Where Online Friends Meet: Social Communities in Location-based Networks

Chloë Brown **Vincenzo Nicosia** **Salvatore Scellato** **Anastasios Noulas** **Cecilia Mascolo**

Computer Laboratory
University of Cambridge
Firstname.Lastname@cl.cam.ac.uk

Abstract

Recent research suggests that, as in offline scenarios, spatial proximity increases the likelihood that two individuals establish an online social connection, and geographic closeness could therefore influence the formation of online communities. In this work we present a study of communities in two online social networks with location-sharing features and analyze their social and spatial properties. We study the places users visit to understand whether communities revolve around places or whether they exist independently. Our results suggest that community structure in social networks may arise from both social and spatial factors, so that exploiting information about the places where people go could benefit the definition of new community detection methods and community evolution models.

Introduction

Millions of people now interact with their friends online using social networking services. Just as their offline counterparts, these online social networks show community structure. Online social communities have been extensively studied, with implications for link prediction systems and recommendation engines (Backstrom et al. 2006; Mislove et al. 2007; Papadopoulos et al. 2011). Much work has addressed the problems of extracting and analyzing social network communities, but the factors that drive their formation are still not well understood. Different theories have been put forward to explain why social groups arise. The main dichotomy is between the common identity and the common bond theories (Back 1951; Ren, Kraut, and Kiesler 2007). The common identity theory says that individuals gather into groups when they share a common interest or purpose, as for the fans of a sport team. The common bond theory proposes that communities are held together by the social ties between their members, as in families or in groups of close friends. The factors that keep communities alive are often not evident in the network structure, so additional information about their members must be exploited to understand why communities form. We propose to analyze the places where people go, in order to understand online social communities better.

Copyright © 2012, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Communities and physical space

Physical space has a fundamental effect on social ties. The probability of friendship quickly decreases with increasing spatial distance between people (Stewart 1941). Social groups tend to be constrained by geography, with smaller communities being tight in space (Onnela et al. 2011). It was thought that the advent of the Internet could change this; researchers suggested the "Death of Distance" (Cairncross 2001): people would interact on a planetary scale, regardless of location. However, recent work has found that geographic distance also affects online relationships (Liben-Nowell et al. 2005; Mok, Wellman, and Carrasco 2009; Backstrom, Sun, and Marlow 2010; Scellato et al. 2011).

The focus theory of social bonds suggests that activities and interactions occurring in physical places foster social ties (Feld 1981). This would suggest that knowing where the members of a community go can reveal much more than just spatial characteristics. A new generation of online services has recently enabled such analysis: location-based social networks are mobile social services built around location sharing. People use mobile applications to *check in* at venues and notify their friends. This exposes a wealth of information about online social ties and the places users visit, ideal for addressing questions about the spatial characteristics of online communities: *Which factors hold an online community together? Are physical places important to online communities? Is an online community focused around a particular place?*

Our contributions

We seek to answer these questions by analyzing communities in two online social services with location-sharing features: Gowalla,[1] and Twitter[2]. We examine communities' social and spatial properties, and show that *social communities tend to arise from different factors on different online platforms*: Twitter communities seem to form around popular users, but places play a crucial rôle in Gowalla. The finding that places can be important to communities in some social networks opens up possibilities for further work on community detection methods and models of community formation, as well as the design of new systems and applications.

[1] www.gowalla.com
[2] www.twitter.com

Dataset	N	K	N_{GC}	$\langle k \rangle$	$\langle c \rangle$	L	C
Gowalla	65,504	295,380	62,260	9.01	0.249	799,116	7,388,401
Twitter	123,665	544,215	120,242	8.80	0.108	1,024,057	3,868,845

Table 1: Properties of the datasets: number of nodes N and edges K, number of nodes in the giant connected component N_{GC}, mean node degree $\langle k \rangle$, mean clustering coefficient $\langle c \rangle$, total number of places L and total number of check-ins C.

Data description

In this work we analyze datasets from two online social services with location-sharing features, Gowalla and Twitter.

Gowalla is a location-based social network created in 2009 and discontinued when the company was bought by Facebook at the end of 2011. Users *check in* at named places and share their location with their friends. In this work we use a complete snapshot of the service acquired in August 2010.

Twitter is one of the most popular online social networks, with over 300 million registered users at the end of 2011. Its main focus is not location sharing, but users of location-based services such as Foursquare[3] share their check-ins publicly through Twitter. Foursquare is the most popular online location-based social network, with over 15 million users in January 2012. We consider Foursquare check-ins pushed to Twitter between May and November 2010. Social ties between users are inferred from their Twitter connections at the end of the measurement period; we take two users to have a social link when each follows the other.

We consider users whose most checked-into location has latitude between 18°N and 72°N and longitude between 66°W and 179°W. This includes the USA, where both networks have many users. Properties of the datasets are shown in Table 1.

Notation and measures

We now define the notation and measures we use in our analysis.

We represent each network as an undirected graph $G = (V, E)$, where the set $V = \{u_1, u_2, \ldots, u_N\}$ is the set of N users, and the set of edges E is composed of pairs of users present in one another's friend lists. We define F_i to be the friends of user u_i. There are L different places $M = \{m_1, m_2, \ldots, m_L\}$ where users have checked in, and c_{ij} represents the number of check-ins that user u_i has made to place m_j. M_j is the set of users who have checked in at place m_j, and U_i is the set of places where user u_i has checked in.

A second network $G^P = (V, E^P)$ is formed from the same nodes V and the set E^P of *placefriends* edges. E^P contains the edge (u_i, u_j) whenever different users u_i and u_j have checked in at one or more of the same places. We call the graph G^P the *placefriends* graph.

Community measures

Finding communities in a social network can be seen as partitioning the set of nodes V into subsets. Given a community C, our aim is to express quantitatively its social and place properties, in order to address the questions of whether communities form around and are held together by people or places, and what rôle, if any, places play for communities.

Social properties *Edge density*: The fraction of possible edges between community members that are actually present:

$$\frac{1}{|C|(|C|-1)} \sum_{u_i, u_j \in C} A_{ij} \qquad (1)$$

where A_{ij} is the ij^{th} entry in the adjacency matrix of the graph (1 when u_i and u_j are connected; 0 otherwise). Communities where many members have ties to many other members have high edge density.

Maximum internal fractional degree: The maximum proportion of other members of C to whom a single user in C has social ties:

$$\max_{u_i \in C} \left(\frac{|F_i \cap C|}{|C|-1} \right) \qquad (2)$$

High values indicate that a member is connected to most of the others. If a community has high maximum internal fractional degree but low edge density, this would suggest that the individual to whom most users are connected is important in holding the community together.

Place properties *Placefriends edge density*: The fraction of possible edges between members of C in the placefriends graph G^P that are actually present:

$$\frac{1}{|C|(|C|-1)} \sum_{u_i, u_j \in C} A^P_{ij} \qquad (3)$$

where A^P_{ij} is the ij^{th} entry in the adjacency matrix of G^P (1 if u_i and u_j have a place in common; 0 otherwise). This value is large when a community member typically shares at least one place with most of the other members, and would indicate that physical places may be important to the community; it does not exist solely online.

Members sharing one place: The maximum fraction of members of C who have visited a particular location:

$$\max_{m_i \in M} \left(\frac{|M_i \cap C|}{|C|} \right) \qquad (4)$$

This captures whether or not *most* members of the community visit the *same* venue.

Minimum check-ins to most-shared place: Let m_v be the venue that the largest fraction of members of community C share. We compute the minimum number of check-ins to m_v by a member of C who has at least one check-in there:

$$\min_{\{u_i \in C | c_{iv} > 0\}} c_{iv} \qquad (5)$$

[3] www.foursquare.com

If the members of C share no places, this measure is defined to be 0. This can show whether a community has formed around a place through its members visiting regularly and meeting one another there, as suggested by the focus theory of social bonds (Feld 1981). For communities where many members share the most-shared location, high values suggest that the community could have formed with that place as its focus.

Analysis and discussion

We used the Louvain algorithm (Blondel et al. 2008) to find communities in each of the two networks. We now discuss the social and spatial properties of these communities.

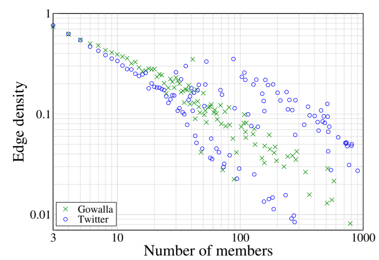
Social properties

Figure 1 reports the mean edge density and mean maximum internal fractional degree for communities of a given size. Edge density falls with increasing size. Twitter shows a marked tendency towards communities with maximum internal fractional degree close to 1, even for large communities of around 100 nodes. This indicates hub nodes connected to almost all of the other community members. Gowalla, on the other hand, shows a steady decrease of maximum internal fractional degree with community size, and lower absolute values. This suggests that in Twitter, communities may form around important individuals. Given the decrease in edge density with increasing community size, the rôle of these popular users becomes more important to hold the community together as the community gets larger. This does not seem to be the case in Gowalla, which raises the question as to whether, in this explicitly location-based network, places may play this part instead.

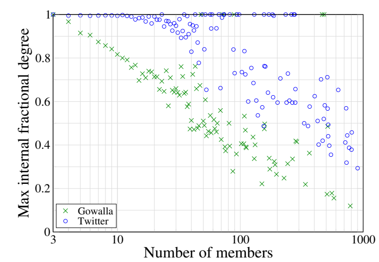
Place properties

Figure 2 shows the mean proportion of members visiting a single place, and mean placefriends edge density, for communities of each size. Both measures have noticeably lower values for Twitter than Gowalla. The relative unimportance of places for Twitter communities may reflect Twitter's focus on content sharing. As proposed by the common identity theory, users may form connections to others with shared interests, possibly independently of location. In contrast, Gowalla communities of 10 or fewer members have mean edge densities of above 0.6: most members share at least one place with most others. Additionally, many communities have a place where a large fraction of their members have been, even for very large communities. This suggests that common places may hold Gowalla communities together, just as popular users do in the Twitter communities. While online social ties hold Twitter communities together, Gowalla communities seem to revolve around places and may represent local groups in the offline world.

We have thus used our social and place measures to give possible answers to our questions: *Which factors hold an online community together? Are physical places important to online communities?* We now use the final place measure to test the focus theory of social bonds described in the introduction, and address the third question: *Is an online community focused around a particular place?*



(a) Edge density



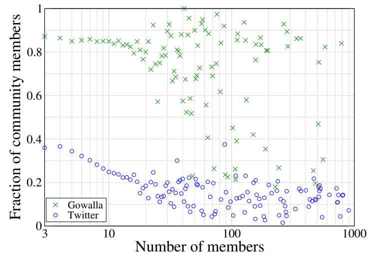
(b) Maximum internal fractional degree

Figure 1: Social properties of communities

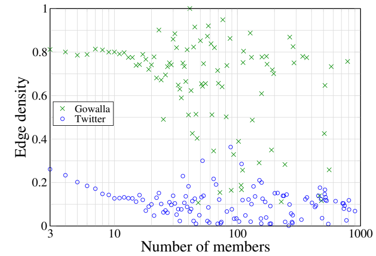
Figures 2(c) and 2(d) report the minimum number of check-ins to the community's most-shared place by a member who has visited that place (0 if a community shares no places). A vast majority of communities have a member with just one check-in to the most-shared place. In Twitter this is not surprising as we have seen that places are not important to communities. For Gowalla, while places are highly important, it seems that these communities are likely *not* indicative of groups that form around *single* significant places, such as groups of colleagues with the same workplace, or a sports team who visit their training ground regularly. If these groups had formed through members' regular interaction at single places, as proposed by the focus theory of social bonds, we would expect members to have visited that place many times. Instead, place sharing in Gowalla communities may indicate *social-driven mobility*, where users visit the same places because they are friends.

Conclusions and future work

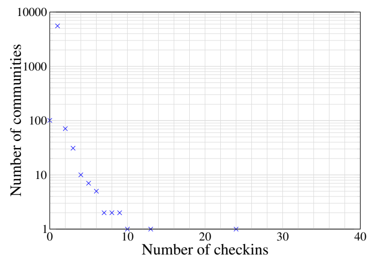
We have presented a study of the social and spatial properties of communities in Gowalla and Twitter, and found that in Twitter popular users hold communities together, while in Gowalla community members tend to visit the same places. Our results demonstrate that online communities can arise for different reasons. For some social networks, Gowalla included, geography is essential to characterize communities fully. This opens up new directions for community detection methods. Methods aiming to cancel out the impact of geography on social ties have been proposed to detect meaningful communities that are independent of physical space (Expert et al. 2011). However, our results imply that in networks such as Gowalla where places are crucially relevant to com-



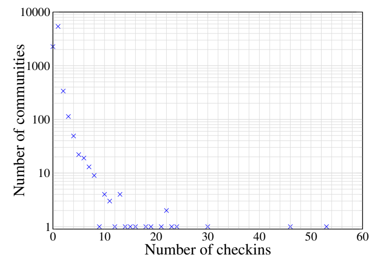
(a) Maximum fraction of members visiting one place



(b) Edge density of placefriends graph



(c) Minimum check-ins to most-shared place: Gowalla



(d) Minimum check-ins to most-shared place: Twitter

Figure 2: Place properties of communities

munities, it might instead be appropriate to *exploit* location information to improve community detection. Furthermore, since places can play a vital rôle for communities, social theories such as the common bond and common identity theories could be used *in conjunction* with location information to capture how communities form in real systems, and to define more accurate models of community evolution.

Acknowledgements

Chloë Brown is a recipient of the Google Europe Fellowship in Mobile Computing, and this research is supported in part by this Google Fellowship.

References

Back, K. W. 1951. Influence through social communication. *Journal of Abnormal and Social Psychology* 46(1):9–23.

Backstrom, L.; Huttenlocher, D.; Kleinberg, J.; and Lan, X. 2006. Group formation in large social networks: membership, growth, and evolution. In *Proceedings of KDD '06*, 44–54. New York, NY, USA: ACM.

Backstrom, L.; Sun, E.; and Marlow, C. 2010. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of WWW '10*, 61–70. Raleigh, North Carolina, USA: ACM.

Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008(10):P10008.

Cairncross, F. 2001. *The Death of Distance: How the Communications Revolution Is Changing our Lives*. Cambridge, MA, USA: Harvard Business School Press.

Expert, P.; Evans, T. S.; Blondel, V. D.; and Lambiotte, R. 2011. Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences* 108(19):7663–7668.

Feld, S. L. 1981. The Focused Organization of Social Ties. *The American Journal of Sociology* 86(5):1015–1035.

Liben-Nowell, D.; Novak, J.; Kumar, R.; Raghavan, P.; and Tomkins, A. 2005. Geographic routing in social networks. *PNAS* 102(33):11623–11628.

Mislove, A.; Marcon, M.; Gummadi, K. P.; Druschel, P.; and Bhattacharjee, B. 2007. Measurement and analysis of online social networks. In *Proceedings of IMC '07*, 29–42. San Diego, California, USA: ACM.

Mok, D.; Wellman, B.; and Carrasco, J. A. 2009. Does distance matter in the age of the Internet? *Urban Studies* 46.

Onnela, J.-P.; Arbesman, S.; González, M. C.; Barabási, A.-L.; and Christakis, N. A. 2011. Geographic constraints on social network groups. *PLoS ONE* 6(4):e16939.

Papadopoulos, S.; Kompatsiaris, Y.; Vakali, A.; and Spyridonos, P. 2011. Community detection in social media. *Data Mining and Knowledge Discovery* 1–40.

Ren, Y.; Kraut, R.; and Kiesler, S. 2007. Applying common identity and bond theory to design of online communities. *Organization Studies* 28(3):377–408.

Scellato, S.; Noulas, A.; Lambiotte, R.; and Mascolo, C. 2011. Socio-spatial Properties of Online Location-based Social Networks. In *Proceedings of ICWSM '11*.

Stewart, J. Q. 1941. An inverse distance variation for certain social influences. *Science* 93(2404):89–90.