What, When and Where: The real-world activities that contribute to online social networking posts

Grant McKenzie
Department of Geography
University of California, Santa Barbara
USA
grant.mckenzie@geog.ucsb.edu

Kathleen Deutsch
Department of Geography
University of California, Santa Barbara
USA
kdeutsch@geog.ucsb.edu

Martin Raubal
Institute of Cartography and Geoinformation
ETH Zürich
Switzerland
mraubal@ethz.ch

Abstract

Understanding the relationship between online social networking posts and real-world activities is key for many areas of research including activity prediction and recommendation engines. This paper presents the results of an exploratory study conducted with the purpose of extracting the types of an activity that are reported in an online post. The spatio-temporal components of activities are explored as well as the categories of the activities being conducted. Results suggest that activities that occur with less frequency are more likely to contribute to online action than those that are more routine.

Keywords: Social Networking, VGI, activities.

1 Introduction

As Online Social Networking (OSN) applications grow in influence and user base, a multitude of questions have arisen focusing on the relationship between our real world and the online virtual one. The OSN application *Facebook* recently reported an average of over 552 million daily active users [12] with almost as many interacting with the application via a mobile device. These statistics indicate that OSNs have become fully integrated into our everyday lives, though questions remain as to the extent of this integration. Given the ubiquity of OSN applications and our desire to increase social worth, the propensity to perform an activity in the real world and then broadcast the accomplishment of that activity through an OSN must be investigated. A better understanding of the relationship between our real-world accomplishments and our online social lives will have significant impacts in research areas ranging from activity behavior to location prediction systems.

A better understanding of the relationship between online activities and non-online activities should also direct the studying of the social structures of daily life. Social factors have been shown to have a significant impact on the activities we perform in the real world [2, 10, 18], but the question remains as to how that translates to the virtual world. Current location-based social networks (LBSN) such as *Foursquare*, *Yelp* and even *Facebook* allow users to broadcast their real-world locations along with status updates and photos related to the activities they are conducting at these locations, and with whom. As the users of these applications grow in number, it becomes more apparent that the desire to share one's activities is at least partially socially motivated. Discussions related to social capital and social worth emerge this notion of activity broadcasting [20]. What effect does publically announcing your activities have on your social relationships? The connection between the types of activities that one perceives to increase social worth and the types of activities that actually do is undoubtedly an area of future research. Taking one step at a time, our research offers a first step in exploring this connection and the activity types.

This paper presents a feasibility study that explores the *types*, *locations* and *time frames* of real-world activities and the likelihood that these activities are reported on an online social network. Using activity surveys and online activity tracking, this research offers an insightful view into the types of daily interactions and events that result in an online announcement. This exploratory study categorizes real-world activities based on established research guidelines and statistically determines which categories of activities are most prone to producing an OSN interaction. Given the spatio-temporal nature of activities (an activity must occur in space and time), we predict that the location and time of an activity play important roles in the types of activities that are reported.

The remainder of this paper is organized as follows. Section 2 introduces related work on the relationship between real and online activities. In Section 3 we describe the data collection and statistical methods used. Section 4 presents the results of our feasibility study while Section 5 and 6 discuss a number of limitations and offer conclusions and next steps for this area of research.

2 Related Work

2.1 Activity Categorization

Chapin defines activities as "classified acts or behavior of persons or households which, used as building blocks, permit us to study the living patterns or life ways of socially cohesive segments of society" [7:21]. These acts of behavior may be categorized into a number of different classes. Researchers investigating time use and activity patterns have produced a plethora of classification schemes [3, 19] each focused on a specific field. In addition, guidelines and suggestions have been developed to aid in classification of activities in endeavors such as activity based travel demand modeling [17]. Activities may be classified based on anything from their frequency, duration and sequence to social interaction [13]. Additionally, the location at which an activity takes place may also form the framework on which activities are categorized.

Chapin developed a classification scheme for aggregating activities into two levels. He categorized activities based on a glossary of approximately 230 activity codes [7]. Szalai et al. [22] built on Chapin's classification and produced a simpler system that they presented in their multinational time-budget study in 1972. These categories ranged from work to cultural events to transportation and travel. It is upon much of this work that the categories defined later are based.

2.2 Online Social Networking

In the area of online social networking, it has been shown that interaction through online social networking applications such as *Facebook* have a tendency to increase social capital [11]. Not surprisingly, measures and types of social capital change with the type of online social interaction. Research in the area of social networks has also explored the role of "friendships" both on- and offline. While there is definite overlap in "friends" that one has on an online social networking application and their real-world social network, the research suggests that online social networks play a role in strengthening different aspects of offline friendships [21]. This research takes the next step in understanding the activities that motivate these online interactions.

2.3 Activity Prediction

One of the primary motivations for conducting this research is to recognize the relationship between online posts and real-world activities. An understanding of this relationship can offer significant insight into activity prediction. A number of studies have explored the usefulness of social network data in predicting future activities [6, 9] while others have investigated the effectiveness of travel trajectories in determining a user's hometown or place of residence [15, 4].

Though most of these studies involve the exploration and use of online communities, surprisingly few have ground-truthed the data with real-world travel or activity data. While it may be possible to predict an individual's location based on her previous location history, was this person at any of the locations previously reported by the social networking application?

Previous studies have explored the idea of activity-based ground-truthing given real-world social survey information [1], but to our knowledge, very little, if any research has investigated the position of online social networking data in determining an individual's real-world location [16].

3 Methods

This section presents the methods used for data collection and analysis.

3.1 Data Collection

3.1.1 Participants

In order to assess the relationship between *Facebook* posts and real-world activities, a representative sample of online social network users was required. A total of 30 participants (15 female) between the ages of 20 and 45 (mean = 28.6, stdev= 5.2) were asked to participate in the research in return for a $20 USD gift card. These participants were sampled from *Facebook* using a snowball sampling method [5]. Social acquaintances of the principal researcher were contacted initially with the offer to participate in the research study. Additional participants were recruited through word-of-mouth interest from the initial contacts. The sole requirement for participation was that each participant have a history (the two weeks before the study) of posting to *Facebook* a minimum of once a day on average. This ensured a reasonable amount of data for analysis and removed bias due to lack of participation.

3.1.2 Collection Methods

The study requested participation over a continuous 3-week period by completion of two components. First, study participants were asked to record their activities (to the nearest hour) through a daily activity diary available online. Participants were asked to record the start time, end time, location and description of the activity performed. Both the location and description fields were presented as free-text boxes, allowing participants to use their own words when describing the location and activity. The instructions asked participants to be as detailed as possible when recording this information since the final goal of the location field was the ability to geocode.

Second, participants were required to install an application developed by the research team, allowing access to basic *Facebook* profile information[1] and social activities. The application gathered profile information on the specific participant (originally provided to *Facebook* by the user) as well as posts made by the user on her own wall (typically only visible to friends). By granting access to this application, online social network data was downloaded for each participant over the same 3-week period in which they were completing the daily activity diary. During the course of data collection, access to 2 of the participant's *Facebook* accounts was interrupted resulting in incomplete datasets, reducing the number of participants to 28. The study took place over a three-month period from October to December 2011.

3.2 Activity Categorization

From the 28 participants in the study, 3,198 activities were recorded to the daily activity diaries (mean=114, stdev=37.6). As discussed previously, the data consisted of activity and location descriptions entered as free-text by participants. In order to include these data in a statistical model, it was necessary to categorize each of the 3,198 activities recorded by the 28 participants. The two levels of classes into which each activity was grouped consisted of the type of activity (*What*) and the location of the activity (*Where*). These classes are discussed in detail in the next sections.

Categorization of activities was achieved through manual processing. The activities were anonymized and 3 researchers independently categorized each activity in both the type and location classes. The principle researcher then reviewed the results in order to ensure consistency in coding. When a disagreement of activity categorization was discovered, conflicts among classification were resolved. The purpose of this multiple-categorization was to remove as much bias as possible from the categorization process.

3.2.1 Class: *What*

The *What* class conveys the type of activity the participant was performing. It is important to note that an activity can be (and often was) classified in multiple categories. Table 1 lists the general categories established based on the data contributed by users. Example activities are shown as well.

It should be noted that the difference between local and distance travel is primarily within a city (commute) and between cities respectively.

Table1: Example activities categorized by *What*

Category	Example activity
Eating	Eating Dinner @ Home
Drinking (non-alcoholic beverages)	Yumm… Coffee @ Starbucks
Drinking (alcoholic beverages)	Drinking beer @ O'Hares
Fitness / Active	Bike Ride @ SB Bike Path
Watch TV / Play Video Game / Surf the Web	Watching Football @ Buddy's house
Local Transportation	Bus to work
Distance Transportation	Catching my flight to Toronto @ YVR Airport
Shopping	Christmas shopping @ Brentwood mall
Errands	Dentist Appointment
School	Classes, working @ UCSB
Work	Sound Design @ Work
Self Maintenance	Shower @ Home
Sporting Event	Go Canucks @ Rogers Arena
Cultural Entertainment (Concert, Museum, Play, etc.)	Sounds of Vienna Concert @ Kursalon, Vienna
Movie Theater	Mission Impossible @ South Edmonton Theaters
Vacation	Relaxing Vaca @ Victoria
Party / BBQ	Christmas Party @ Work
Other	Visiting @ Friend's house

[1] Profile information consisted of name, gender, birth date, location, email, education, hometown and username. It is important to note that Facebook considers email as the only required field.

3.2.2 Class: *Where*

The location at which each activity took place was categorized within the *Where* class. Unlike the *What* class, activities were categorized into a single location type (an activity could not take place at multiple locations). From a data-entry perspective, participants were asked to enter the name of a single location for each activity they performed. The example given was a spatial point described as an intersection of two streets. Actual entries from participants ranged from residential address to city level precision. For this reason, it was decided to simply group the *Where* activity tags into very general categories. Table 2 shows the categories as well as an example activity.

Table 2: Example activities categorized by *Where*

Category	Example activity
Home	Dinner with Family @ Home
Work/School	Working @ Westjest campus, Calgary
Transportation/Trip	Bus Home @ Waterfront Station
Other	Grocery Shopping @ West 4th & Vine

3.2.3 Class: *When*

The temporal component of activity posts is also worth exploring. The online activity diary required that participants enter both a start and an end time to their activities. The diary allowed participants to specify times to the nearest 30 minutes. Given the temporal bounds of these real-world activities, we are able to compare them to the online posts to which they are related.

3.3 Facebook Posts

For the purposes of this research, we define *Facebook Post* as a contribution of digital content to a user's social "wall." To refine it further, our research only analyzed textual input in the form of a status update made by one participant on her own wall. The total number of status updates for all users was 352 over the three-week period (mean=12.6, stdev=9.9). Of the 3,198 activities entered through the activity diary by the 28 participants, 75 of the activities were linked to one or more online post. The linking of these activities to posts was again achieved by manually matching an individual's online contribution to their real-world activity.

3.4 Analysis

The purpose of this research is to determine what types of activities at which locations are most likely to result in a post on the online social network *Facebook*. In order to achieve this goal, the two classes above were analyzed independently based on descriptive statistics.

3.4.1 Analysis of *What*

In exploring the original categories, we combined a number of the categories that were often tagged together. These categories roughly follow the *Time Budget Classification of Activities* framework developed by Szalai et al. [22] and built upon Chapin and Logan's activity classes [8]. Merging activity types was also a result of the frequency of matched-tags based on the participant-contributed descriptions.

Eating and drinking (non alcohol related activities) were grouped together as were common activities often done in sequence at home (sleeping, watching television, showering, etc.). *Cultural entertainment* was also combined ranging from visits to a museum to concerts and attending sporting events. Lastly, *shopping and running errands* were combined as the two are often done together, or could be synonymous. Table 3 lists the number of occurrences categorized from the self-reported activity diaries as well as the number of *Facebook* posts in which an activity of that category was reported.

3.4.2 Analysis of *Where*

While the categories related to *What* type of activity were aggregated, the *Where* categories were not. This was primarily due to the vagueness of locations reported in the self-reported activity diaries. As was the case with the categorical *What* data, each location category resulted in a number of online posts. Table 4 displays these data in a format mirroring the *Where* categories.

Table 3: Activities categorized by *What*

Category	Count	FB Post	Percentage reported on FB
Distance Travel	60	11	18.33
Vacation	30	3	10.00
Cultural Entertainment (Concerts, Museum, Theater, Sporting Events)	41	3	7.32
Drinking alcohol	96	5	5.21
Party / BBQ	70	2	2.86
Fitness	161	4	2.48
Shopping & Errands	453	9	1.99
Sleeping, watching TV, video games, browsing internet, self maintenance	969	17	1.75
Eating & Drinking (non-alcohol)	1017	17	1.67
Local Travel	240	4	1.67
School & Work	851	14	1.65

Table 4: Activities categorized by *Where*

Category	Count	FB Post	Percentage reported on FB
Transportation/Trip	298	14	4.70
Other	1235	38	3.08
Home	1022	16	1.57
Work/School	642	7	1.09

3.4.1 Analysis of *When*

Given the temporal bounds of a real-world activity, as provided by a participant, we are able to evaluate the relationship to the related online post in terms of time. As mentioned in previous sections, of the 3,198 daily activities, 86 *Facebook* posts were judged to be directly related to participants' real-world activities. It is important to note that these 86 posts include duplicate real-world events as a number of participants posted more than once regarding a specific activity.

In exploring the relationship between real-world activities and online *Facebook* posts from a temporal perspective, we ask, how close to an activity does a post occur in relation to the start of an activity? The data shows that on average a post occurs approximately 9.31 hours before the start of an activity, with a median of 4.78 and standard deviation of 32.89 hours.

4 Results & Discussion

4.1 *What*

Both *Distance Travel* and *Vacation* show the highest percentage of posts related to real-world activities. Activities categorized as *Drinking Alcohol* and *Cultural Entertainment* presented lower percentages indicating that they only slightly influence the likelihood of an OSN user posting online. Not surprisingly, the more mundane activities such as *Sleeping, watching TV and Self Maintenance* are the least influential in contributing to an online post, along with *Shopping & Errands*. More surprising is the fact that activities related to *BBQs & Party* had little to no sway on online posts. This could be largely due to the small sample size upon which this model was built.

The categories that demonstrated the highest influence on *Facebook* posts both related to events that occur less frequently than most other categorized occurrences. It follows that OSN users feel some increased sense of social value from traveling, perhaps as it presents a divergence from their average routine. Travel as it relates to vacation (and work for that matter) symbolizes financial stability, recreational activities and freedom from our daily routine that most cultures desire. This is slightly mirrored in the activities related to *Cultural Entertainment* such as concerts and sporting events.

4.2 *Where*

Again, we explore these data based simply on the number of online posts stemming from each location category. In looking at table 4, we see that the vast majority of activities is categorized as either *Home* or *Other* with both Transportation and Work producing less. Given the shear number of activities categorized as *Other* (occurring outside of home, work or travel), it is not surprising to see it resulting in the most number of online posts. However, it is interesting to note that the percentage of influence is much higher for *Other* and *Transportation* than *Home* or *Work*. Again, as with the *What* categorization, this increased influence on online contributions reflects a divergence from the routine activities that most likely occur at home and work. The *Transportation* category echoes the results from the *What* categorization showing that *Distance Travel* has a large influence on posts. This fits with our preconceived notions that activities done at work, school and home are not as interesting as those completed at other locations and therefore less worthy of being broadcast to our social circle.

4.3 *When*

The analysis of the temporal relationship between posts and activities indicates that on average, posts occur 9.31 hours before the start of an activity. In total, 79% of posts were written before the activity, ranging from approximately 9 days prior to the activity, to the exact start time of an activity (remember we are dealing with 30 minute resolution). In fact, the vast majority (91%) of posts in our sample set are within 24 hours of the start of an activity. These results suggest that access to an individual's online activity may offer insight into that individual's future real-world activities. These results imply significant value to research in the areas of activity prediction and recommendation engines.

5 Limitations & Next steps

The method and results presented in this paper have a number of limitations. Given that this research presents an exploratory study, the number of both participants and *Facebook* posts made by those participants is small. This small sample size should be taken into account when interpreting the results. For example, the category *Party & BBQ* shows 70 occurrences and only 2 posts tagged to this category. The first step in expanding this to a full study would be to increase the number of participants as well as the duration of the study. Moving from a participant pool of 30 individuals to 300 (for example) would allow for further statistical exploration using existing as well as new and more robust methods. Given this increased sample size, next steps would involve evaluating variable correlations and employing a binary choice model [14] to explore the influence of the different categories. It is expected that an increased number of participants would act to strength the results presented in this paper.

The self-reported activity survey should also be enhanced to include some level of participant tracking (e.g., GPS enabled mobile phones) along with strongly typed category choices for activities and locations. The free-text entry method undertaken in this research required considerable manual classification that could be avoided with standardized drop-down lists, or multiple-choice with an open ended "other" category. A more extensive, formalized list of categories combined with the increased number of participant activities would offer more insight into types of activities broadcast through an OSN. Alternatively, natural language processing approaches could potentially be employed to extract location information from the updates themselves. Machine learning techniques and geographic information retrieval methods could both be of considerable value to this area of research.

Additionally, a wider range of social networking applications will provide more breadth to the study and enable the generalization of results on a larger scale. *Facebook* is by far the most ubiquitous online social network today and the perfect source for a preliminary study such as this, but future research in this area can make use of the abundance of online applications surfacing every day.

6 Conclusions

This paper presented methods and results from an exploratory study investigating the relationship between daily activity schedules and online social networking posts. This first step showed that it is possible to conduct a study that explores the interaction between the real world and the virtual one. While this is an introductory stride in the much larger research agenda of ground-truthing online social networking data, the methods produced encouraging results. This study suggests that activities that occur with less frequency are more likely to contribute to online action than more routine activities. These results also intimate that users place high social value on activities that diverge from the norm such as vacations and travel. In summary, this exploratory study offers encouraging results for understanding the relationship between the real and the online social world.

References

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

[2] Ahas, R., Aasa, A., Silm, S., Aunap, R., Kalle, H., & Mark, Ü. (2007). Mobile Positioning in Space–Time Behaviour Studies: Social Positioning Method Experiments in Estonia. *Cartography and Geographic Information Science*, 34(4), 259-273.

[3] Axhausen, K., (2008). Social Networks, mobility biographies and travel: survey challenges. *Environment and planning B, Planning & design*, 35(6), 981.

[4] Backstrom, L., Sun, E., & Marlow, C., 2010, Find me if you can: improving geographical prediction with social and spatial proximity. *WWW 2010*, 61-70

[5] Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological Methods & Research*,10(2), 141-163.

[6] 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*. 74-80

[7] Chapin, F. S., (1974). *Human activity patterns in the city: Things people do in time and in space* (13-26). New York: Wiley

[8] Chapin, F. S., & Logan, T., (1967). Patterns of time and space use. *Contributions to the Future Forum on New Resources in an Urban Age*. Washington, D.C., 1-2 Nov.

[9] 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.

[10] Elliott, A., & Urry, J. (2010). *Mobile Lives*. London, UK: Routledge.

[11] Ellison, N. B., Steinfield, 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(4), 1143–1168.

[12] Facebook Statistics, 2012, http://newsroom.fb.com/content/default.aspx?NewsAreaId=22. Accessed May 2, 2012

[13] Golledge, R. G., & Stimson, R. J. (1996). *Spatial behavior: A geographic perspective*. Guilford Press.

[14] Lee, L. F. (1979). Identification and estimation in binary choice models with limited (censored) dependent variables. *Econometrica: Journal of the Econometric Society*, 977-996.

[15] 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

[16] McKenzie, G., & Raubal, M. (2012). Ground-truthing spatial activities through online social networking data. *Geographic Information Science 2012*, Extended Abstracts.

[17] National Cooperative Highway Research Program (NCHRP) (2008) Report 571 Standardized Procedures for Personal Travel Surveys Washington D.C.

[18] Páez, A. and Scott, D., (2007) Social influence on travel behavior: a simulation example of the decision to telecommute. In Environment and Planning: Part A. 39, 647-665.

[19] Parkka, J., Ermes, M., Korpipaa, P., Mantyjarvi, J., Peltola, J., & Korhonen, I. (2006). Activity classification using realistic data from wearable sensors. *Information Technology in Biomedicine, IEEE Transactions on*, 10(1), 119-128.

[20] Pultar, E., Winter, S., & Raubal, M. (2010). Location-Based Social Network Capital. Paper presented at the *Geographic Information Science - 6th International Conference, GIScience 2010*, Extended Abstracts Volume, Zurich, Switzerland.

[21] Subrahmanyam, K., Reich, S. M., Waechter, N., & Espinoza, G. (2008). Online and offline social networks: Use of social networking sites by emerging adults. *Journal of Applied Developmental Psychology*, 29(6), 420-433.

[22] Szalai, A. (1972). The use of time: daily activities of urban and suburban populations in twelve countries. *The use of time: daily activities of urban and suburban populations in twelve countries*.