Socio-spatial Properties of Online Location-based Social Networks

Salvatore Scellato
Computer Laboratory
University of Cambridge
salvatore.scellato@cam.ac.uk

Anastasios Noulas
Computer Laboratory
University of Cambridge
anastasios.noulas@cl.cam.ac.uk

Renaud Lambiotte
Deparment of Mathematics
Imperial College London
r.lambiotte@imperial.ac.uk

Cecilia Mascolo
Computer Laboratory
University of Cambridge
cecilia.mascolo@cl.cam.ac.uk

Abstract

The spatial structure of large-scale online social networks has been largely unaccessible due to the lack of available and accurate data about people's location. However, with the recent surging popularity of location-based social services, data about the geographic position of users have been available for the first time, together with their online social connections.

In this work we present a comprehensive study of the spatial properties of the social networks arising among users of three main popular online location-based services. We observe robust universal features across them: while all networks exhibit about 40% of links below 100 km, we further discover strong heterogeneity across users, with different characteristic spatial lengths of interaction across both their social ties and social triads. We provide evidence that mechanisms akin to gravity models may influence how these social connections are created over space. Our results constitute the first large-scale study to unravel the socio-spatial properties of online location-based social networks.

Introduction

Online Location-based Social Networks (LBSNs) have recently attracted millions of users, experiencing a huge popularity increase over a short period of time. Thanks to the widespread adoption of location-sensing mobile devices, users can share information about their location with their friends. Among the biggest providers there are Foursquare and Gowalla, while other hugely popular social networking services such as Facebook and Twitter have also introduced location-based features.

Location is increasingly becoming a crucial facet of many online services: people appear more willing to share information about their geographic position with friends, while companies can customize their services by taking into account where the user is located. As a consequence, service providers have access to a valuable source of data on the geographic location of users, as well as to online friendship connections among them. The combination of these two factors offers a groundbreaking opportunity to understand and exploit the spatial properties of the social networks arising among online users, but also a potential window on real human socio-spatial behavior.

Spatial networks have been extensively studied, particulary when dealing with transportation and mobility networks, Internet router connections, power grids, urban road networks and other systems where nodes are embedded in a metric space (Barthélemy 2011). In general, metric distance directly influences the network structure of such systems by imposing higher costs on the connections between distant entities. Social networks, instead, have been largely studied from a purely topological perspective, focusing on the structural position of their nodes and on structural mechanisms that describe their evolution.

Sociologists have studied the effect of distance on social ties with the underlying assumption that most individuals try to minimize the efforts to maintain a friendship link by interacting more with their spatial neighbors (Mok, Wellman, and Carrasco 2009; Goldenberg and Levy 2009). Nonetheless, the connection cost that heavily affects other types of spatial networks may not be as important in social systems, particularly when focusing on online interactions. As proposed by Cairncross (2001), distance may cease to play a role because of the increasing availability of affordable long-distance travel and new communication media, resulting in the inevitable "Death of Distance".

Social Ties and Geographic Distance

One fundamental spatial property of social networks is the probability $P(d)$ of having a social connection between two individuals as a function of their distance d. Even though there is universal agreement on the fact that $P(d)$ decreases with distance, the exact relationship between these two variables is still unclear.

Lambiotte et al. (2008) have found that it decays as $P(d) \sim d^{-2}$ in a mobile phone communication network, while Liben-Nowell et al. (2005) have found a different relationship $P(d) \sim d^{-1} + \epsilon$ among online bloggers on LiveJournal in the USA, ϵ being a constant probability which acts on online communities regardless of distance. In another study, Backstrom et al. (2010) have similarly found spatial scaling $P(d) \sim 1/d$ among online interactions: they show how this association appears so strong and important that it can be safely exploited to infer where Facebook users are only from the location of their friends (Backstrom, Sun, and Marlow 2010). It has also been proposed that the spatial structure of social networks might be scale-invariant, with a universal distribution $P(d) \propto d^{-1}$ (Hu et al. 2011): nonetheless, the exact nature of this relationship in a spatial social network often conveys interesting information about how geographic distance constrains social ties.

At the same time, the effect of online communication tools on such relationship is still under debate, even though initial results tend to confirm that distance is still an important factor that shapes human online interaction, with some individuals engaging preferentially with spatially close acquaintances across different online and offline communication media (Mok, Wellman, and Carrasco 2009; Goldenberg and Levy 2009; Scellato et al. 2010).

Analysis of Location-based Services

Location-based social platforms represent the ideal systems to investigate the spatial properties of social networks arising among online users for three main reasons. First of all, they uniquely provide data on both social connections and geographic locations, making socio-spatial analysis possible. Then, user location information in these services is often more accurate than text-based descriptions usually available in other online systems (Hecht et al. 2011), as it is acquired through sensing devices whenever users willingly *check-in*, that is when they share with their friends information about the place where they are. Finally, they have quickly accumulated hundreds of thousands, and sometimes millions, of users, thus enabling large-scale studies which can uncover general properties and trends.

Among the many research questions that arise, apart from the fundamental one about understanding the effect of distance on online relationships, there is the need to understand whether space homogeneously affects users or if, instead, some individuals prefer connecting to people further away, leading to a heterogeneous system. Moreover, social networks are often characterized by a large number of social clusters, where triads of individuals are mutually connected. Although social triangles seem to generally appear across different geographic scales (Lambiotte et al. 2008), different users may exhibit varied preferences towards short-range or long-distance triads (Scellato et al. 2010).

Spatial Properties of Online Social Networks

We will address these questions by analyzing three different popular LBSNs: Brightkite, Foursquare and Gowalla. We have collected data about all of them and extracted the social networks among their users. We are able to assign a "home location" to each user, in order to embed the nodes in a 2-dimensional metric space.

Then, we design two *randomized null models* of a spatial network which allow us to investigate the statistical significance of the empirical properties found in these networks. We observe *heterogeneity in the characteristic distance of interaction across users*, with some of them exhibiting preference towards short-range rather than long-distance ties.

In addition, we study the geographic properties of social triads. Again, we find non-trivial heterogenities across users, with some of them belonging mainly to geographically small triads and others to wider ones, spanning thousands of kilometres. In particular, *users with more friends tend to create triangles with individuals further apart* far more than expected by chance. We discuss how *user heterogeneity seems compatible with mechanisms akin to gravity models*, with the likelihood of connection between two users depending both on their popularity (i.e., number of friends) and on their distance.

This work constitutes the first large-scale study to investigate the spatial structure of online LBSNs, observing robust and universal properties across three of these social services: the observed features may be the signature of social processes happening regardless of the particular online tool adopted by users. While previous research has been focusing on defining new measures to take geographic distance into consideration when dealing with social networks (Scellato et al. 2010) and on exploiting simple socio-spatial properties to predict user location (Backstrom, Sun, and Marlow 2010), our contributions are different: we shed new light on how these socio-spatial properties arise from social and spatial factors and how user heterogeneity is related to both of these aspects. We believe location-based features will become ubiquitous in online social services: our findings may then inspire how systems and applications are designed and implemented on these services.

Data Collection

In this work we study three spatial social networks acquired from different popular online location-based social services. We extract the social networks arising among users and a single *geographic home location* for each user.

Brightkite Brightkite was founded in 2007 as a social networking website which allows users to share their location with their friends: it is available worldwide and it is based on the idea of making check-ins at places, where users can see who is nearby and who has been there before. Brightkite users can establish mutual friendship links and they can push their check-ins to their Twitter and Facebook accounts.

We study a dataset collected in September 2009 which includes the whole Brightkite user base at that time, with information about 54,190 users (Scellato et al. 2010). Since this dataset was collected, Brightkite has gathered more than 2 millions members: nonetheless, this dataset represents a complete snapshot of a popular location-based service in its initial evolution phase.

Foursquare Foursquare was created in 2009 and it has quickly risen as the most popular location-based service, with more than 6 million users as of January 2011. Users utilize the Foursquare application on their mobile devices, which allows them to check-in, sharing with their friends the place where they are. Foursquare provides game features, since the user with the highest number of check-ins in the last 60 days becomes the *mayor* of a place.

Acquiring Foursquare data requires user authorization to collect personal information and has rate limitations set in place. However, many Foursquare users choose to automatically push their check-in messages to Twitter, which provides a public API to search and download these messages.

Dataset	N	K	N_{GC}	$\langle k \rangle$	$\langle C \rangle$	D_{EFF}	$\langle D \rangle$	$\langle l \rangle$
Brightkite	54,190	213,668	50,896	7.88	0.181	5.73	5,651	2,041
Foursquare	258,706	2,854,957	254,532	22.07	0.191	5.90	8,494	1,442
Gowalla	122,414	580,446	117,361	9.48	0.254	5.44	5,663	1,792

Table 1: Properties of the datasets: number of nodes N and edges K in the social network, number of nodes in the giant connected component N_{GC}, average node degree $\langle k \rangle$, average clustering coefficient $\langle C \rangle$, 90-percentile effective network diameter D_{EFF}, average geographic distance between nodes $\langle D \rangle$ [km], average link length $\langle l \rangle$ [km].

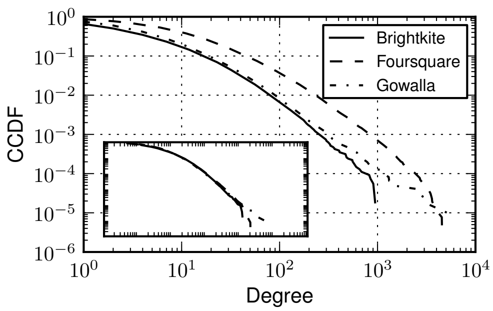


Figure 1: Empirical Complementary Cumulative Distribution (CCDF) of the number of friends in Brightkite, Foursquare and Gowalla. The inset shows the same distributions rescaled by dividing for the average number of friends in each network: the three datasets fall on the same curve.

Thus, we have recorded approximately 4 million tweets, each one containing a check-in sourced by a Foursquare user during June 2010. Those messages come from about 250,000 different users and cover about 1.5 million locations on the planet. We estimate that our sample contains approximately 20% to 25% of the entire Foursquare user base at collection time. Each tweet provides a URL to the Foursquare website, where information about the geographic location of the venue can be acquired. Since Foursquare does not provide unathorized access to user friends list, we have acquired friendship ties that Foursquare users have among them on Twitter, where they are publicly available, extracting a social network. While the resulting social graph is not expected to be identical to the original Foursquare graph, it provides a reasonable approximation and we will show how it conveys meaningful information, comparable to the other datasets. Finally, we extract as home location of each user the geographic location of the place where he/she has more check-ins overall.

Gowalla Gowalla is a location-based social network created in 2009: its users check-in at places through their mobile devices. Check-ins are shared with friends: as a consequence, friends can check where a user is or has been; conversely, it is possible to see all the users that have recently been in a given place. The friendship relationship is mutual, requiring each user to accept friendship requests to allow location sharing. However, there is a small number of user accounts that represent companies or other organizations and appear to automatically accept every friendship request. These accounts can become hugely popular and collect thousands of connections.

Gowalla provides a public API to let other applications integrate with their service: in particular, they provide information about user profiles, friend lists, user check-ins and place details. We have collected a complete snapshot of Gowalla data in August 2010. For every user we have gathered the user profile, the friends list and the list of all the check-ins the user has made. Finally, for each place we have collected its geographic location, as specified in Gowalla, described as a latitude-longitude pair. Since users are identified by consecutive numeric IDs, we were able to exhaustively query all user accounts. As in Foursquare, we define the home location of each user as the place with the largest number of check-ins.

Network Socio-spatial Properties

We first address the spatial properties of the social networks under analysis, focusing on the main topological and geographic measures. We discuss the fundamental relationship between likelihood of friendship and geographic distance and, finally, we define two randomized spatial networks models which will help later assessing the statistical significance of the properties we observe in these systems.

Socio-spatial properties

More formally, a *spatial social network* is a social network whose actors are positioned in a space equipped with a metric (Barthélemy 2011). In our case, online users are located over the 2-dimensional surface of the Earth and we adopt the great-circle distance as metric: the distance D_{ij} between any two nodes i and j is then computed given their geographic coordinates. Then, the social network can be represented as an undirected graph G with N nodes and K links, with users as nodes and where a link exists for each social tie (e.g., a person lists another user as one of his/her friends). We associate a length l_{ij} to each social link so that $l_{ij} = D_{ij}$.

The general properties of these three datasets are reported in Table 1. The social networks are heterogeneous in size, ranging from 54,190 nodes in Brightkite to 258,706 in Foursquare; the average degree is lower in Brightkite and Gowalla, respectively 7.88 and 9.48, than in Foursquare, where users have on average 22.07 friends. Thus, Foursquare presents a much denser and bigger social network, a consequence of its dominance of the LBSN market. All networks present a giant connected component which contains the vast majority of the users. The degree distributions for the three networks are reported in Figure 1:

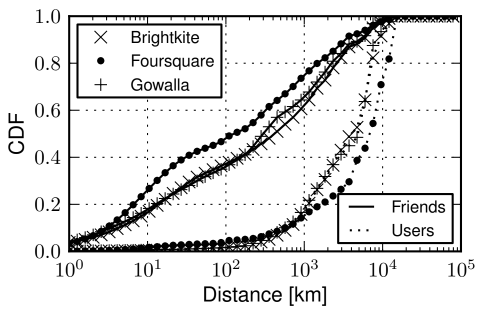


Figure 2: Empirical Cumulative Distribution (CDF) of the geographic distance between all users (dotted line) and between connected friends (solid line) for the three datasets.

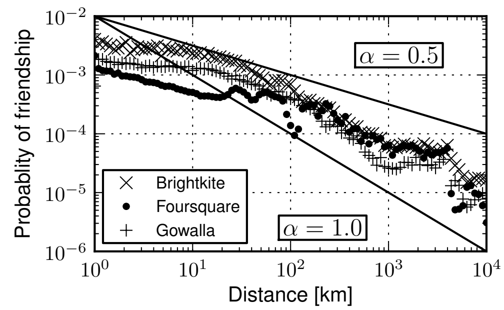


Figure 3: Probability of friendship between two users as a function of their geographic distance for the three datasets under analysis. The two straight lines represent probability $P(d) \sim d^{-\alpha}$ with two different exponents $\alpha = 0.5$ and $\alpha = 1.0$.

they all show a heavy-tail, with some users having thousands of friends. Rescaling the degree distributions by their average values results in a common trend, as shown in the inset. These networks also exhibit high values of average clustering coefficient, between 0.18 and 0.26, and short topological distances among their nodes, with 90% of all couples being less than 6 hops away. These properties confirm the small-world nature of LBSNs, as found in other online social systems (Leskovec and Horvitz 2008).

The average geographic distance between users $\langle D \rangle$ is consistently larger than the average distance between friends $\langle l \rangle$ across all the datasets: while the first value ranges between 5,600 and 8,500 km, the latter has much shorter values, between 1,400 and 2,000 km. This already provides evidence that the probability of having a social link between two users decreases with distance: we will further investigate this relationship later. The distribution of social link length is comparable across the three datasets, as shown in Figure 2: about 40%-50% of all couples of friends are within 100 km, with more than 3% of all links being shorter than 1 km. Instead, the distribution of distances among users, also shown in Figure 2, has a different behavior: about 50% of users are at distances larger than 4,000 km across all the networks.

Online Friendship and Distance

To further investigate how social links appear more likely to exist between close rather than distant users, we study the probability of friendship $P(d)$ as a function of distance d by counting L_d, the number of social links with length d, and by estimating N_d, the number of pairs of users at distance d. This gives us $P(d) = L(d)/N(d)$. As discussed before, this relationship has been found to be close to a law $P(d) \sim d^{-\alpha}$, with values of α ranging between 1 and 2 (Liben-Nowell et al. 2005; Lambiotte et al. 2008; Backstrom, Sun, and Marlow 2010; Goldenberg and Levy 2009).

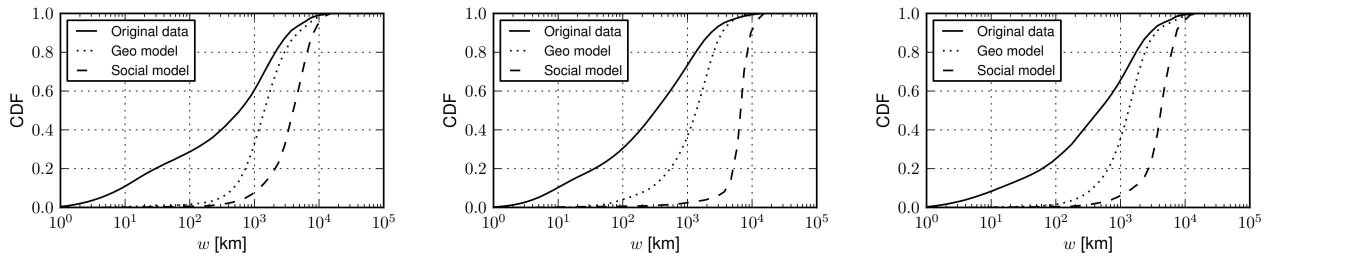
As shown in Figure 3, our datasets present noisy patterns and, furthermore, Brightkite and Gowalla exhibit an almost flat probability in the range 1-10 km, while all curves then decrease as distance grows and then they reach another steady probability between 1,000 and 4,000 km, maybe denoting a background probability that affects individuals within this distance threshold. Similar constant trends at short and long geographic ranges have also been found in other online systems (Backstrom, Sun, and Marlow 2010; Liben-Nowell et al. 2005). The appearance of social ties longer than 4,000 km becomes constrained by the fact that both Europe and North America, where a large part of users are based, are not large enough to allow such long-range connections and their mutual distance is about 6,000 km.

Surprisingly, we find that our data are closer to a law with $\alpha = 0.5$, whereas larger exponents have been found in other similar studies: hence, in LBSNs long-range social ties have a higher probability of occurrence than in other social systems. A potential explanation of this behavior is that LBSNs are relatively new, so they have mainly attracted early adopters. These users tend to be tech-savy, with many already existing long-distance online friendship ties which they bring to these services. This might not happen in other social networks, such as mobile phone networks or Facebook, which have already undergone through an initial phase and are already mature. Indeed, mobile phone connections show a larger exponent α than online social networks: phone conversations are much more constrained by geographic distance than interactions on Facebook. It might be possible that as location-based services become more mainstream their user audience may broaden and include individuals which are affected by distance in a stronger way.

Network Randomization

After these initial investigations, we will assess the statistical significance of the empirical spatial properties of these networks using two *randomized models*, which capture either the geographic or the social properties of the original social networks and randomize everything else:

- **Geo**: this null model keeps the user locations unmodified and then assigns every social link between two users at distance d according to the relative probability of friendship $P(d)$ (as reported in Figure 3).
- **Social**: this null model keeps the social connections as they are, shuffling at random all user locations.



(a) Brightkite (b) Foursquare (c) Gowalla

Figure 4: Empirical Cumulative Distribution (CDF) of the average friend distance w for each user in the social network, together with the distributions obtained in the randomized models.

The overall properties of these models are a direct consequence of their definition. Both models result in a network with exactly the same number of nodes and, on average, the same number of edges. The Social model has the same social properties of the original network, such as degree distribution, clustering coefficients and topological network distances, but link lengths are distributed as the pairwise user distances: as a result, the average link length becomes higher than in the original network, with $\langle l_{SOC} \rangle = \langle D \rangle$. On the other hand, the Geo model has the same distribution of link lengths of the original network, so that $\langle l_{GEO} \rangle = \langle l \rangle$, but the social properties are now lost: the degree distribution is peaked and has no heavy-tail, while the average clustering coefficient is much lower, since there are less social triads. Nonetheless, the two network models present similar distribution of topological distances, with 90% of all couples of nodes always within 6 hops.

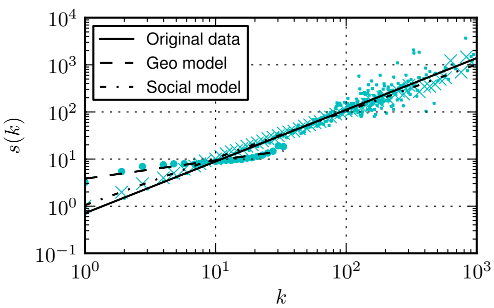
We will exploit these two null models in the following sections by comparing their properties to the ones of the real networks, in order to better understand whether the observed socio-spatial characteristics might be explained in terms of simple geographic or social factors. Every analysis performed on a randomized model will be averaged over 50 different realizations of the model itself.

Socio-spatial Properties of Individual Users

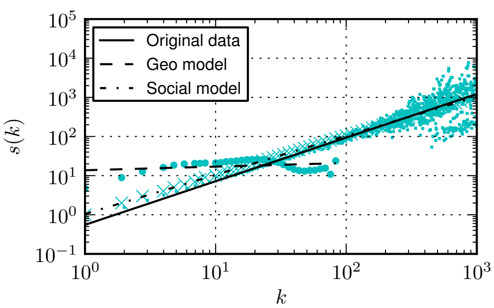
We now focus on individual users, studying how their social ties stretch across space. We define

$$w_i = \frac{1}{k_i} \sum_{j \in \Gamma_i} l_{ij} \qquad (1)$$

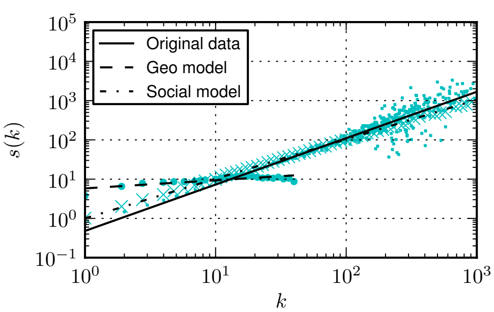
as the average friend distance of user i, where Γ_i is the set of neighbors of node i and $k_i = |\Gamma_i|$ is its degree. The overall distribution of w is reported in Figure 4 for the original social network and for the two randomized versions. The existence of values over all geographic scales is due to the existence of users with different characteristic lengths of interaction. For instance, about 10% of users have connections with an average length of just 10 km, whereas about 20% of users exhibit distances above 2,000 km. Since this distribution closely matches the aggregated link length distribution in Figure 2, links with different geographic lengths do not appear homogeneously across all users. Instead, there is



(a) Brightkite

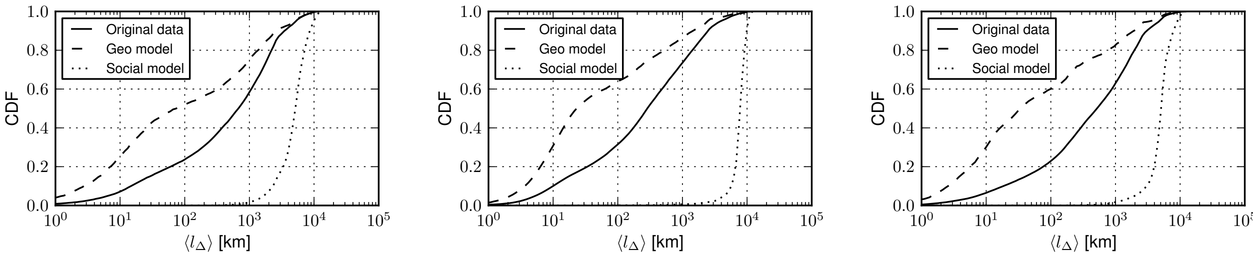


(b) Foursquare



(c) Gowalla

Figure 5: Average distance strength $s(k)$ as a function of node degree k for the original network and for its two randomized versions. Each trend is fitted by a law $s(k) \sim k^{\beta}$.



(a) Brightkite (b) Foursquare (c) Gowalla

Figure 7: Empirical Cumulative Distribution of average triangle link length $\langle l_\Delta \rangle$ for the original network and for the two randomized models.

heterogeneity between users, with some of them with only short-range connections and others with long-distance ties. These correlations are stronger than one would expect by chance: in fact, the two randomized models show how values of w should be more peaked around the average, instead of over a large range of magnitudes.

Another interesting result is obtained by studying the correlation between the average friend distance w_i and the degree k_i. We study the user *distance strength* (Barrat et al. 2004), defined as

$$s_i = \sum_{j \in \Gamma_i} l_{ij} = k_i w_i \quad (2)$$

and then we compute the average distance strength $s(k)$ for all users with degree k. In absence of any correlation, this measure should be scaling linearly with the degree $s(k) \sim k\langle l \rangle$, while a relation of the form $s(k) = Ak^{\beta}$ with $\beta \neq 1$ or $A \neq \langle l \rangle$ would imply correlation between the distance strength and the degree. In particular, $\beta > 1$ signals that users with more friends tend to have longer connections than users with fewer friends, while $\beta < 1$ would imply the opposite correlation, with users with more friends having shorter social links.

This relationship is reported in Figure 5 for the three datasets under analysis: we obtain values of β in the range 1.10-1.18 across the different networks, showing weak positive correlation. Real data show a pattern much closer to the Social model, which has $s(k) \sim k$, with $\beta = 1$, rather than to the Geo model, which instead has much lower values of β in the range $0.2 - 0.4$, denoting negative correlation between node degree and average friend distance. As users add more and more friends, on average their link length slightly increases, in contrast to what found in the null models, providing evidence that users with more connections tend to have friends further away.

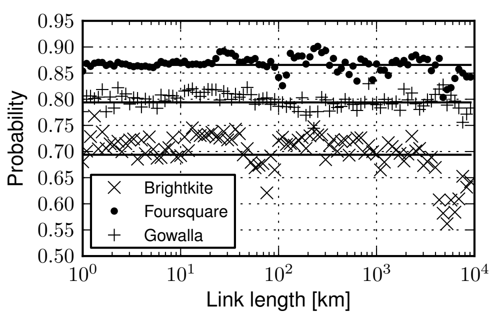


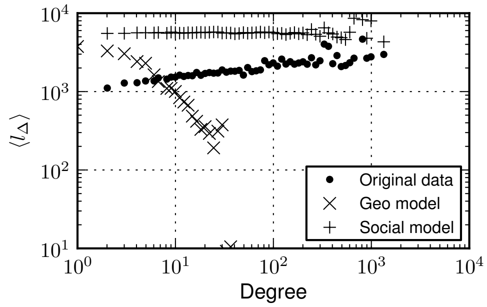
Figure 6: Probability that a social link belongs to a triangle as a function of its geographic length for the three datasets under analysis. The solid lines show the average probability that a link belongs to a triangle for each network.

Socio-spatial Properties of Social Triangles

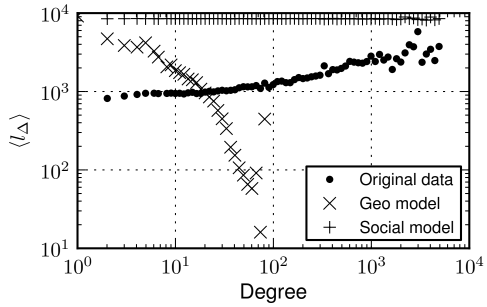
We now shift our attention to understanding the geographic properties of social triangles. Social networks usually present several triads, resulting in high values of clustering coefficient. Our networks also exhibit similar patterns, with clustering values between 0.18 and 0.26. We extract 377,438 triangles in the Brightkite social networks, 18,764,129 in Foursquare and 1,327,559 in Gowalla. Between 70% and 86% of all links in each social network belong to at least one triangle, given their highly clustered structure.

We find that social triangles arise at a wide range of geographic lengths: however, investigating the probability that a link belongs to a triangle as a function of its length provides a surprising result, since this probability is largely unaffected by distance, as shown in Figure 6. As a consequence, a link is equally likely to belong to a social triangle regardless of its length. A related result was found by Lambiotte et al. (2008): many spatially local clusters of people appear in mobile phone communication networks, with social links below 40 km more likely to belong to social triads, but then this likelihood reaches a constant value for longer links. As we have already seen, online behavior appears less sensitive to distance than mobile phone communication. Overall, the trend that longer links equally participate to social triangles holds also in our datasets.

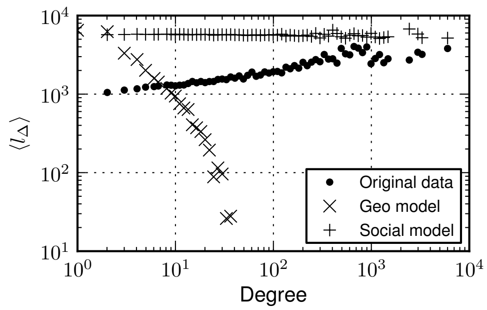
To assess user heterogeneity, we compute the geographic mean length l_Δ of the three links of each triangle and then we compute the average triangle geographic length $\langle l_\Delta \rangle_i$ for each user i by considering all the triangles he/she belongs to. This value does not take into account how many triangles users belong to, as the clustering coefficient does: instead, we aim to assess merely the geographic span of a user's social triangles, whatever their number might be. In Figure 7 we show the distribution of $\langle l_\Delta \rangle$ over all users: triangles



(a) Brightkite



(b) Foursquare



(c) Gowalla

Figure 8: Average triangle link length $\langle l_\Delta \rangle$ users with the same degree as a function of user degree for the original network and for the two randomized models.

with different geographic span are not equally arising among all users, but instead there are users with smaller triads and users with wider ones. For example, there are at least 20% of users with an average triangle length less than 100 km, while the top 20% have values above 2,000 km. This heterogeneity is much higher than one would expect if space did not matter, as the Social model shows mainly values in the range 1,000-10,000 km. Nonetheless, if social mechanisms were not taking place at all, then social triads should have been smaller, as the Geo model exhibits. The existence of both local, short-range triads and global, long-distance ones needs to be related to both the influence of geographic distance and of social processes such as homophily, triadic closure and focus constraint (McPherson, Lovin, and Cook 2001; Granovetter 1973; Feld 1981).

We further study this heterogeneity arising among users by computing the average $\langle l_\Delta \rangle$ as a function of the node degree. In these social networks $\langle l_\Delta \rangle$ increases with the number of friends, as shown in Figure 8. This effect is not present in the randomized networks: the Social model shows no correlation at all, while the Geo model exhibits the opposite trend, with smaller triangles appearing among users with higher degrees. Apparently, there are both social and geographic factors influencing social triangles, since having only one type of factors does not capture the empirical data.

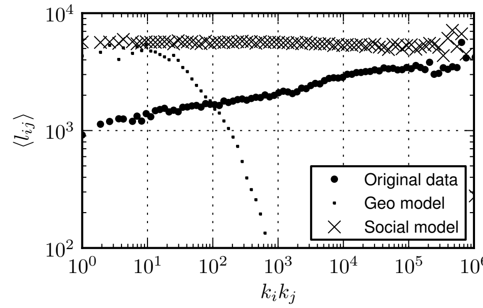


Figure 9: Average link length $\langle l_{ij} \rangle$ as a function of the product of the node degrees $k_i k_j$ for the original network and for its two randomized versions in Gowalla. The other two networks exhibit similar patterns.

This signals that users with less friends tend to generate social triangles on a smaller geographic scale, while users with many more friends belong to triangles with longer links. This suggests that there might be strong connection between the social properties of a given user and the geographic distance of his/her friendship connections.

Discussion

We have seen how users exhibit different characteristic geographic scales of online interaction, with weak positive correlation between number of friends and their average distance. Also, a similar heterogeneity appears with respect to social triads, with users participating in geographically wider triangles as their degree increases. Our findings are robust across the three LBSNs under analysis, as they arise regardless of the particular service we consider, the data collection methodology, the time elapsed since the creation of the service or the size of the social network. However, the properties we observe in the real systems are not appearing in the two randomized versions of these networks: therefore, *their socio-spatial structure can not be explained by taking into account only geographic factors or social mechanisms.*

Indeed, this claim can be further supported by considering the average length of a link l_{ij} as a function of the *product of the degrees* $k_i k_j$. As observed in Figure 9, longer links tend to arise between users with more friends, while links connecting users with fewer friends tend to be much shorter. This effect signals significant correlations between users social properties and their spatial behavior. In fact, it is not appearing at all in the Social model: this denotes how there might be an underlying spatial process taking place which results in this correlation, since social ties are not equally likely to appear regardless of their geographic length. On the other hand, the Geo model exhibits the opposite trend, with shorter links appearing mainly among well-connected users. Hence, distance is not the only factor affecting the link formation process: in other words, when only mechanisms which depend on geographic distance are in place, a user accumulates many friends only where there are many potential friends living nearby, i.e., if he/she is located in an area with high density of users. Furthermore, such geographic model can not explain how some users accumulate thousands of friends, creating a heavy-tail in the degree distribution.

A more accurate modelling of these networks requires the incorporation of processes mingling social and spatial factors. An interesting possibility is related to *gravity models*, which have long been used to model connections in spatial networks such as trade flows across countries (Bhattacharya et al. 2008). In this formulation the intensity of interaction between two spatial nodes i and j is proportional to $N_i N_j f(d_{ij})$, where N_i is the importance of node i, d_{ij} is their distance and $f(d)$ is a deterrence function which captures spatial effects. A gravity model balances the effect of spatial distance with other node properties: the underlying assumption is that longer (and more expensive) ties will appear mainly between important entities, while a node will connect to an unimportant one only if they are close to each other.

Gravity models are only a first tentative step, as they are expected to fail at reproducing some of our observations. In particular, gravity models only focus on pairs of nodes, without taking into account social effects such as triadic closure and focus constraint (Granovetter 1973; Feld 1981). Furthermore, any notion of "node importance" in a social network appears vague and uncertain, thus making the definition of a sound social gravity model hard to specify. Such individual importance may be an exogenous variable which affects the socio-spatial structure, such as being a well-known celebrity or any other type of individual popularity or social influence measure.

As people spend more time online, more and more data will be available regarding their spatial behavior and their social connections, allowing more reliable and precise models to be built. Such models present many potential applications in the design of any type of location-based service, but also important implications for other systems such as security mechanisms, user identification techniques and recommendation engines.

Conclusions and Future Work

In this work we have studied the socio-spatial properties of users of location-based services. Our methodology is based on two randomized null models and highlights how observed properties deviate from what would be expected by chance with purely social or geographic mechanisms. We find that LBSNs present universal spatial features across them, regardless of the service, its number of users or the adopted sampling method. We observe and discuss heterogeneity in user socio-spatial behavior: users exhibit friendship connections across a wide range of geographic distance, showing similar variability in the social triads they belong to.

An interesting direction for further work is understanding how such heterogeneity arises in correlation with the temporal evolution of the social network, as users spend more time on the service. This may lead to a better understanding of the generative mechanisms behind these properties and to new predictive models.

References

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.

Barrat, A.; Barthélemy, M.; Pastor-Satorras, R.; and Vespignani, A. 2004. The architecture of complex weighted networks. *PNAS* 101(11):3747–3752.

Barthélemy, M. 2011. Spatial Networks. *Physics Reports* 499:1–101.

Bhattacharya, K.; Mukherjee, G.; Saramäki, J.; Kaski, K.; and Manna, S. S. 2008. The International Trade Network: weighted network analysis and modelling. *Journal of Statistical Mechanics* 2008(02):P02002+.

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

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

Goldenberg, J., and Levy, M. 2009. Distance Is Not Dead: Social Interaction and Geographical Distance in the Internet Era. *ArXiV* abs/0906.3202.

Granovetter, M. S. 1973. The strength of weak ties. *The American Journal of Sociology* 78(6):1360–1380.

Hecht, B.; Hong, L.; Suh, B.; and Chi, E. H. 2011. Tweets from Justin Bieber's Heart: The Dynamics of the "Location" Field in User Profiles. In *Proceedings of CHI '11*.

Hu, Y.; Wang, Y.; Li, D.; Havlin, S.; and Di, Z. 2011. Possible Origin of Efficient Navigation in Small Worlds. *Physical Review Letters* 106(10):108701+.

Lambiotte, R.; Blondel, V.; Dekerchove, C.; Huens, E.; Prieur, C.; Smoreda, Z.; and Vandooren, P. 2008. Geographical dispersal of mobile communication networks. *Physica A* 387(21):5317–5325.

Leskovec, J., and Horvitz, E. 2008. Planetary-scale views on a large instant-messaging network. In *Proceedings of WWW' 08*.

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

McPherson, M.; Lovin, L. S.; and Cook, J. M. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27(1):415–444.

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

Scellato, S.; Mascolo, C.; Musolesi, M.; and Latora, V. 2010. Distance Matters: Geo-social Metrics for Online Social Networks. In *Proceedings of WOSN'10*.