

# Far from the eyes, close on the Web: impact of geographic distance on online social interactions

Andreas Kaltenbrunner<sup>\*</sup>  
kaltenbrunner@gmail.com

Yana Volkovich<sup>\*</sup>  
yana.volkovich@barcelonamedia.org

Dave Currie<sup>‡</sup>  
dave@tuenti.com

Erik J. Jutemar<sup>‡</sup>  
ejutemar@tuenti.com

Salvatore Scellato<sup>†</sup>  
salvatore.scellato@cl.cam.ac.uk

David Laniado<sup>\*</sup>  
david.laniado@barcelonamedia.org

Cecilia Mascolo<sup>†</sup>  
cecilia.mascolo@cl.cam.ac.uk

<sup>\*</sup> Barcelona Media Foundation  
Barcelona, Spain

<sup>‡</sup> Tuenti  
Madrid, Spain

<sup>†</sup> Computer Laboratory  
University of Cambridge, UK

## ABSTRACT

Online friendship connections between users often tend to be not representative of social relationships or shared interest, but merely provide a public display of personal identity. A better picture of online social behavior 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 analyze 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 behavior and open new directions for future research and applications.

## 1. INTRODUCTION

Online social platforms 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 analyzed these properties [9, 1]. Some of these works shed light on whether user behavior is purely social or, instead, more influenced by other non-social factors, resulting in online behavior appearing different than what is observed in “offline” real-life social ties [15, 10]. In particular, not all the online ties declared by users on social platforms 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 [20, 8]. 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 online social networks. Hence, this gives access to a new layer of spatial information about where users live and where they go: this has ignited a new thread of works addressing the effect of geographic distance on social ties [12, 2, 16]. Recent results show how geographic distance still matters even in online social platforms: *users tend to connect preferentially with spatially close acquaintances*

rather than with individuals further away [11, 2, 18]. Hence, the first law of geography seems to hold even on online social networks: “*everything is related to everything else, but near things are more related than distant things*” [19].

### 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, 21, 17].

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 anonymized dataset of the full social network among Tuenti members, to their online interactions with each other and to their home locations, discretized 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 close 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.

## 2. DATASET

In this section we present the dataset under analysis 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 analyze a large sample of Spanish, invitation-only social networking service, Tuenti<sup>1</sup>. 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

<sup>1</sup>www.tuenti.com

to other popular social networking platforms: it allows users to set up a profile, connect with friends, share web links and media items and write on each other’s walls. Our dataset is based on an anonymized snapshot of Tuenti’s friendship connections as of November 2010. It includes about 9.8 million registered users, more than 580 million friendship links and about 500 million interactions (via message exchanges) during a 3 months period. For every user we have the self-reported city of residence selected from a predefined list.

### 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.

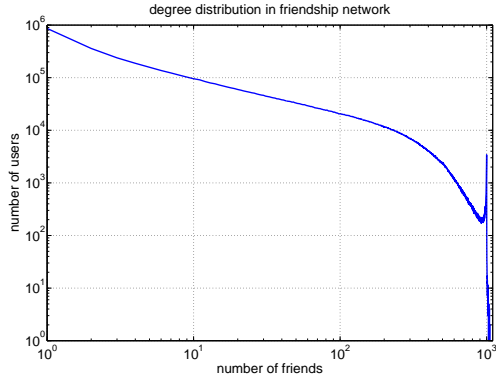
### 2.3 Social properties

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, there are still few users that manage to evade this limit.

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 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 num-

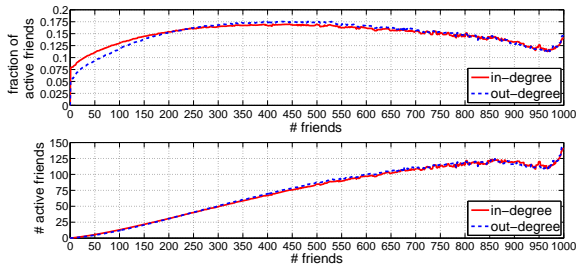
Network	$N$	$K$	size GC	$\langle k \rangle$	$\langle C \rangle$	$d_{eff}$	$d_{max}$	$\langle d \rangle$	$\langle D \rangle$	$\langle l \rangle$
Friendship	9 769 102	587 415 363	99.47%	126	0.200	5.8	9	5.2	531.2	98.9
Wall	6 487 861	111 503 001	99.56%	34	0.137	6.8	10	6.1	531.2	79.9

**Table 1: Properties of the networks: number of nodes  $N$  and edges  $K$ , size of the giant connected component GC, average node degree  $\langle k \rangle$ , average 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  $\langle d \rangle$ , average geographic distance between nodes  $\langle D \rangle$  [km], average link length  $\langle l \rangle$  [km].**



**Figure 1: Degree distribution in the Tuenti social network.**

ber of active connections never exceeds 150 users. This value is in perfect agreement with Dunbar’s number [6], which is an alleged theoretical cognitive limit to the number of people with whom one can maintain stable social relationships.



**Figure 2: The fraction and the number of friends users interact with as a function of the number of friends.**

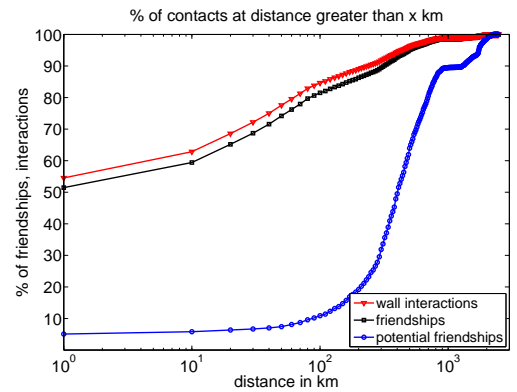
### 3. GEOGRAPHIC PROPERTIES

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

#### 3.1 Friendship and distance

As found in many other online social networks [2, 18], 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 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.

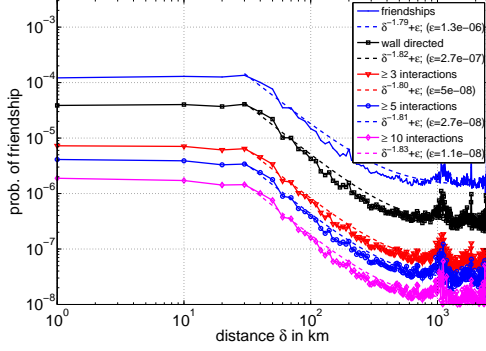


**Figure 3: Cumulative distribution function (CDF) of geographic distance of social links, interaction links and all pairs of users.**

#### 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 two individuals are connected as a function of their spatial distance. Since the fraction of short-range social link 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 fact, in Figure 4 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, then it quickly decreases as  $d^{-\alpha + \epsilon}$ , with  $\alpha \approx 1.8$ . The constant value  $\epsilon$  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].



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

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  $\theta$ . We observe a power-law decay  $d^{-\alpha} + \epsilon$  with similar exponents  $\alpha$  even for different values of  $\theta$ : the only difference we notice is in the initial constant value of the probability for distances below 30 km and in the final constant  $\epsilon$ , which decreases as we increase the threshold  $\theta$ . 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.*

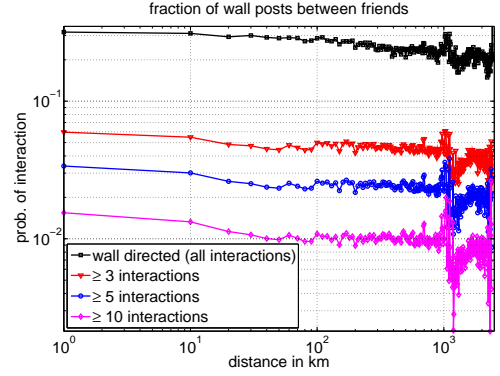
## 4. INTERACTION ANALYSIS

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

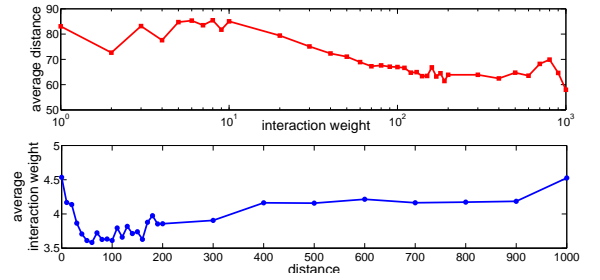
### 4.1 Interactions and distance

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



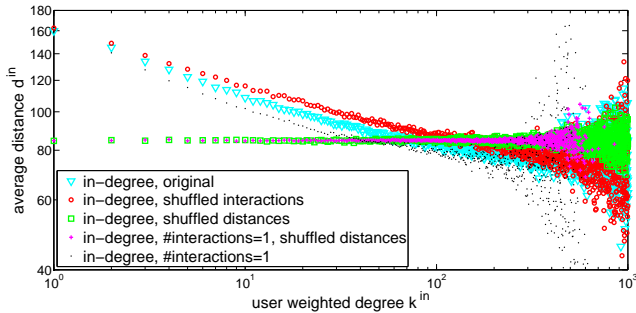
**Figure 5: Probability of interaction with a friend as a function of geographic distance for the weighted wall network and for the thresholded networks.**



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

Figure 5, the probability of interaction 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 remarkable stable, supporting the hypothesis that *while*



**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.

geographic distance heavily influences how users establish social connections, its effect on social interactions is only weak. In other words, once users choose their social connections, spatial factors are not important any more.

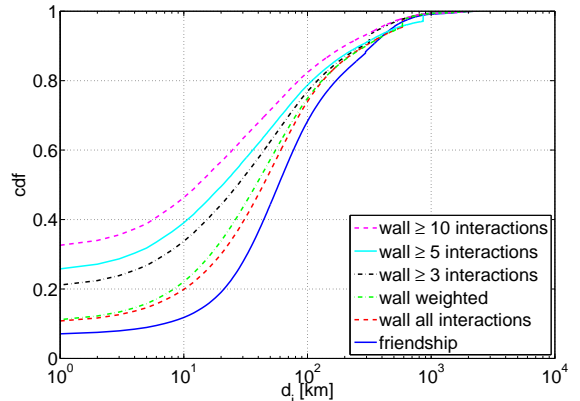
## 4.2 User properties

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**:

$$s_i^{in} = \sum_{j \rightarrow i} w_{j,i} d_{j,i} \quad s_i^{out} = \sum_{i \rightarrow j} w_{i,j} d_{i,j},$$

where as before  $w_{i,j}$  is the number of interactions from user  $i$  to user  $j$  and  $d_{i,j}$  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.  $k_i^{in} = \sum_j w_{i,j}$  and  $k_i^{out} = \sum_j w_{j,i}$ . 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 compare the original Tuenti data we introduce null models: we maintain the network structure as in the original Tuenti graph but either the interaction or the distance weights are shuffled, destroying any existing correlation. As baseline models we also consider models where all interaction weights are set to 1. In Figure 7 we plot the weighted average distance of incoming interaction 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



**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.

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. We see in Figure 8 that as we threshold the graph more and more, keeping only links with higher levels of interaction, the probability distribution shifts accordingly to lower values. 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 behavior 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. Such category of people could exhibit a noticeable propensity to interact more with other 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 analyzed 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.

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. Systems and architectures that rely on geographic locality of interest to serve online social networking traffic should take our results into account: while distant friends are rare, their social connections equally generate traffic load. Our findings are also likely to help other domains such as link prediction, tie strength modeling and user profiling: the observed spatial patterns can be also included in security mechanisms to detect malicious and spam accounts.

## 6. ACKNOWLEDGEMENTS

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*, pages 835–844, New York, NY, USA, 2007. ACM.
- [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*, pages 61–70, 2010.
- [3] A. Barrat, M. Barthélemy, 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. *The American Journal 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*, pages 369–382, New York, NY, USA, 2010. ACM.
- [9] R. Kumar, J. Novak, and A. Tomkins. Structure and Evolution of Online Social Networks. In *Proceedings of KDD '06*, pages 611–617, New York, NY, USA, 2006. ACM.
- [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*, pages 591–600, New York, NY, USA, 2010. ACM.
- [11] J. Leskovec and E. Horvitz. Planetary-scale views on a large instant-messaging network. In *Proceedings of WWW' 08*, 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, Cambridge, MA, 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*, pages 29–42, New York, NY, USA, 2007. ACM.
- [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] 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*, Hyderabad, India, 2011.
- [18] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo. Socio-Spatial Properties of Online Location-Based Social Networks. In *Proceedings of ICWSM'11*, 2011.
- [19] W. R. Tobler. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46:234–240, 1970.
- [20] 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*, pages 205–218, New York, NY, USA, 2009. ACM.
- [21] 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*, New York, NY, USA, 2010. ACM.