Far from the Eyes, Close on the Web: Impact of Geographic Distance on Online Social Interactions

Andreas Kaltenbrunner[*]
kaltenbrunner@gmail.com

Salvatore Scellato[†]
salvatore.scellato@cl.cam.ac.uk

Yana Volkovich[*]
yana.volkovich@barcelonamedia.org

David Laniado[*]
david.laniado@barcelonamedia.org

Dave Currie[‡]
dave@tuenti.com

Erik J. Jutemar[‡]
ejutemar@tuenti.com

Cecilia Mascolo[†]
cecilia.mascolo@cl.cam.ac.uk

[*]Barcelona Media Foundation
Barcelona, Spain

[‡] Tuenti
Madrid, Spain

[†] Computer Laboratory
University of Cambridge, UK

ABSTRACT

Online friendship connections are often not representative of social relationships or shared interest between users, but merely provide a public display of personal identity. A better picture of online social behaviour can be achieved by taking into account the intensity of communication levels between users, yielding useful insights for service providers supporting this communication. Among the several factors impacting user interactions, geographic distance might be affecting how users communicate with their friends. While spatial proximity appears influencing how people connect to each other even on the Web, the relationship between social interaction and spatial distance remains unexplored.

In this work we analyse the relationship between online user interactions and geographic proximity with a detailed study of a large Spanish online social service. Our results show that while geographic distance strongly affects how social links are created, spatial proximity plays a negligible role on user interactions. These findings offer new insights on the interplay between social and spatial factors influencing online user behaviour and open new directions for future research and applications.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data mining*; H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web based services*

Keywords

social networks, geo-spatial data, wall interactions

WOSN'12, August 17, 2012, Helsinki, Finland.

1. INTRODUCTION

Online social networks (OSN) have become the most popular destination for Web users, sparking off related systems and applications that take advantage of the data generated by user interactions to offer better recommendations, better tailored advertising or, simply, to promote commercial brands to devoted supporters.

The structural properties of the social graphs arising among users are of great interest, in particular as they influence the traffic load that service providers experience: hence, many studies have analysed these properties [1, 9]. Some of these works shed light on whether user behaviour is purely social or, instead, more influenced by other non-social factors, resulting in online behaviour appearing different than what is observed in "offline" real-life social ties [10, 15]. In particular, not all the online ties declared by users on OSN are the same: even if some users have hundreds of connections, due to the finite amount of resources available, such as time [14], communication tends to be biased towards those relationships that are deemed more important [4].

As in real life, where tie strength is an extremely important facet of social interactions and where weak ties with "familiar strangers" often appear predominant [7, 13], *online friendship connections exhibit heterogeneous intensity*, with a large fraction of users interacting mainly with a small subset of acquaintances [8, 23]. In addition, social ties established online are often carefully chosen and displayed by users to represent their status and identity, supporting the hypothesis that social links often fail to signal real social proximity, mutual trust or even shared interest [5]. Failing to take these factors into account when studying the development of online social interactions one is unlikely to uncover the true social properties of these platforms.

A more recent but equally important development is the increasing offer of location-aware services by OSN. They give access to a new layer of spatial information about where users live and where they go, which has ignited works addressing the effect of geographic distance on social ties [2, 12, 16, 22]. Recent results show how geographic distance still matters even in OSN: *users tend to connect preferentially with spatially close acquaintances rather than with individ-*

Table 1: **Network properties: number of non-isolated nodes N and edges K, size of the giant component GC, average degree $\langle k \rangle$, average local clustering coefficient $\langle C \rangle$, 90-percentile effective network diameter d_{eff}, maximal distance d_{max} between two nodes in the network. Average path-length between nodes $\langle d \rangle$, geographic distance between nodes $\langle D \rangle$ [km] and link length $\langle l \rangle$ [km] over all pairs of nodes. Values for the wall network correspond to the undirected dense version of the network.**

Network	N	K	size GC	$\langle k \rangle$	$\langle C \rangle$	d_{eff}	d_{max}	$\langle d \rangle$	$\langle D \rangle$	$\langle l \rangle$
Friendship	9 356 588	587 415 363	99.47%	126	0.200	5.8	9	5.2	531.2	98.9
Wall	6 487 861	111 503 001	99.56%	34	0.137	6.8	10	6.1	531.2	79.9

uals further away [11, 2, 19]. Hence, the first law of geography seems to hold even on OSN: *"everything is related to everything else, but near things are more related than distant things"* [20].

Our work.

Given how social links on online networking platforms are likely to represent a wide range of social interaction levels, and given that the effect of geographic distance on such online social networks appears present but still not fully understood, the main research question we address in this work is: *are actual online social interactions affected by geographic distance, with high-intensity social relationships more constrained than weaker ties?* This question has important implications for service providers, because the availability of geographic information for popular online social networks opens unprecedented opportunities to enhance engineering of world-wide systems based on human communication and interaction, as demonstrated by some initial recent attempts [2, 18, 24].

We aim to address this question through a detailed study of the large-scale social network Tuenti, which is widely popular in Spain. We have access to an anonymised dataset of the full social network among Tuenti members, to their online interactions with each other and to their home locations, discretised across more than 7 000 Spanish cities. Our results support the idea that geographic distance strongly affects the friendship connections that users establish on online social networks: however, *the intensity of interaction on social ties seems unaffected by distance*, with negligible differences in how users interact with nearby friends and friends far away. Furthermore, even though users tend to allocate their interactions in a highly skewed way, sending a large fraction of their messages to few important friends, geographic distance does not play a strong role in this allocation. This finding supports the idea that geography affects *whom* we interact with, but it does not influence *how much* we interact.

In [22] we furthermore explore this finding in relation to the structural position of friendship-ties in the network.

2. DATASET

In this section we present the dataset we use to study the effect of geographic distance on online social interactions and introduce the notation we will use throughout our work.

2.1 Tuenti

We analyse a large sample of Spanish, invitation-only (at the time our dataset was extracted) social networking service, Tuenti[1]. Founded in 2006, thanks to its widespread popularity in the country, Tuenti is now sometimes referred to as the "Spanish Facebook". Tuenti provides many features common to other popular social networking platforms: it allows users to set up a profile, specify the location where they live, connect with friends, share web links and media items and write on each other's walls. Our dataset is based on an anonymised snapshot of Tuenti's friendship connections as of November 2010. It includes about 9.8 million registered users (9.35 million with at least one friend), more than 580 million friendship links, about 500 million interactions (via message exchanges) during a 3 months period and the user's the self-reported city of residence (selected from a predefined list). Tuenti users have an incentive to specify the real city where they live, to be discoverable by other potential friends. The location where users live is an important discriminator when a search returns a list of users with the same name.

[1]www.tuenti.com

2.2 Notation

A goal of our work is to study how social interactions is related to users' geographic locations. These interactions either correspond to explicitly declared connections such as friendship links in a social network or implicit ones retrieved from interactions via wall comments. We note that Tuenti only allows users that are friends to exchange wall messages: thus, we can model the social network among Tuenti users as a directed weighted graph $G = (V, E)$, where nodes are users and edges are friendship connections on Tuenti. We refer to this graph as the *friendship network*.

The weight w_{ij} of the edge from user i to user j is equal to the number of messages user i posted on the wall of user j: in general $w_{ij} \neq w_{ji}$. Two users may be connected to each other but never exchange a message, hence $w_{ij} \geq 0$. If we remove all the edges with $w_{ij} = 0$ and all nodes which have not sent nor received any message, we are left with a smaller *wall network*. Furthermore, we define d_{ij} as the geographic great-circle distance between the cities of residence of user i and user j: we define $d_{ij} = 0$ if they report the same city of residence. In Table 1 we report the main properties for both the friendship and wall networks.

2.3 Social properties

In Figure 1 we plot the distribution of the number of friends in the friendship network. We see a peak at 1 000 friends, that is a friendship limit defined by Tuenti. Nonetheless, Tuenti opens this limit occasionally for users with special merits (e.g. celebrities).

Recall that by interaction we mean a post written by a user on the wall of a friend. Hence, a given user will interact with a subset of friends, while having no interactions at all with the remaining portion. In Figure 2 we show the average fraction of friends and the average absolute number of friends a user interacts with as a function of the number

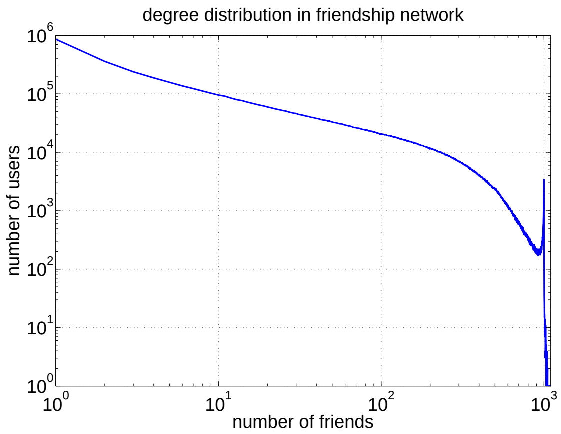

Figure 1: The friendship degree distribution.

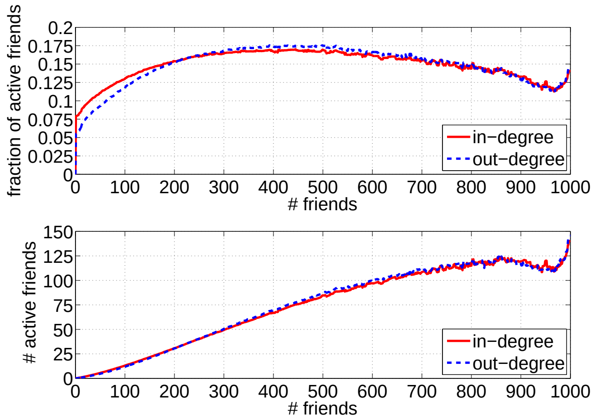

Figure 2: Fraction and number of friends users interact with as a function of friendship degree.

of friends. Surprisingly, as the fraction of friends a user interacts with initially increases for users with more friends, it quickly reaches a plateau and then it slightly decreases for users with more than 500 friends, denoting how additional friendship links are unlikely to generate high levels of interaction. In particular, we observe that the absolute number of active connections never exceeds 150 users. This is in perfect agreement with Dunbar's number [6], an alleged theoretical cognitive limit to the number of people with whom one can maintain stable social relationships.

3. GEOGRAPHIC PROPERTIES

In this section we analyse the spatial properties of the Tuenti social network.

3.1 Friendship and distance

As found in many other online social networks [2, 19], Tuenti users tend to preferentially connect to closer users. In fact, as depicted in Figure 3, the distribution of geographic distance between connected users shows much lower values than for random pairs of users (i.e. potential friendships). About 60% of social links between users are at a distance of

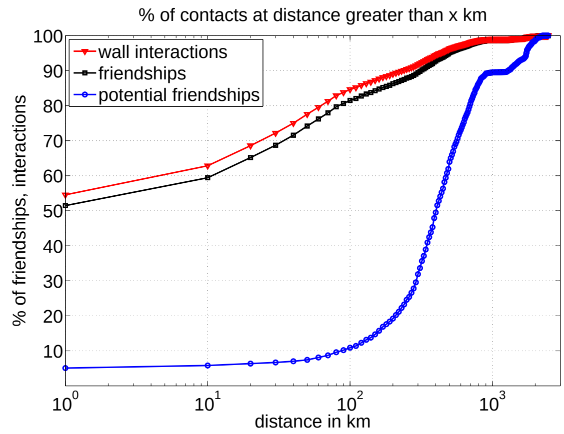

Figure 3: Cumulative distribution function (CDF) of geographic distance of social links, interaction links and all pairs of users.

10 km or less, while only 10% of all distances between users are below 100 km. If we restrict this analysis to the wall network we see a similar trend, though with slightly shorter distances.

3.2 The effect of distance

A better way to assess the constraining effect of geographic distance on social ties is to compute the probability that any two individuals are connected as a function of their spatial distance. Since the fraction of short-range social links is high, and since there are many more users at a large distance than close by, the probability of connection must be decreasing with distance.

In Figure 4 we plot this probability together with the probabilities that any two users interact more than a certain number of times. We observe a strong effect of distance d on the probability of connection $P(d)$: while the probability has a flat trend below 30 km, it quickly decreases as $d^{-\alpha} + \epsilon$, with $\alpha \approx 1.8$. The constant value ϵ becomes non-negligible only at large distance, denoting a constant background probability of connection between individuals that does not seem affected by distance. Similar patterns containing a constant offset, although with different exponents, have been also found on other online social networks [12, 2].

To our surprise, the same functional form of the probability of connection $P(d)$ does not change when we remove links with an interaction weight $w_{i,j}$ lower than a threshold θ. We observed a power-law decay $d^{-\alpha} + \epsilon$ with similar exponents α even for different values of θ: the only difference we notice is in the initial constant value of the probability for distances below 30 km and in the final constant ϵ, which decreases as we increase the threshold θ. These results suggest that *while distance strongly constrains how social links are established, there seems to be only a uniform effect on all user interactions, unrelated to the geographic length they span.*

4. INTERACTION ANALYSIS

In this section we focus on the spatial properties of user interactions.

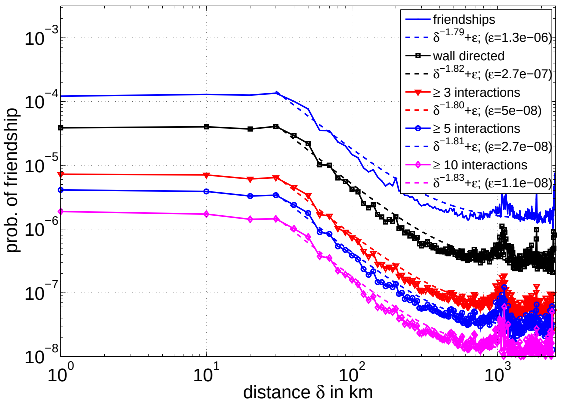

Figure 4: Probability of friendship and of wall interaction between two users as a function of their geographic distance.

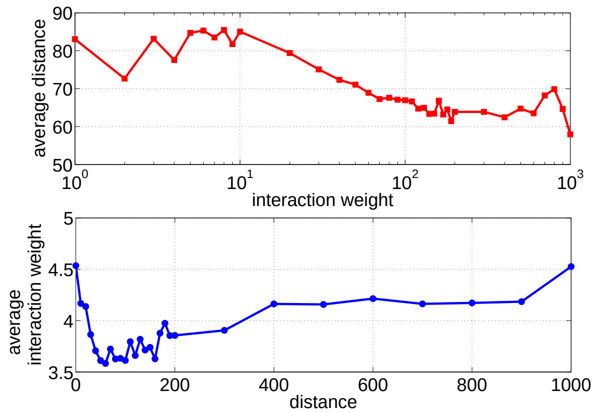

Figure 6: Relations between the number of interactions and spatial distances.

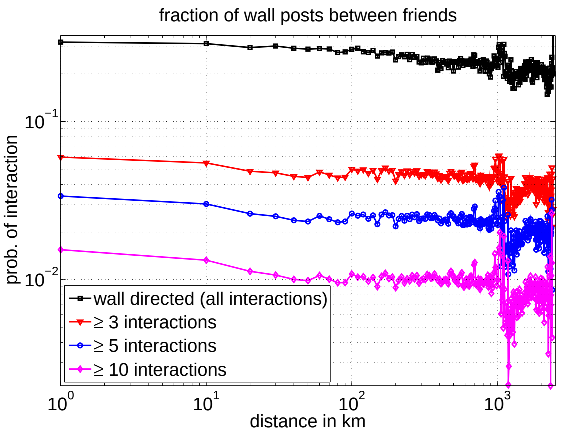

Figure 5: Probability of interaction with a friend as a function of geographic distance for the weighted wall network and for the thresholded networks.

4.1 Interactions and distance

As discussed in the previous section, Figure 4 provides evidence that the probability of connection between individuals is affected by geography in the same way across different levels of user interaction. In other words, it seems that there are two processes taking place. One process, strongly affected by geographic distance, influences how users connect to each other, i.e. their friendship links; another process impacts the level of interaction among connected users and appears unrelated to spatial proximity.

In order to better understand the relationship between social interactions and spatial distance we compute a different property: the probability that a message is exchanged over an existing social link as a function of geographic distance. If spatial distance affects interactions as it affects social ties, then we would expect another relationship with a strong decay: to our surprise, this is not the case. In fact, as highlighted in Figure 5, which depicts the probabilities of interaction between any pair of friends at a given distance. These probability ranges between 0.35 and 0.15 even when geographic distances increases from 0 to 1 000 km. Moreover, if we consider only links with increasingly larger interaction weights we see that the large-distance tail becomes flatter: high-intensity communication takes place on social connections regardless of their geographic distance. Thus, even if we see a decreasing trend, *geographic constraints on online interactions do not appear nearly as strong as for social connections.*

The analysis of individual social links conveys the same message: the number of messages sent over a certain social link exhibits only a weak dependence on the geographic length of the link itself, as shown in Figure 6. The average number of interactions between two users is unrelated to their geographic distance and, at the same time, the average distance between two individuals is only slightly related to the number of messages they exchange. We observe that there is a slight decay from an average distance of around 90km for a lower number of interactions to 70km if the users interact more than 90 times. Nevertheless, both indicators are remarkably stable, supporting the hypothesis that *while geographic distance heavily influences how users establish social connections, its effect on social interactions is only weak.* In other words, once users choose with whom they will interact, spatial factors are not important any more.

4.2 User properties

To identify how different users are affected by geographic distance, we adopt a methodology based on distance strength.

The distance strength was introduced in [3] as a measure of correlation between the degree of a node and the geographic distance of its links in spatial networks. We modify the original definition for the case of directed weighted networks. Thus, for every user i we compute two **directed and weighted distance strengths**:

$$s_i^{in} = \sum_i w_{ji} d_{ji} \qquad s_i^{out} = \sum_i w_{ij} d_{ij},$$

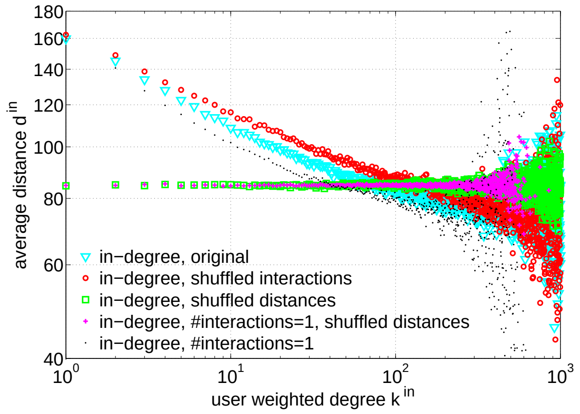

Figure 7: Average weighted distance d^{in} as a function of user weighted in-degree k^{in}: results are shown as well for null models.

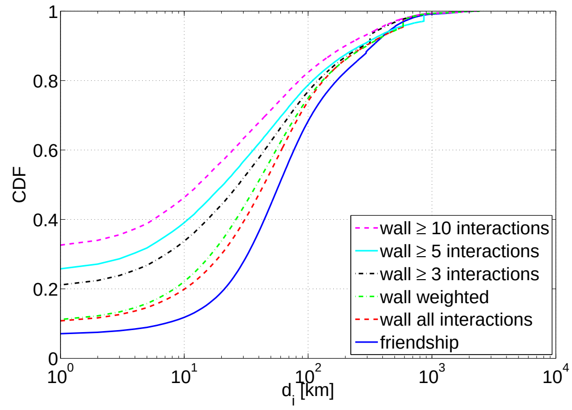

Figure 8: Distribution (CDF) of the average friend or wall-interaction distance d_i for each user. Different curves for the wall-interaction distances reflect the results when thresholds are applied on the amount of user interaction.

where as before w_{ij} is the number of interactions from user i to user j and d_{ij} is the distance between users i and j. In the absence of any correlation these measures should scale linearly, respectively, with the weighted in- and out-degree, i.e. $k_i^{in} = \sum_j w_{ji}$ and $k_i^{out} = \sum_j w_{ij}$. We also introduce the **average directed weighted distances**: $d_i^{in} = s_i^{in}/k_i^{in}$ and $d_i^{out} = s_i^{out}/k_i^{out}$, where k_i^{in} and k_i^{out} are non-zero. Again, these values should be unrelated to the degrees in absence of the correlation.

To contrast the original Tuenti data we introduce null models: we maintain the network structure as in the original Tuenti graph but shuffle either the interaction or the distance weights, destroying any existing correlation. As baseline models we also consider models where all interaction weights are set to 1. Figure 7 plots the weighted average distance of incoming interactions k^{in} versus the number of incoming interactions. We notice that as users have more and more incoming interactions the average weighted distance goes down: this does not happen when shuffling spatial distances in the null models. Even neglecting interaction weights the correlation remains strong, confirming that users with more friends have also shorter links. We found similar results for the out-degree version of the distance strength. Overall, users with a higher number of friends tend to have their interactions on spatially shorter social ties.

This finding is confirmed when looking at the distribution of values for the average friend distance, and its weighted version, across all users. Figure 8 shows that as we threshold the graph more and more, keeping only links with higher levels of interaction, the probability of lower average distances increases . In other words, *while online interactions on individual links do not appear affected by spatial distance, individual users with more interactions tend to have short-range links.* A potential explanation for this behaviour might be that more active users, with a greater number of friends, could be younger individuals, which are notoriously highly active on online social services. This kind of people could exhibit a noticeable propensity to interact more with friends living nearby. Instead, older users might exhibit more long-range connections because those were established between individuals when they were close in the past. Yet, the true reason behind such phenomenon deserves further investigation.

5. CONCLUSIONS

In this paper we have presented a study on the effect of geographic distance on online social interactions. We have analysed data collected from Tuenti, a Spanish social service with millions of users, containing information about social links and messages exchanged. While spatial proximity greatly affects how users establish their connections on online social platforms, we have found that social interactions are only weakly affected by distance: this suggests that once social connections are established other factors may influence how users send messages to their friends. On the other hand, more active users tend to preferentially interact over short-range connections. Although our dataset is restricted to a single country, we expect the result to be valid as well in an international setting.

There are many implications of our results. First of all, while users tend to have fewer long-range connections, the level of interaction can be as high on these ties as on short-range ones. This observation is crucial for architectures that optimise distributed storage of data related to online social platforms based on users' geographic locations [17]. Similarly, it is important for systems that exploit geographic locality of interest to serve content items requested through online social network services [18, 21]. In all these scenarios, our results suggest that while distant friends are rare, their social connections equally generate traffic load.

In addition, our findings are also likely to help other domains such as link prediction, tie strength inference and user profiling: the observed spatial patterns can be also included in security mechanisms to detect malicious and spam accounts [2]. Currently, storage solutions adopted by Tuenti are optimised using techniques that directly take advantage of the self-reported users' locations. Their main aim is to cluster together data related to users living in the same geographic area, because online friendship ties are also formed around geographical areas. A future goal is to further improve such architecture by replicating data generated over long-distance social ties across multiple storage locations.

6. ACKNOWLEDGEMENTS

This work was partially funding by the Centro para el Desarrollo Tecnológico Industrial (CDTI). Yana Volkovich acknowledges support from the Torres Quevedo program from the Spanish Ministry of Science and Innovation, co-funded by the European Social Fund.

7. REFERENCES

[1] Y. Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. In *Proceedings of WWW' 07*. ACM, 2007.

[2] L. Backstrom, E. Sun, and C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of WWW '10*. ACM, 2010.

[3] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *PNAS*, 101(11):3747, 2004.

[4] K. Dindia and D. J. Canary. Definitions and theoretical perspectives on maintaining relationships. *Journal of Social and Personal Relationships*, 10(2):163–173, 1993.

[5] J. Donath and D. Boyd. Public Displays of Connection. *BT Technology Journal*, 22(4):71–82, 2004.

[6] M. Gladwell. *The tipping point: How little things can make a big difference*. Little, Brown and Company, 2000.

[7] M. S. Granovetter. The strength of weak ties. *Am. J. of Sociology*, 78(6):1360–1380, 1973.

[8] J. Jiang, C. Wilson, X. Wang, P. Huang, W. Sha, Y. Dai, and B. Y. Zhao. Understanding latent interactions in online social networks. In *Proceedings of IMC '10*. ACM, 2010.

[9] R. Kumar, J. Novak, and A. Tomkins. Structure and Evolution of Online Social Networks. In *Proceedings of KDD '06*. ACM, 2006.

[10] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *Proceedings of WWW' 10*. ACM, 2010.

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

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

[13] S. Milgram. *The familiar stranger: an aspect of urban anonymity*. Addison-Wesley, 1977.

[14] G. Miritello, E. Moro, and R. Lara. Dynamical strength of social ties in information spreading. *Phys. Rev. E*, 83:045102, 2011.

[15] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and Analysis of Online Social Networks. In *Proceedings of IMC '07*. ACM, 2007.

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

[17] J. M. Pujol, V. Erramilli, G. Siganos, X. Yang, N. Laoutaris, P. Chhabra, and P. Rodriguez. The little engine(s) that could: scaling online social networks. In *Proceedings of SIGCOMM'10*, 2010.

[18] S. Scellato, C. Mascolo, M. Musolesi, and J. Crowcroft. Track Globally, Deliver Locally: Improving Content Delivery Networks by Tracking Geographic Social Cascades. In *Proceedings of WWW'11*. ACM, 2011.

[19] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo. Socio-Spatial Properties of Online Location-Based Social Networks. In *Proceedings of ICWSM'11*. AAAI, 2011.

[20] W. R. Tobler. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46:234–240, 1970.

[21] S. Traverso, K. Huguenin, V. Trestian, I. Erramilli, N. Laoutaris, and K. Papagiannaki. TailGate: Handling Long-Tail Content with a Little Help from Friends. In *Proceedings of WWW'12*, 2012.

[22] Y. Volkovich, S. Scellato, D. Laniado, C. Mascolo, and A. Kaltenbrunner. The length of bridge ties: structural and geographic properties of online social interactions. In *Proceedings of ICWSM'12*. AAAI, 2012.

[23] C. Wilson, B. Boe, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. In *Proceedings of EuroSys '09*. ACM, 2009.

[24] M. P. Wittie, V. Pejovic, L. Deek, K. C. Almeroth, and B. Y. Zhao. Exploiting locality of interest in online social networks. In *Proceedings CoNEXT '10*. ACM, 2010.