography and Geoinformation, ETH Zürich, Switzerland

Abstract. Online social networks (OSN) have experienced substantial growth over the past few years. Contributions to these OSNs offer a rich source of contextual information capable of augmenting the existing location based service model. Taking into consideration the inherent uncertainty in these new sources of online communication, the data can be exploited to expose social constraints from which a predictive model of user activities can be generated.

Keywords. Social Network, Time Geography, Location based services

1. Introduction

In recent years there has been considerable growth in the area of online social networking (OSN). Online applications such as *Facebook* have exceeded 800 million active users (Facebook Statistics 2011) and the micro-blogging service *Twitter* recently reported over 200 million tweets a day (Twitter Blog 2011). Current advances in mobile technology have allowed these platforms to expand into the area of location based services (LBS), providing users with the ability to add spatial context to their social interaction.

While social communication has been enhanced through location-aware technology, LBS have yet to truly exploit this rich new source of shared data. Context relevant information provided via LBS need not be based solely on a user's spatial and temporal location (Richter et al. 2010) but can also include social constraints. At the very least, exploration of online communication networks may provide social details ranging from gender and age to shopping trends or common modes of transportation. Further

analysis may offer insight into travel behavior and even the ability to predict an individual's spatio-temporal movement.

In terms of human-computer interaction, this plethora of social data will decrease reliance on user involvement, allowing the results of social data analysis to answer questions that would otherwise require manual input. The latest collaboration between social networking services and online search engines make use of this methodology, providing socially relevant search results (and advertisements) to end-users (Facebook Blog 2010). For example, *ten of your friends found this website interesting [Facebook like]* or, *my social profile lists surfing as one of my interests so surfing equipment is advertised along with my search results*. Location based services can take this a step further. Given a user's interest in surfing, a navigation service might suggest a route along the coast rather than through the mountains. Similarly, analysis of a mobile-based conversation between two acquaintances could predict user locations and recommend a local restaurant at which the contributors could meet.

Many of the LBS on the market today simply account for the physical limitations of the environment while incorporating less of the social-institutional or mental constraints. For example, most mobile guides would generally not suggest a route that forces a pedestrian to cross a major highway. Even fewer existing LBS acknowledge social-institutional impediments such as a bank's hours of operation or the social graces of wearing black to a funeral. While physical and institutional limitations of the environment are important, further emphasis needs to be placed on social and mental affordances, the reasons why certain travel decisions are made.

This research focuses on the social aspects of social-institutional limitations and explores the social factors that lead LBS users to make the decisions they do. This involves examining how existing social networking data can provide additional, context-relevant information to a LBS. The data contributed by many OSN members not only offers insight to the reasons they travel from location to location, but also offer a means of predicting future activities and travel.

2. 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 (Hägerstrand 1970, Miller 2004). The activities require certain capabilities and affordances in order to be utilized. These affordances exist in the physical, social-institutional, and mental realms (Raubal 2001). While

significant LBS research has focused on physical limitations (Raper et al. 2007), surprisingly little has explored the social-institutional dimensions.

Existing travel behavior research has already shown that we are creatures of habit. Individuals typically exhibit the same travel behavior day after day (Schlich et al. 2003) and predicting future movement given past transportation trajectories (González et al. 2008) can often be reasonably reliable. At present, very few of these predictive models incorporate social communication data. One of the reasons for this is that, until recently, social communication data have not been readily available. Admittedly, another motive may lie in the inherent uncertainty afforded to social communication (Antheunis 2010). In view of the recent rise of online social networks, we believe it is time to revisit the role of social data as constraints to a predictive location model.

This is not to say that no research is being done in this area. On the contrary, research in the area of online social information continues to explore many of these issues. Chang et al. (2011) explored past *Facebook* user check-ins and their role in predicting future check-ins. User mobility based on georeferenced tweets and *Foursquare* check-ins was investigated by Cheng et al. (2011) showing that online check-ins follow cyclical behavior. The power of the social network structure has also been the focus of a number of location-based studies, showing that an individual's location may be inferred based on the locations of their network links (Liben-Nowell et al. 2005, Backstrom et al. 2011).

A significant amount of research has gone into *extracting* relevant information from online social networks (Chakribarti et al. 2011, Hecht et al. 2011). While most of these studies involve the examination and use of online communication, surprisingly few have ground-truthed the data with real-world travel or activity data. Related research in transportation and travel behavior has looked at real-world social network interaction and how social relationships influence travel behavior (Carrasco & Miller 2006). Likewise, an individual's movement patterns and spatial behavior give insight into his/her social intentions (Kiefer et al. 2010). In contrast, studies have explored the issue of social exclusion, showing that movement in time and space is restricted for disadvantaged groups (Miller 2005). This proposed research also builds on Ahas & Mark's (2005) investigation of the *Social Positioning Method (SPM)*. The SPM associates individuals' mobile phone movements with their basic social data, provided through questionnaires. The finer (though fluctuating) resolution of OSN data allows for more accurate prediction of a user's spatio-temporal movement as it incorporates not only a user's social profile, but also activities that she determines to be socially relevant.

3. Predictive Model

Current LBS models consume spatial and temporal information in an attempt to provide context for the service being provided (Küpper 2005). This research focuses on accessing socially relevant information unique to the LBS user and incorporating additional contextual data into an existing LBS architecture with the purpose of enhancing a user's LBS experience. While a considerable amount of the material obtainable through online communication is subjective, much of this information can also be used to *explain* spatio-temporal movement. For example, an individual's GPS activity places her at the mall on a Tuesday afternoon. Her recent social network *post* illustrates excitement that a new album from her favorite musical artist is being released today. These social data allow us to understand her motivation for going to the mall on a Tuesday afternoon as well as offer further refinement to the location provided via GPS (the music store for example).

Not surprisingly, one observes that many online social communications (posts, tweets, etc.) present information in the future tense: "I can't wait for the new album", "Jane is looking forward to this weekend." These types of broadcasts generally feature upcoming events. Because of this knowledge, it is possible to predict where and when a user will be at a particular location.

The purpose of this research is to build a probability model with the intention of predicting an individual's activity location. We hypothesize that given previous social network data and travel history, this computation model will predict a user's spatial and temporal movement with significantly greater accuracy than estimates based solely on previous travel history.

4. Proposed Methods

4.1. Participants

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 of gathering data (Barabási 2003, Caverlee & Webb 2008). The study will involve two pools of 100 participants, equal parts female and male between the ages of 18-35. Participants must have an account and actively contribute¹ to the online social network *Facebook*.

¹An active contributor is defined as a user that contributes a minimum of 10 updates a week (status, links, photos, videos, comments or notes).

Participants will also be asked to keep a detailed activity diary, recording the date, time, location and reason for travel for the length of the study.

The first pool of participants will be randomly sampled with an attempt to avoid participants with more than five common network links (*friends*). Sampling will be achieved through advertisements on the social network platform itself. The second pool of participants will be randomly sampled from a small geographic and academic subset of the online social network. Students (undergraduate and graduate) from the University of California, Santa Barbara campus will be asked to participate. Given the social climate of a university campus, it is expected that each of the participants will have a minimum of ten friends in common with at least one other participant.

The purpose of the two sampling pools is to prove that tighter social networks (high average of common network links) result in more accurate location predictions. Since spatial proximity is often mirrored by social proximity (Butts 2003), there is a decent likelihood that proposed travel by one participant would be reflected in the online broadcasts of at least one of their network links. It is common knowledge that an individual's interests and likes are often reflected in the people with whom they associate.

4.2. Data Mining

An online social networking application has been built on the facebook platform. Once a participant grants access to the application, her OSN profile² and activities³ will be downloaded to a secure server for further analysis. Natural language processing and entity extraction algorithms will be used to extract location and transportation information as well as subjects of interest, such as sports, companies and persons/groups of interest (political, musical etc.).

4.3. Model Building

Figure 1 shows an actual workflow of the process from natural language to extraction of relevant contextual information. The model is based on both information provided through a participant's timeline (temporally significant social updates), and a participant's profile information (biographical and personal interests). The process involves determining

² Profile data include information related to location, gender, age, education, employment, relationships, languages, religious views, politics and interests.

³ Any activity posted on a participant's wall either from her or any of her network-links (status updates, photos, links, notes, videos, location updates, comments) as well as any posts to friend's walls.

common themes within a participant’s online profile and using these data as input to existing online application programming interfaces (APIs).

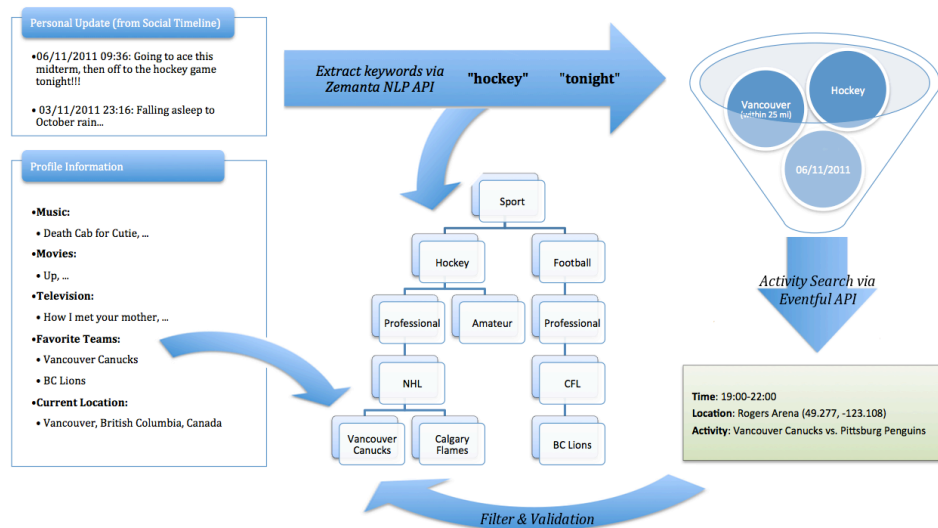


Figure 1 – Event Extraction

As seen in the example above, the participant’s profile lists a number of interests ranging from musical acts to sports teams. Given this information, when she posts a status update referencing an upcoming hockey game she plans on attending, there is a high probability⁴ that this game will involve one of the sports teams listed in her *interests*. The spatial and temporal inputs can be determined from the social timeline *update* as well. “Tonight” implies the same date as the posting and granted no other spatial context, the model defaults to the user’s “hometown” or “current city” listed in her profile information. These data is then assigned as input to an *event search API* (e.g. eventful.com) and filtered based on the interests listed on her social profile. The results of this process will have significantly reduced the possible locations at which the individual may be at any given time.

⁴ In this case, high probability refers to a higher probability of this event happening than another event (based on information provided).

4.4. Topic Modeling

Topic modeling provides an excellent framework from which to determine the general topics and themes in which individuals are interested. Clustering algorithms and topic modeling methods examine the entire participant pool as a whole and establish common subjects and themes inferred from across the entire dataset (Blei 2003, Ramage 2009). Each individual can then be defined as a distribution of these computationally extracted topics. This permits one to reason, for example, that Joe has more of an interest in outdoor activities than Jane. These distributions of topics across individuals also allow us to explore themes temporally. For example, last week Jane showed significant interest in music while this week revealed a slight increase in sports. These data combined with profile information allows assumptions to be made about the general likes and interests of a particular individual. In turn, this information provides the foundation on which the probabilistic model is built.

4.5. Network Weights

Lastly, the importance of online relationships must not be undervalued. Weights will be applied to nodes (friends/followers) on the social graph indicating strength of relationship. Information provided by close friends (frequent bi-direction communication) should be valued higher than those of the mere acquaintances. The power of a social *network* is that often nodes with links to the node in question (friends) are able to provide additional insight into behavior (in this case activities) that the individual node itself may not.

5. Preliminary Study

As a first step towards testing the above hypothesis, we conducted a preliminary study with five participants. All participants were between the ages of 25-35, two females and three males. The participants resided in United States (2), Canada (2) and Australia (1).

Participants were asked to provide a travel diary (hourly) for a one-week time span. After submitting the travel diary, participants allowed the researchers access to their facebook "walls" for the same one-week time span. The results showed that participants traveled to an average of 4.43 individually unique locations over a seven-day period. On average, participants' Facebook wall activity indicated 1.20 interactions over the same seven-day period. Of these online activities, an average of 0.19 contained information related to the current, past or future location of the

participant. Over the entire set of participants, this resulted in online interaction accounting for 5.2% of real-world activities.

While these are merely preliminary results based on an extremely small sample size, they do offer encouragement. Notably, of the five participants, the two females accounted for 100% of the location related posts (32% of all posts were related to their current, past or future locations), while one of the male participant's *walls* simply consisted of two statements, neither of which related to location (one link & one comment). Again, these initial results lead to questioning the role of gender, age and publishing frequency (to name a few) in assessing user location through online social networks.

6. Expected Outcomes

The expected outcome of this research is a probabilistic model that consumes natural language text from online social communications and outputs spatial and temporal location predictions along with a set of socially constrained probability values. It is anticipated that this model will supply significantly more accurate estimates than if OSN data were not exploited.

Figure 2 shows an overly simplified, graphical example of the model's expected outcome: 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) with a slight probability at *Starbucks* (around 09:00). 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 shows no results. The graph reflects this lack of information by showing equal probabilities at all known establishments at roughly 18:00. Finally, analysis of her OSN shows 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.

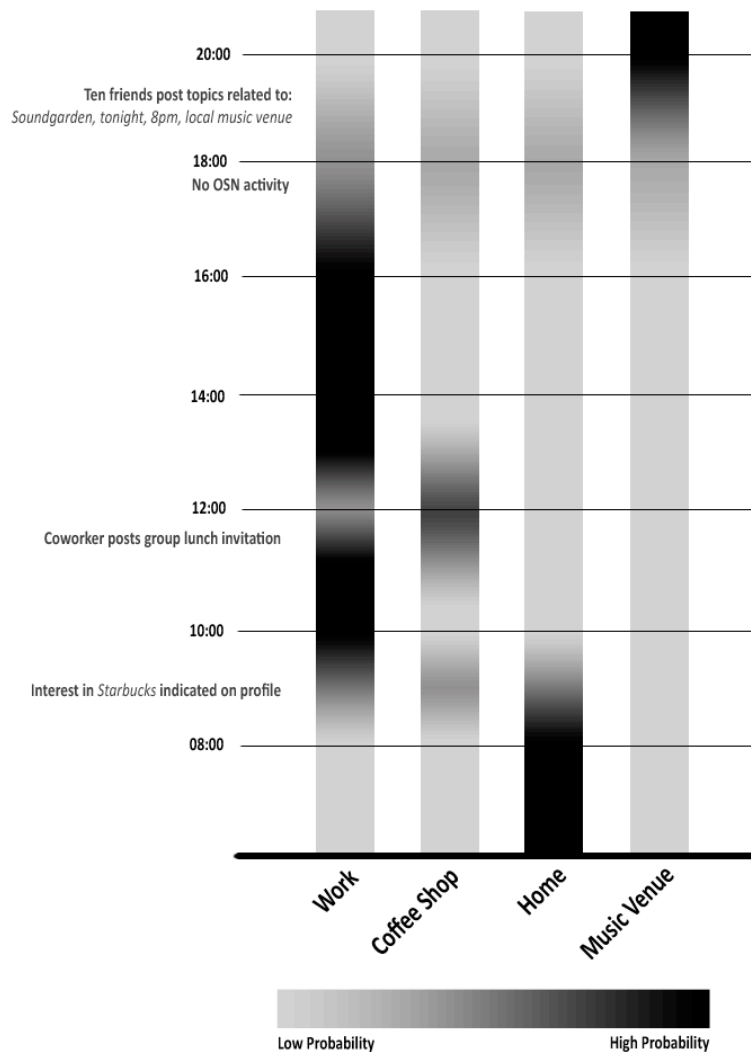


Figure 2 – Spatio-temporal probability of activities

Validation of this predictive model will be a key component of the research. The spatial and temporal predictions will be compared against participant travel diaries to ascertain the statistical accuracy of the simulation. The travel diaries will be taken as a true measure of a participant’s spatio-temporal movement. It is expected that the frequency of posts, as well as the amount of “usable data” contained within the post, will directly correlate with the accuracy of the model output.

7. Conclusion & Concerns

The effect of the dramatic increase in online social data contributions has yet to be fully understood. Humanity has been awarded with an unprecedented amount of shared data, most of which are publically available. Not only is this information entertaining to read and discuss, but it also offers a rich new source of contextual information—information that can be used to enhance our everyday lives. Location based services have been, and continue to be on the forefront of technology research. Many mobile devices coming to market today contain some form of location system (Berg Insight 2010) as well as some way of accessing at least one online social network. These social network data can enhance existing LBS by increasing spatio-temporal accuracy, contributing to a better user experience and a more contextually informed user.

This type of research does not come without limitations and concerns. Privacy and data access are two major concerns when dealing with these types of data (Barkuus & Dey 2003, Bulgurcu et al. 2010). Most social networking data available online are subject to strict privacy guidelines and are often inaccessible to the general public. Unfortunately, much of the data are 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.

Data accuracy is also a limitation in this environment. Information contributed to online social networks is inherently biased, volatile and ambiguous. Humans can be indecisive and untrustworthy and online contributions reflect this. These facts must not be overlooked in executing these types of studies, though we intend to show that regardless of the uncertainty in the data, its usefulness in providing social constraints to location based services persists.

References

- Ahas, R., & Mark, U. (2005) Location based services-new challenges for planning and public administration? *Futures*, 37(6):547-561.
- Antheunis, M. L., Valkenburg, P. M., & Peter, J. (2010). Getting acquainted through social network sites: Testing a model of online uncertainty reduction and social attraction. *Computers in Human Behavior*, 26, 100-109.
- Backstrom, L., Sun, E., & Marlow, C. (2010). Find me if you can: improving geographical prediction with social and spatial proximity. *WWW 2010* pp.61-70

- Barabási, A. (2003) Linked: How Everything Is Connected to Everything Else and What It Means. Plume, New York, NY.
- Barkuus, L., Dey, A. (2003) Location-Based Services for Mobile Telephony: a Study of Users' Privacy Concerns. *Proceedings of the INTERACT 2003, 9TH IFIP TC13 International Conference on Human-Computer Interaction*. 9
- Berg Insight (2010) <http://www.berginight.com/ReportPDF/Summary/bi-gps4-sum.pdf>. Accessed October 24, 2011.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. (J. Lafferty, Ed.) *Journal of Machine Learning Research*, 3(4-5):993-1022
- Bulgurcu, B., Cavusoglu, H., Benbasat, I. (2010) Understanding emergence and outcomes of information privacy concerns: A case of Facebook. *ICIS 2010 Proceedings*. 1:230.
- Butts, C. (2003) Predictability of large-scale spatially embedded networks. In P. P. Ronald L. Breig, Kathleen M. Carley (Ed.), Dynamic Social Network Modeling and Analysis
- Carrasco, J. A., & Miller, E. J. (2006) Exploring the propensity to perform social activities: a social network approach. *Transportation*, 33(5):463-480.
- Caverlee, J., & Webb, S. (2008) A large-scale study of MySpace: Observations and implications for online social networks. *Proc. of ICWSM*, 8.
- Chakrabarti, D., & Punera, K. (2011). Event summarization using tweets. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. pp.66-73
- Chang, J., & Sun, E. (2011). Location 3: How Users Share and Respond to Location-Based Data on Social Networking Sites. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. pp.74-80
- Cheng, Z., Caverlee, J., Lee, K., & Sui, D. Z. (2011). Exploring Millions of Footprints in Location Sharing Services. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. pp.81-88.
- Facebook Statistics (2011) <http://www.facebook.com/press/info.php?statistics>. Accessed October 24, 2011
- Facebook Blog (2010) <http://blog.facebook.com/blog.php?post=437112312130> Accessed October 24, 2011
- González, M. C., Hidalgo, C. a, & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196), 779-82
- Hägerstrand, T. (1970) What about people in regional science? *Papers in Regional Science*, 24(1): 6–21.
- Hecht, B., Hong, L., Suh, B., & Chi, E. H. (2011) Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. *Proceedings of the 2011 annual conference on Human factors in computing systems* pp.237–246
- Kiefer, P., Raubal, M., & Schlieder, C. (2010) Time Geography Inverted : Recognizing Intentions in Space and Time. *18th ACM SIGSPATIAL*

International Conference on Advances in Geographic Information Systems (ACM GIS 2010), San Jose, California, USA, pp. 510-513.

- Küpper, A. (2005) Location-Based Services: Fundamentals and Operation, John Wiley & Sons, Ltd, Chichester, UK.
- Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P., & Tomkins, A. (2005). Geographic routing in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(33):11623-11628
- Miller, H. J. (2004) Activities in Space and Time. *Handbook of Transport 5: Transport Geography and Spatial Systems*. 647-660.
- Miller, H. J. (2005) A Measurement Theory for Time Geography. *Geographical Analysis*, 37(1): 17-45.
- Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009). Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: 1:248-256*
- Raper, J., Gartner, G., Karimi, H., Rizos, C. (2007) Applications of location-based services: a selected review. *Journal of location-based services*, 1(2): 89-111
- Raubal, M. (2001) Ontology and epistemology for agent-based wayfinding simulation. *International Journal of Geographical Information Science* 15(7): 653-665.
- Richter, K.-F., Dara-Abrams, D., Raubal, M. (2010). Navigating and Learning with Location Based Services: A User-Centric Design. *7th International Symposium on LBS & TeleCartography*. G. Gartner and Y. Li. (Eds.). Guangzhou, China, pp. 261-276.
- Schlich, R., & Axhausen, K. (2003). Habitual travel behaviour : Evidence from a six-week travel diary. *Transportation*, 30, 13-36.
- Twitter Blog (2011) <http://blog.twitter.com/2011/06/200-million-tweets-per-day.html> Accessed October 24, 2011