Geographic Community Analysis of Mobile Social Network

Nan Li
Department of Computer Science
University of Massachusetts Lowell
1 University Avenue, Lowell, MA 01854
nli@cs.uml.edu

Guanling Chen
Department of Computer Science
University of Massachusetts Lowell
1 University Avenue, Lowell, MA 01854
glchen@cs.uml.edu

Abstract

Location-based Mobile Social Networks (MSNs) are becoming increasingly popular given the success of Online Social Networks (OSNs), such as Facebook and MySpace, and recent availability of GPS-embedded mobile platforms, such as Apple iPhones and Google Android phones. MSNs extend existing OSNs by allowing a user to know where her friends are and to meet new people nearby who share her interests. There are few studies, however, on how users are connected through these emerging MSNs.

In this paper, we present analysis results of a commercial MSN for which we applied the community-finding algorithm on geographic graphs of user movements and identified geographic characteristics of different locations. The discovered relationship between physical locations provides insights on user travel patterns that have implications on city planning and user similarity studies.

Keywords

Location-Based Mobile Social Networks, Geographic Community

1 Introduction

In recent years, we have witnessed that the Online Social Networks (OSNs) have been grown fast. The two of the most popular OSNs, Facebook[1] and Myspace,[2] have achieved tremendous successes. They attracted more than 225 million and 264 million registered users by June 2009, respectively.

Besides the traditional social networking services, micro-blogging sites are also growing fast. Twitter[3] is the leading service in this area. Instead of writing a long article, the micro-blogging sites allow users to share short messages (typically 140 characters) quickly and easily. Due to the convenience of posting these short messages on mobile devices, sharing user location also becomes an interesting feature for the micro-blogging. For example, Brightkite[4] enables users to attach current location information with their update messages. One of the driving factors of the location-based mobile social networking (MSN) is the availability of advanced mobile platforms, such as Apple iPhones and Google Android phones, which provide rich-media interactions with the social networking services from anywhere and at anytime. These mobile platforms leverage the GPS or signal triangulation technologies to automatically sense user's location, making location sharing much easier.

In this paper, we study the geographic characteristics of users' movements exposed through Mobile Social Network (MSN) updates. We selected the Brightkite as the target site, collected users' update messages with attached location information for 6 months, and analyzed how these users traveled or commuted during the 6-month period by applying community-detection algorithms. The analysis results show the different characteristics across different cities in the United States.

The rest of this paper is organized as follows. We will briefly discuss the related work in Section 2. In Section 3 we describe how we collected the user updates and how to calculate the community structures in the graph. The analysis results are presented in Section 4. The discussion and conclusion are presented in Section 5.

2 Related work

There have been numerous studies on OSNs. Mislove et al. studied extensively the structural graph properties of several popular OSN services, such as Flickr, YouTube, LiveJournal, and Orkut [12]. Nazir et al. studied how social gaming and utility applications were used on Facebook [13]. For micro-blogging related research, Java et al. studied Twitter, focusing on the topological and geographical properties of its social graph [6]. While our methodology is similar with these projects, we are particularly interested in the relationship of geographic locations instead of the user friendships.

The Reality Mining project at MIT studies mobile social applications by distributing to about 100 users mobile phones with customized software that records call logs for real social links and Bluetooth scans for geographical device proximity [3]. Ludford et al. studied how people shared the location knowledge thorough different location types using two small-scale controlled experiments [10]. These passively collected measurements were used to compare with users' self-reported social net-

[1] www.facebook.com
[2] www.myspace.com
[3] www.twitter.com
[4] www.brightkite.com

work structures. Our study, however, focuses on the location analysis at a much-larger scale (over United State).

One way to study user mobility is to passively observe how users' mobile devices associate with wireless access points (APs), assuming that a device's re-association indicates the user movement. This measurement-based approach has been used to study user mobility in a Computing Science department building [16], on academic campuses [8, 5, 11], on a corporate campus [1], and in a metropolitan area [17]. These studies have shown limited mobility that means users spent most of their time in their home location, while PDA users are more mobile than laptop users [11] and academic users are less mobile than corporate users [5]. Other researchers have conducted mobility studies using location traces from GPS-equipped vehicles [7], inter-device contact history from specially-designed sensors [2], or class schedules for students [15]. Contract to our studies of geographic relations, all these studies focus on mobility patterns of individual users.

3 Geographic Community

3.1 Data Collection

Brightkite is a Denver-based startup, founded in 2005, that allows users to post notes and upload photos through a number of interfaces, including Web, SMS, and Email. Each update is attached with user's location information, which can either be set by the user manually or be probed by website based on the IP address. The recently released native client applications on Apple iPhone and Google Android phones leverage GPS and other on-device technologies for automatic location sensing, though still requiring users to hit "check in" button to update their current location.

Brightkite website provides a RESTful API for integration with third-party applications. The interface is implemented by HTTP query, so an application sends a specified HTTP request and the website sends an XML or JSON response. An example XML-represented user text update is shown in Figure 3.1.

To collect users' updates, we first queried all updates that were posted between December 9, 2008 to January 9, 2009. There were 18,951 users who made at least one update during that period. Our analysis thus focused on these users, because we believe that the other users are inactive since they had not make any update for a month. After deciding the user population, we then queried all updates from each of these users. Brightkite returned 1,505,874 updates in total, from March 21, 2008 to January 9, 2009.

3.2 Community Calculation

While the Brightkite data trace covers location points over the global area, our analysis in this paper focuses on the data set covering the United States. We further restricted the vertices in the geographic graph by identifying 213 large cities in the United State, whose city populations were larger than 200,000 as 2008.

Since each Brightkite update is tagged with location information, we can find the nearest city for each update


```
<note>
  <created_at_ts>1234822696.004165</created_at_ts>
  <body>in the umpteenth surreal moment between nice
  memories and a sweet doubt. the requirements are
  good to have a nice sleep...</body>
  <creator>
    <fullname>Gennaro Del Giudice</fullname>
    <login>gdelgiudice</login>
  </creator>
  <public type="boolean">true</public>
  <place>
    <scope>address</scope>
    <display_location>Viale Fulvio Testi, 280,
      20126 Milano, Italy</display_location>
    <longitude type="float">9.21059</longitude>
    <name>Viale Fulvio Testi, 280</name>
    <id>2d3335eedb6b11ddbf3f003048c10834</id>
    <latitude type="float">45.523338</latitude>
  </place>
  <created_at type="datetime">2009-02-16T22:18:16Z
      </created_at>
  <id>b5162494fc7711ddb34c003048c10834</id>
</note>
```


Figure 1. An example of text update with some entries omitted.

by calculating the geographical distance between the update and each city. Brightkite provides different location granularity depending on the users' configuration, such as city level, street level, and address level. In this calculation, we ignored those differences, because we are only interested in the relationship between the cities. We also eliminated the updates that were more than 30 miles away from any of the 213 cities. We assumed that the updates from 30 miles away have no meaningful relationship with the city. By applying these rules, we pruned the updates to 648,188, and mapped each update to a city.

An edge between two vertices (cities) represents that a user sent two consecutive updates in these two cities. We identified 47,240 such update pairs. Each edge is labeled with a weight that is the count of total update pairs for the two cities of that edge. So the location graph is a weighted non-directed graph. The total edge number is 2,907. Assuming that there are a lot of people traveled between city A and city B (sending one update in A and next one in B or vice versa), there will be a large-weight edge between A and B to indicate the tight relationship between them.

We then calculated the community structures (vertex clusters) in the geographic graph by applying the *Girvan and Newman Algorithm* [4]. The algorithm's basic idea is that in a graph, a few edges that lie between communities can be thought of as forming "bottlenecks" between the communities. By removing those "bottlenecks" edges, the community structures will be isolated. In order to find those edges, they calculated the betweenness of the edges, which is defined to be the number of geodesic (i.e., shortest) paths between vertex pairs that run along the edge in question, summed over all vertex pairs [14]. This metric can be calculated for all edges in time of $O(mn)$ on a graph with m edges and n vertices.

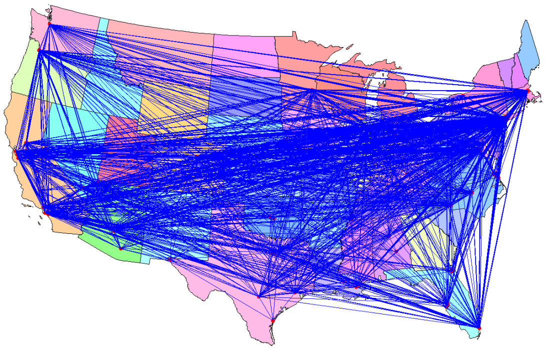

Figure 2. The original geographic graph

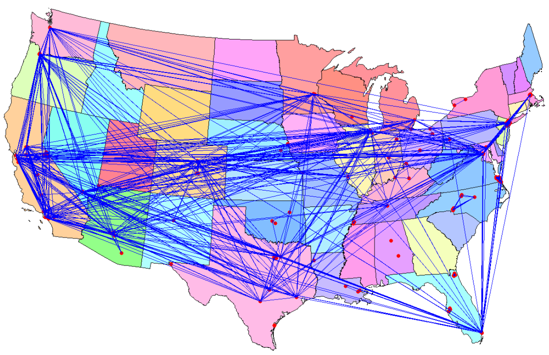

Figure 3. The geographic graph with 1000 edges left

4 Results

Figure 2 shows the geographic graph before applying the community finding algorithm. As we described, there are 213 vertices and 2,907 edges in the graph, whose average edges (outbound or inbound) of each vertex is 27.29. It is a relatively dense graph.

Figure 3 shows the geographic graph with 1000 edges left (after 1907 edges removed). We found that the edges between large cities are still quite strong. It makes sense since the majority of travel trips are concentrated on those edges. There is no obvious community separated at this step.

We kept applying the community finding algorithm to remove additional edges until there are 500 edges left. Figure 4 shows the geographic graph. The eastern and

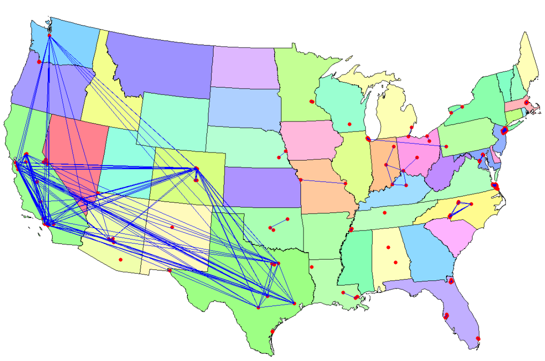

Figure 4. The geographic graph with 500 edges left

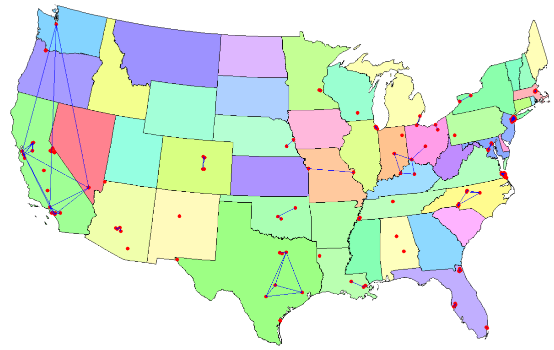

Figure 5. The geographic graph with 300 edges left

the western cities clearly have different community characteristics. The eastern cities form communities in close proximity. The adjacent cities around Boston, New York City, Washington DC, and Virginia Beach, form communities and their distances to each other are no more than 50 miles. These communities are shown as a big red point due to short distance between community cities (Figure 4). In the state of North Carolina, the Greensboro, Raleigh, and Charlotte form a community, in which city distances are a bit farther but still less that 150 miles. The only eastern community of relatively long-distance cities is the one formed by Chicago and Pittsburgh. It will be interesting to study why these two cities have such a strong bond. On the other hand, most of the western and southwest cities are still connected as communities in large geographical range. All of these cities form a large-distance community, whose geographical diameter is up to 2,000 miles, from Seattle to Houston.

Figure 5 shows the geographic graph with 300 edges left. In the west coast, Seattle, San Francisco, Los Angeles, Las Vegas and the adjacent cities around them form a big tight community. In Texas, Dallas, Houston, Austin, and San Antonio form a community in a relatively large area.

One of the potential applications for OSNs using above results is friend recommendation. It is helpful to recommend non-friends to users based on their shared interests, hobbies, and travel patterns, especially for large OSNs. In [9], we have developed a multi-layered friend recommend model for mobile social networks. We used the user's geographical distance as one of the metrics to recommend friends. As we have seen from the results above, however, the geographical distance is not always a good metric for users' closeness. Some cities have much closer community relationship than others. It is thus reasonable to recommend friends belong to these cities even if they are geographically far apart.

5 Conclusion

In this paper, we present a community analysis of the large cities in the United States. The connections between the cities are derived from real user movements reported through mobile social networks. We found that the geographical distance between the cities does not necessarily reflect the closeness between people in those cities. Some cities have tight relationship even when they are far away. These characteristics are especially different between the eastern and western cities. As a future work, the analysis results can be used as a metric for location-based friend recommendation in mobile social networks.

6 References

[1] M. Balazinska and P. Castro. Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network. In *Proceedings of the First International Conference on Mobile Systems, Applications, and Services*, pages 303–316, San Francisco, CA, May 2003.

[2] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Pocket switched networks: Real-world mobility and its consequences for opportunistic forwarding. Technical Report UCAM-CL-TR-617, University of Cambridge Computer Laboratory, Feb. 2005.

[3] N. Eagle, A. S. Pentland, and D. Lazer. Inferring social network structure using mobile phone data. In *Social Computing, Behavioral Modeling, and Prediction*. Springer, Mar. 2008.

[4] M. Girvan and M. E. Newman. Community structure in social and biological networks. *Proc Natl Acad Sci U S A*, 99(12):7821–7826, June 2002.

[5] T. Henderson, D. Kotz, and I. Abyzov. The changing usage of a mature campus-wide wireless network. In *Proceedings of the Tenth Annual International Conference on Mobile Computing and Networking*, pages 187–201, Philadelphia, PA, Sept. 2004.

[6] A. Java, X. Song, T. Finin, and B. Tseng. Why we Twitter: Understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, pages 56–65, Aug. 2007.

[7] T. Kobayashi, N. Shinagawa, and Y. Watanabe. Vehicle mobility characterization based on measurement and its application to cellular communication systems. *IEICE Transactions on Communications Special Issue on Multimedia Mobile Communication Systems*, E82-B(12), Dec. 1999.

[8] D. Kotz and K. Essien. Analysis of a campus-wide wireless network. In *Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking*, pages 107–118, Atlanta, GA, Sept. 2002.

[9] N. Li and G. Chen. Multi-layer friendship modeling for location-based mobile social networks. In *Proceedings of the Sixth Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous)*, Toronto, Canada, July 2009. To appear.

[10] P. J. Ludford, R. Priedhorsky, K. Reily, and L. Terveen. Capturing, sharing, and using local place information. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1235–1244, New York, NY, USA, 2007. ACM.

[11] M. McNett and G. M. Voelker. Access and mobility of wireless PDA users. *Mobile Computing Communications Review*, 9(2):40–55, Apr. 2005.

[12] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the Seventh ACM Internet Measurement Conference*, San Diego, CA, Oct. 2007.

[13] A. Nazir, S. Raza, and C.-N. Chuah. Unveiling Facebook: A measurement study of social network based applications. In *Proceedings of the Seventh ACM Internet Measurement Conference*, Vouliagmeni, Greece, Oct. 2008.

[14] M. E. J. Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical Review E*, Volume 64, June 2001.

[15] V. Srinivasan, M. Motani, and W. T. Ooi. Analysis and implications of student contact patterns derived from campus schedules. In *Proceedings of the 12th Annual International Conference on Mobile Computing and Networking*, pages 86–97, Los Angeles, CA, Sept. 2006.

[16] D. Tang and M. Baker. Analysis of a local-area wireless network. In *Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking*, pages 1–10, Boston, MA, Aug. 2000.

[17] D. Tang and M. Baker. Analysis of a metropolitan-area wireless network. *Wireless Networks*, 8(2/3):107–120, March–May 2002.