


The Impact of Geographic Distance on Online Social Interactions

David Laniado¹  · Yana Volkovich² · Salvatore Scellato³ · Cecilia Mascolo⁴ · Andreas Kaltenbrunner¹

© Springer Science+Business Media, LLC 2017

Abstract Online social networking services entice millions of users to spend hours every day interacting with each other. The focus of this work is to explain the effect that geographic distance has on online social interactions and, simultaneously, to understand the interplay between the social characteristics of friendship ties and their spatial properties. We analyze data from a large-scale online social network, Tuenti, with about 10 million active users: our sample includes user profiles, user home locations and online social interactions among Tuenti members. Our findings support the idea that spatial distance constraints whom users interact with, but not the intensity of their social interactions. Furthermore, friendship ties belonging to denser connected groups tend to arise at shorter spatial distances than social ties established between members belonging to different groups. Finally, we show that our findings mostly do not depend on the age of the users, although younger users seem to be slightly more constrained to shorter geographic distances. Augmenting social structure with geographic information adds a new dimension to social network analysis and a large number of theoretical investigations and practical applications can be pursued for

online social systems, with many promising outcomes. As the amount of available location-based data is increasing, our findings and results open the door to future possibilities: researchers would benefit from these insights when studying online social services, while developers should be aware of these additional possibilities when building systems and applications related to online social platforms.

Keywords Online social networks · Geographic properties · Online interactions · User behavior · Age factors

1 Introduction

Every day, more and more people spend an increasingly larger amount of time online, interacting on social networking platforms. Studying online social networks has allowed researchers to extend the scope of traditional social network analysis, scaling up to millions of individuals and billions of social links. The combination of large-scale data analysis and insights provided by sociological theories has resulted in a plethora of applications and systems that mine online social interactions to provide suggestions, offer recommendations and filter information. This has impacted the Web in an unprecedented way, as these features are profoundly different from the predominantly static content lacking any personalization, users were restricted to only two decades ago. In fact, insights into the properties of online social services can be exploited to design novel applications that, for example, provide recommendations about items (Golbeck 2008), answer Web search queries (Evans and Chi 2008; Horowitz and Kamvar 2010), and reduce spam (Garriß et al. 2006). Information related to online social ties can even be used to improve existing distributed systems and

✉ David Laniado
david.laniado@eurecat.org

¹ Eurecat - Technology Centre of Catalonia, Barcelona, Spain

² AppNexus, New York, USA

³ Google, London, UK

⁴ University of Cambridge, Cambridge, UK

applications: for instance, by taking into account how people use online social services to share and consume content items, it becomes possible to optimize delivery and storage of online content (Scellato et al. 2011a; Traverso et al. 2012).

In this work we aim to explore effects that geographic distance has on online social interaction and to comprehend the interplay between the social characteristics of online social ties and their spatial properties. Unveiling the role of the geographic factors in online social interactions is a first step towards a better design and understanding of many real-life applications and phenomena ranging from location-aware social dating platforms to online social movements (Conover et al. 2013; Bastos et al. 2014; Borge-Holthoefer et al. 2011). Given the increasing importance of online social networks as information pathways in our society, and ultimately as a new public sphere (Castells 2008), understanding the geographic properties of interactions on these services is essential to study information propagation dynamics, and to interpret phenomena such as rumour spreading or the echo chamber effect (Colleoni et al. 2014).

Geographic distance and social connections More recently, the widespread adoption of powerful mobile devices has led to a dramatic change in the way the Web is accessed. In particular, every day hundreds of millions of individuals use a smartphone to interact with their friends. An important related aspect is the increased access to social networking services through mobile devices: mobile users spend more time interacting with social applications than desktop users (Mascolo 2010). Simultaneously, mobile Web access has caused a substantial shift in the feasibility of pervasive and ubiquitous services. The deployment of location-based services has been made possible by the location-sensing capabilities of these devices; they are able to generate location-tagged information and enable users to share their physical whereabouts. As a result, online services are increasingly becoming *location-aware*.

Hence, the access to information about user location has paved the way for a new thread of research investigating the relationship between geographic distance and social ties (Liben-Nowell et al. 2005; Backstrom et al. 2010a; Scellato et al. 2010; Onnela et al. 2011; Spiro et al. 2016). Sociologists have widely discussed how individuals tend to minimize their social efforts by interacting more with close friends (Mok et al. 2009; Goldenberg and Levy 2009), suggesting that the principle of least effort stated by Zipf may make long-distance ties fade away (Zipf 1948). Nonetheless, as popularly suggested by Cairncross (2001), the cheap and fast communication tools provided by the Web might have instead completely removed the costs imposed by distance. However, several studies show how geographic distance still matters even in online social platforms: users

tend to connect preferentially to spatially close acquaintances rather than with individuals further away (Leskovec and Horvitz 2008; Backstrom et al. 2010a; Scellato et al. 2011b; Spiro et al. 2016). These results suggest that 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*” (Tobler 1970).

Interestingly, the exact relationship between geographic distance and social ties is still under debate: while the probability that two individuals are connected has always been found to decrease with their mutual geographic distance, different systems exhibit different properties. Social networks extracted from mobile phone interactions appear much more constrained by distance than mainstream online social services such as Facebook (Lambiotte et al. 2008; Backstrom et al. 2010a), denoting how online communication on the Web might be less affected by spatial proximity. Furthermore, the new generation of location-based social services exhibit properties that suggest how they are even *less* constrained by geographic distance, even though spatial proximity still plays an important role in shaping the social networks among users (Scellato et al. 2010; Scellato et al. 2011b).

The importance of online social interactions Overall, online social networks represent ideal systems to study social phenomena at a large scale: in fact, they offer fine-grained information about social behavior of large amounts of users. Hence, several studies have addressed the structural properties of the social networks arising among users of these services (Kumar et al. 2006; Ahn et al. 2007). In addition, these properties 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 (Mislove et al. 2007; Kaltenbrunner et al. 2011; Kwak et al. 2010b).

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 (Miritello et al. 2011), communication tends to be biased towards those relationships that are deemed more important (Dindia and Canary 1993). 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 (Granovetter 1973; Milgram 1977), *online friendship connections exhibit heterogeneous intensity*, with a large fraction of users interacting mainly with a small subset of acquaintances (Wilson et al. 2009; Jiang et al. 2010). In addition, online social connections are usually carefully chosen and displayed by users to represent their status and identity, supporting the hypothesis that such links often fail to signal social proximity,

mutual trust or even shared interest (Donath and Boyd 2004). Furthermore, reciprocity and transitivity of interaction have been shown to be important factors for the perceived importance of a social tie (Friedkin 1980; Krackhardt and Kilduff 1999). Failing to take these factors into account when studying the development of online social links is unlikely to uncover any true social property of these platforms.

Social properties, tie strength and spatial distance The importance of tie strength is directly connected to the hypothesis that weak ties, which are more likely to connect together otherwise separated portions of a network, play a meaningful role in information diffusion, as confirmed by a study on the Facebook social network (Bakshy et al. 2012). In turn, space is constraining network structure as well: in fact, social communities tend to be limited in their geographic span (Onnela et al. 2011; Expert et al. 2011).

The three properties – tie strength, social overlap and spatial distance – represent three different facets of a single system which combines spatial and social factors and binds together individuals, affecting complex processes such as the spreading of information over social links (Rogers 1995; Newman et al. 2006) or the ability to navigate the social networks to route a message to a particular individual (Kleinberg 2000). Nonetheless, the research community still lacks a broad understanding of the interplay between the structure of a social network, the strength of its ties and the space that embeds it.

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 questions we address in this work concern *the effect that geographic distance has on online social interactions* and, simultaneously, *the interplay between the social features of online friendship ties and their spatial properties*. We aim to address these questions through a detailed study of a large-scale social network service, Tuenti, at the time of data collection¹ the most popular online social network in Spain with almost 10 million active users. We have access to a complete snapshot of the social network including friendship ties among Tuenti members, the number of their online interactions with each other and their home locations, discretized across more than 7 000 Spanish cities. This allows us to study the social graph focusing on the spatial distances between users and on the level of interaction between them. In this paper we deepen the analysis of the interplay between geographic distance and social interactions started in our

previous work Kaltenbrunner et al. (2012) and Volkovich et al. (2012), and extend it to account for age factors and technology adoption patterns.

Our results support the idea that geographic distance strongly affects technology adoption patterns and 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 spatially 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 significant role in affecting interaction strength. This finding supports the idea that geography affects *whom* we interact with, but it does not influence *how much* we interact.

In turn we show that *how much we interact* depends strongly on social overlap, which counts the number of friends two connected users have in common. This social feature reflects how likely these users form part of a particularly well connected community. Furthermore, we discuss how age of a user might be influencing the spatial and social properties of online social interactions. We find that our results are mostly independent of this feature and can be observed across all age groups. We find however that *younger users tend to be both the most active and the most spatially constrained* while older users tend to have more friendship ties over long-range distance and exhibit lower levels of online interactions. Finally, we notice that the popularity of the social networking service goes hand in hand with its ability to penetrate significantly into layers of older users.

The rest of this article is organized as follows: in Section 2 we provide a review of results related to our work, in Section 3 we describe the online social service we study and introduce basic notations. In Section 4 we study the impact of geographic distance on technology adoption patterns, online social ties and levels of interaction. Further, in Section 5 we analyze the effect of user age on the socio-spatial properties and in Section 6 the interplay between interaction strength, structural and spatial properties of a social tie. Finally, we draw conclusions and outlook in Section 7.

2 Related Work

There are two main areas of research that are directly related to our work: studies of user interactions on online social services and works on the geographic and spatial properties of online social networks.

Online User Interactions Online social services initially sparked many works on understanding the network structure

¹Data collected by November, 2010.

among users. One of the initial studies focused on social connections on the Web between Stanford students, finding high levels of clustering and evidence of small-world behavior in the result network (Adamic et al. 2003). With the rising popularity of massive online services it became possible to study social interactions at larger and larger scale. Studies have found that these systems often show power-law distribution in the number of links and strongly connected cores of high-degree users (Kumar et al. 2006; Ahn et al. 2007; Mislove et al. 2007). Even if in some online social services, such as Twitter, user behavior may appear more interest-driven than purely social (Kwak et al. 2010b; Cha et al. 2010), connections between online users represent rich source of information to understand these systems.

Nonetheless, as users spend more time and accumulate more links in these services, establishing friendship connections may be largely different than actually interacting and exchanging information over these connections. Hence, researchers have studied not only how social links are created but also how they are used: in fact, online interactions appear affected by external factors such as age, gender, geographic distance, thus social links exhibit heterogeneous tie strength (Leskovec and Horvitz 2008). In particular, users seem to allocate in a skewed way their attention, devoting a large fraction of their online interactions with few key friends: as a result, it has been proposed that *interaction graphs* should be extracted from the underlying social network, considering only connections that present a certain given amount of explicit, or implicit, communication events (Wilson et al. 2009; Jiang et al. 2010). Our work further extends our knowledge on online social interaction by addressing the influence of geographic factors on users and on relationships.

With respect to this body of works, our paper brings a new factor into account in the analysis of user interactions: geographic distance. Thanks to our dataset we are able to discuss how, despite strongly affecting social links, spatial proximity has a negligible influence on online user interactions, while other non-geographic factors like age have a stronger influence on this variable.

Spatial Properties of Online Social Networks Even though the spatial properties of several technological networks such as Internet router connections or electric power grids have been widely studied (Barthélemy 2011), social analysis has often neglected the spatial dimension: very few works have tried to correlate social interaction and geographic characteristics, mainly because of the lack of data containing these features or because self-reported user locations are often misleading (Hecht et al. 2011). A seminal work on the effect of geographic distance on online social links investigated the probability of connection as a function of

distance between bloggers in LiveJournal, exploring the navigability of the resulting spatial social network and finding that online social ties seem affected both by spatial proximity and by other non-geographic factors (Liben-Nowell et al. 2005). Online communication through instant messaging also appears more biased towards users living in spatial proximity: however, even though there are more conversations between geographically close users, the amount of interactions over a social link increases with distance (Leskovec and Horvitz 2008).

The only study of the Facebook social network based on the whole graph (Ugander et al. 2011) considers geography at the granularity of countries, and reports that about 84% of the friendship ties fall within countries, with a high value of modularity (~ 0.75) for the partitioning of the users into countries. Other studies based on samples of data from the Facebook social graph confirmed the importance of distance at lower levels of granularity: the limiting effect of geographic distance on social links, with the probability of connection increasing with spatial proximity, seems so strong that it is possible to predict a user's location given only the friends' position. Likewise, in Twitter reciprocity and interactions in the form mentions have been used as predictors of spatially closer connections (McGee et al. 2013; Rout et al. 2013; Sadilek et al. 2012; Jurgens et al. 2015). However, distance affects users in different ways: there is a wide heterogeneity in the socio-spatial properties of individual users, with important correlations between the number of social connections and their geographic length of suggesting how social and spatial factors are jointly influencing social network structure (Scellato et al. 2011b). Beyond purely dyadic relationship, the effect of spatial constraints on more complex network structures such as triangles and communities appears less trivial: social triangles are not strongly constrained by spatial proximity (Lambiotte et al. 2008), while connected groups of individuals tend to be geographically tight when they have few members and more spatially diverse once their number of members grows (Onnela et al. 2011).

While building up on several of these results, this paper continues the investigation, started in our previous works (Kaltenbrunner et al. 2012; Volkovich et al. 2012), on how space and distance constrain user interaction and social tie strength on online social services. While most previous work focuses on services like Twitter, where social interactions may typically just reflect a common interest on specific topics (Kwak et al. 2010b), here we delve into an online social network where relationships are likely to reflect pre-existing offline ties, and in this context we address not only the effect of geographic distance on online social interactions but also analyze the interplay between these characteristics and the social and age properties of the users.

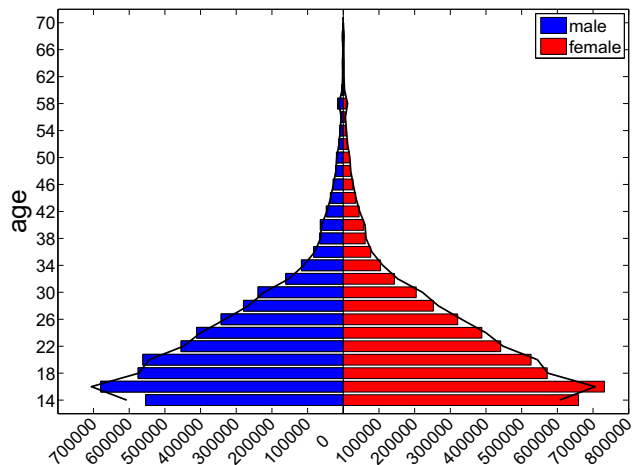


Fig. 1 Tuenti's population pyramid: the continuous line shows the average between male and female users across the different age groups

3 Dataset

Here we present the dataset under analysis and introduce notations we will use throughout our work. We also discuss basic demographics and social properties of the online service we study, Tuenti.

3.1 Tuenti

We use a large sample of Spanish, invitation-only social networking service, Tuenti² for our analysis. Founded in 2006, thanks to its widespread popularity in the country, Tuenti was commonly referred to as the “Spanish Facebook”. Since 2012 it has been rebranded as virtual mobile operator. Tuenti provided many features common to other popular social networking platforms: it allowed users to set up a profile, connect with friends, share web links and media items and write on each other's walls. In particular, the terms of agreement specified by Tuenti did not allow kids younger than 14 to join the service and oblige users to specify a place of residence located in Spain³.

Our dataset is based on a full anonymized snapshot of Tuenti friendship connections as of November 2010. It includes about 9.8 million registered users, more than 580 million bidirectional friendship links, and 500 million directed interactions (an interaction is an exchange of a wall message) during a 3 months period and the user's self-reported city of residence (selected from a predefined list)⁴.

In Fig. 1 we present the population pyramid for Tuenti. We observe a small difference between numbers of male

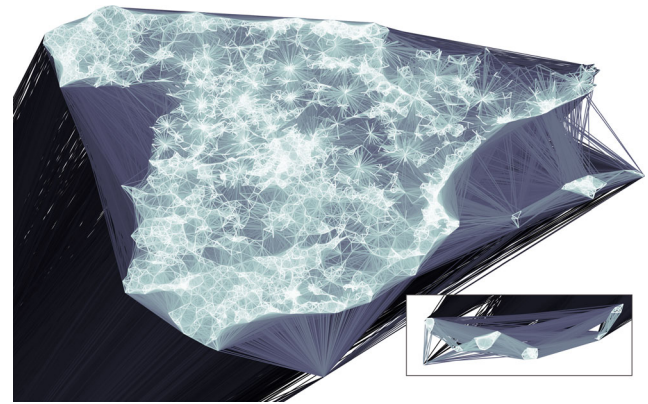


Fig. 2 Map of connections between Tuenti users: the Canary Islands archipelago is visible in the inset on the right. The shorter the geographic distance the brighter the connection

(4 899 659) and female (4 784 975) participants. Tuenti is particularly popular among teenagers: there are 4 376 472 (45% of the total) participants aged 14 to 20 with a Tuenti profile. A comparison with the official census data coming from the Spanish National Statistical Institute shows that, surprisingly, there are only 3 232 012 officially registered teenagers in Spain. Possible explanations for this discrepancy might be: (a) users that have several profiles, (b) users that live outside Spain and misreport their real location or (c) users that lie about their age. The latter is the most likely since the social networking service under study did not allow people under 14 to register, so we assume that many users reported falsely their age to be able to open an account. Indeed, according to a report (Livingstone et al. 2013) around 30% of Spanish kids aged 9–12 have a Tuenti account, i.e. lie about their age when using social media services. We therefore advise the reader that, in what follows, the results about usage patterns of the youngest user-groups might likely contain data of even younger users.

The location where users live is an important discriminator when a search returns a list of users with the same name. Since the top ten surnames cover about 20% of the population⁵, in general, Tuenti users have an incentive to specify the exact city where they live to be discoverable by the people they know.

Tuenti originated in Madrid, and further gained popularity in Seville, Valencia, Malaga and Gran Canaria, progressively gaining traction in Spain: in fact, the service has become pervasively used in many cities. In the map in Fig. 2 we plot a line between the geographic positions of two cities if the total number of friendship connections between these two cities exceeds a threshold of 50. The color of the lines reflects the distance: dark blue for the longest distance friendship links and dark red for the shortest. We

²www.tuenti.com

³This requirement was later changed, but after our dataset was collected.

⁴In other words all users are assigned to a city (represented by a point)

⁵en.wikipedia.org/wiki/Spanish_naming_customs

clearly observe that the existence of many short distance links allows to deduce the geographic structure of Spain.

3.2 Basic Notations

In this work we explore how users' geographic locations correspond to their online social interactions. These interactions may be either explicitly declared connections such as friendship links in a social network or implicit ones retrieved from interactions, i.e. from wall comments. We note that Tuenti only allowed users that are friends to exchange wall messages: thus, we model the social network among Tuenti users as an undirected graph $G = (V, E)$, where nodes are users and edges are friendship connections on Tuenti. We refer to this graph as the undirected **friendship network**. In Fig. 3 we plot the distribution (pdf) of the number of friends in the friendship network. We see a peak at 1 000 friends, that was the friendship limit defined by Tuenti. Nonetheless, Tuenti opened this limit occasionally for users with special merits (e.g. celebrities), which explains why some users managed to overcome the limit.

In line with Gilbert and Karahalios (2009), we take the frequency of interaction as a proxy for interaction strength. Every edge $e_{ij} \in E$ between users i and j is assigned with corresponding **interaction weights** w_{ij} and w_{ji} , equal to the number of messages user i posted on the wall of user j and vice versa. Two users may be connected to each other but never exchange a message, hence $w_{ij} \geq 0$. In general $w_{ij} \neq w_{ji}$, hence we introduce an associated undirected **balanced interaction weight**:

$$\bar{w}_{ij} = \min(w_{ij}, w_{ji}) + (1 - \delta_{w_{ij}, w_{ji}})/2,$$

where $\delta_{x,y}$ is Kronecker delta (returns 1 if arguments are equal and 0, otherwise). The reason behind this definition is that, following Friedkin (1980) as well as Krackhardt

and Kilduff (1999), we are most interested in measuring bidirectional interactions (since non-reciprocated communication may indicate spam or anecdotal interactions) but we also want to take into account (to a much lesser extent) these non-reciprocated interactions. Therefore, for the non-reciprocated interactions we only add 1/2 no matter the difference in the numbers of messages exchanged. The balanced interaction weight assigned to each social link captures how likely a social tie is to be used to spread information.

Leveraging the edge balanced weights \bar{w}_{ij} we define graph G_m in such way that there is a link between user i and j only if $\bar{w}_{ij} \geq m$. We call G_1 as **wall network**, as it connects only users who have exchanged wall posts in both directions.

In Table 1 we report the main properties for the friendship and wall networks.

In our previous work (Volkovich et al. 2012) we have shown that additional friendship links are unlikely to significantly increase the levels of interaction. In particular, we found that the absolute number of active connections does not exceed 150 users. This value is in good agreement with Dunbar's number (Dunbar 1998), which is an alleged theoretical cognitive limit to the number of people with whom one can maintain stable social relationships.

4 Geographic Properties

The goal of this section is first to describe the geographic features of online (Tuenti) and offline (Spanish) population. Second, we discuss how distance between users affects their making friends behavior. Finally, we study the interplay between friendship strength and spatial properties of social ties.

4.1 Geographic Footprint

To provide a quick overview of the Tuenti population placed on the real geographic map we draw a histogram for the top 50 cities by number of users together with their population in Fig. 4. We observe that these top cities are, in general, the most populated ones. We look at the fraction of the number of registered users and find that for many of these top cities the fraction of involved users exceeds half of the population. In particular we note a very high proportion of registered users in medium-size cities with large university campus, such as Ciudad Real or Granada: this is probably due to the high number of young people who move in and actually live and study in these cities, although their official residence is still in their original hometown. For other cities, e.g. Barcelona and Palma, the percentage of Tuenti population is very small. This might be caused by the fact

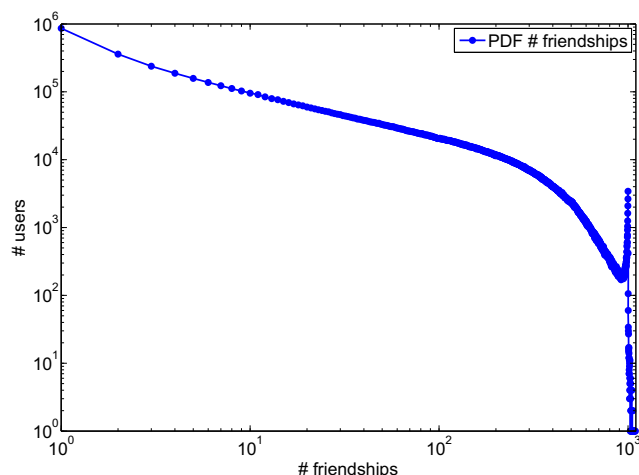


Fig. 3 Probability distribution function for the number of friends in Tuenti

Table 1 Properties of the wall and friendship 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 \mathcal{D}_{eff} , maximal distance Δ_{max}

Network	N	K	size GC	$\langle k \rangle$	$\langle C \rangle$	\mathcal{D}_{eff}	Δ_{max}	$\langle l \rangle$	$\langle D \rangle$	$\langle d \rangle$
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

that Tuenti only later extended its language support to the other Spanish co-official languages, in particular to Catalan, an important co-official language in these cities. Barcelona also has a large international community which might have brought in Facebook earlier.

Apart from language issues a possible explanation for a low proportion of users in a city can also be the fact that many people move away for study or work and get in contact with other social networking services. For example in Zaragoza a significant amount of people migrate to Barcelona for better job opportunities. It is likely that they signed up with other social networking services more popular in Barcelona and help to spread and increase the usage of these services (and decrease the usage of Tuenti) in their age-group back in their original hometown.

To investigate this idea and better understand the popularity of the service in different age groups, we compare the average age for people registered in the top cities on Tuenti with the population fraction. For these 50 cities we observe large correlation (0.64, p -value $< 10^{-6}$), which can be also seen on top plot in Fig. 4. Therefore, for cities with low average user ages we find, in general, that a smaller fraction of the population was using Tuenti.

To summarize it seems that *the popularity of the social networking service depends on whether it is able to penetrate significantly into layers of older users*. This statement brings us to an open question about the relation between popularity of the social network platform and the number of teens present in it. In Section 5 we further study age factors and their connection with the geographic properties of online social interactions.

4.2 Distance between Friends

For every edge $e_{ij} \in E$, $G = (V, E)$, we define d_{ij} as the geographic great-circle **distance**⁶ between the cities of residence of user i and user j . We set $d_{ij} = 0$ if users report the same city of residence. In other words we approximate

⁶The shortest distance between two points on the surface of a sphere measured along the surface of the sphere: $d_{ij} = r \arccos(\sin \phi_i \sin \phi_j + \cos(\lambda_i - \lambda_j) \cos \phi_i \cos \phi_j)$, where ϕ_i , λ_i and ϕ_j , λ_j are geographic latitude and longitude cities of user i and j . We use the mean Earth radius $r \approx 6378$ km.

between two nodes in the network, average path-length between nodes $\langle l \rangle$, average geographic distance between nodes $\langle D \rangle$ [km], average link length $\langle d \rangle$ [km]

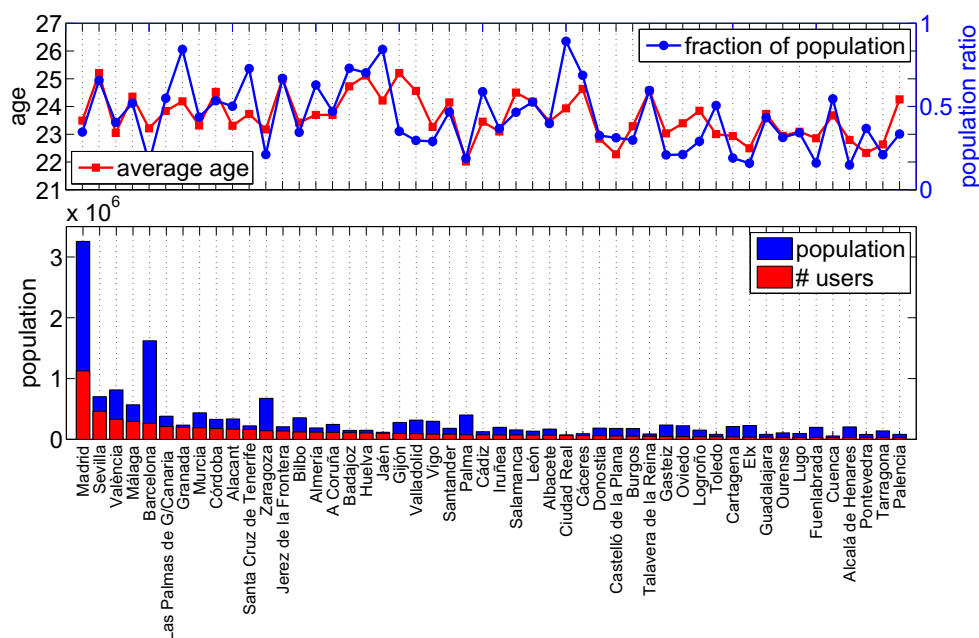
spatial closeness between users by considering “as the crow flies” distance. Note that distance between towns does not always reflect perfectly actual distances in km between persons. In particular, when distance is zero the users might live in the same city but a few kilometers away from each other.

As an alternative solution to absolute geographic distance, to take into account differences in population density (especially marked in the case of Spain), we experimented the ranked distance introduced by Liben-Nowell et al. (2005). With such definition, the distance between a user and a given point in space is not defined in number of kilometers, but in terms of the number of people who live closer to that point than the given user. While the model is useful to account for the different impact of distance in differently populated areas, its application raises several issues. As the granularity of our data corresponds to the municipality, for cities like Madrid or Barcelona with population in the order of millions, this model entails an extremely high number of tied values, and a user living in the surroundings of a big city may result to be very far from friends who actually live just across the municipality border. Due to the possible biases and issues introduced by this model, we preferred to rely on the absolute geographic distance, which has also the advantage of representing a straightforward and immediately understandable measure.

As found in other online social networks (Backstrom et al. 2010a; Scellato et al. 2011b), Tuenti users tend to preferentially connect to closer users. In fact, as depicted in Fig. 5, the distribution of geographic distance between connected users show much lower values than for random pairs of users. There are about 60% of social links between users at a distance of 10 km or less (blue line), while only 10% of all distances between users are below 100 km (black line). If we restrict this analysis to the wall network G_1 we see a similar trend as in the friendship network, though with slightly shorter distances.

We now focus on the distribution of geographic distance between friends. Different behavioral patterns associated with large, intermediate or short distance relationships are revealed. Figure 6 depicts the distribution of the geographic distance for all pairs of friends (black stars) and for all the friendship ties where at least one of the two users is living in one of the 4 largest cities in Spain: Madrid, Barcelona,

Fig. 4 Top cities by number of users (red bars) together with their populations (blue bars) and the fractions of the population present on Tuenti (blue line, top plot). Average age for Tuenti users registered in these cities (red line, top plot)



Valencia, and Seville. Since Madrid is, roughly, in the center of Spain one might expect that it is on average “closer” than other cities. We fit these distributions partially to power-law distributions and find low exponents (α_1) for short distance friendship links (distances less than 300 km). The exponents in this region oscillate between 1.05, if we only consider friendship links with one end in Barcelona, and 1.45 for friendships which involve people from Seville. The lower this exponent is the larger is the proportion of longer distance friendships. The distribution of all the distances has an exponent of 1.41 in this distance region. For friendships over distances between 300 and 1 000 km a much steeper slope is observed. The exponent (α_2) then oscillates

between 2.44 for the global distance distribution and 4.56 for Barcelona.

We see a clear difference in the proportion of short and large distance friendships between the main cities in Spain. The fraction of large distance friendships is much larger in Barcelona than in Madrid (and the remaining cities shown in Fig. 6). Only 30% of all friendships are within a radius of 300 km for users from Barcelona, while this fraction is 76% for users from Madrid and even 87% for Seville. The latter percentage also coincides with the overall fraction of friendships in Spain which can be found within a 300 km radius. If we just take into account the closer neighborhood, we find that only 17% of all friendships are within a radius of 10 km for Barcelonians, while in the overall case this number is elevated to about 56%. In particular, we find about 47% of all friendships within this distance for Madrid.

In Fig. 6a we observe a jump around 500 km in the distribution of Barcelona indicating connections with users from Madrid, a much smaller jump around 300 km for Valencia indicating friendships both from Barcelona and Madrid, and an even smaller jump for Seville again reflecting connections to Madrid. We observe a much smoother distribution for Madrid implying for example that the friendship links between Madrid and Barcelona occupy a smaller share of the total connections of Madrid than they do for Barcelona.

In general, we can conclude that *the Tuenti friendship network has a clear predominance for short distance friendships*, even if a different pattern arises in Barcelona. It seems that at least in Barcelona Tuenti was mainly used to maintain links with distant friends but not as a medium to interact with the local circle of friends. Given the low penetration of the social network in Catalonia, and the attractiveness

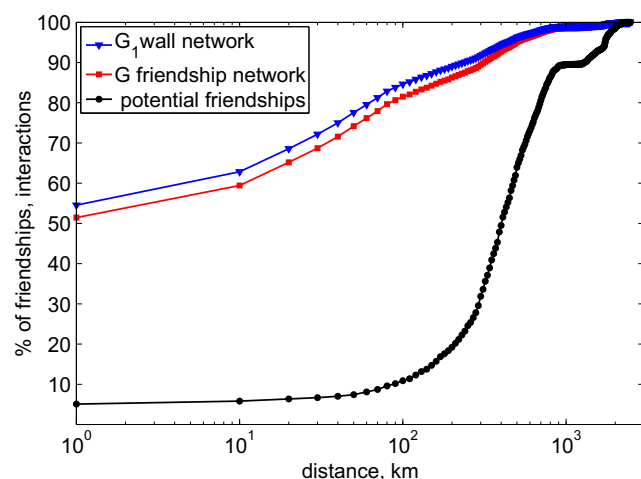


Fig. 5 Cumulative distribution function (cdf) of geographic distance of social links (red squares), mutual interaction links (blue triangles) and all pairs of users (black circles)

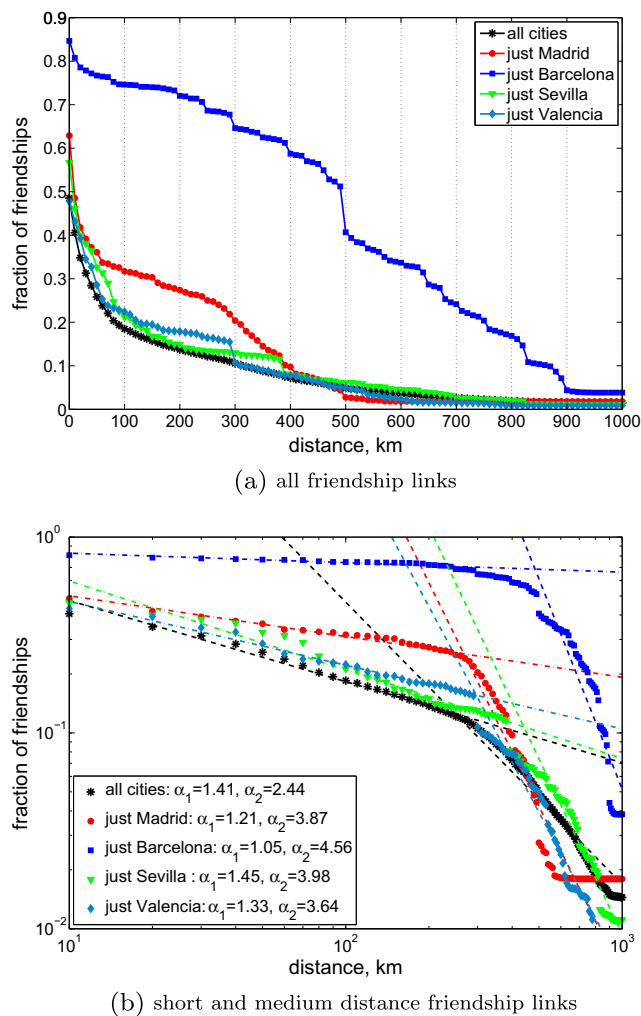


Fig. 6 Distribution (ccdf) of friendship distances and break-down for friendship ties originating from four Spanish cities (10 km binning)

of Barcelona in the internal migration flows for study and work, we conjecture that a remarkable part of Tuenti users located in Barcelona may be people from other parts of Spain, who live in Barcelona and use Tuenti to keep in contact with friends in their home region, where the service is widespread. Given the low penetration of the social network in Catalonia, and the attractiveness of Barcelona in the internal migration flows for study and work, we conjecture that a remarkable part of Tuenti users located in Barcelona may be people from other parts of Spain, who live in Barcelona and use Tuenti to keep in contact with friends in their home region, where the service is widespread.

4.3 Distance and Technology Adoption

We now take a look at how distance affects technology adoption patterns. Previous work showed that gender plays an important role at the time of joining a social network service, with 72% of female users joining Tuenti on invitation

of other women (Laniado et al. 2016). Here we inspect how joining the social network service is affected by distance.

Tuenti was an invitation only social network service at the time of data collection, therefore we assume that the first friend of each user is the friend who successfully invited her/him to join the platform. In Fig. 7 we show the cumulative distribution of distance for such connections representing successful invitations to join the service (blue line with triangles). The figure also shows for comparison all social links (red line with squares, corresponding to the red line in Fig. 5). On the one hand, we observe that short distances are more frequent in invitations with respect to the overall friendship graph, which indicates that distance is generally more relevant at the moment of joining the service through the invitation of the first friend than at the moment of adding other friends. On the other hand, after 50 km there is an inversion in the trend, indicating that longer distances are also more frequent among accepted invitations than among the overall friendships, which may point out the behavior of users who join the service to connect with friends who live in another part of the country.

4.4 The Effect of Distance

To assess the constraining effect of geographic distance on social ties we now compute the probability that 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 fact, in Fig. 8 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

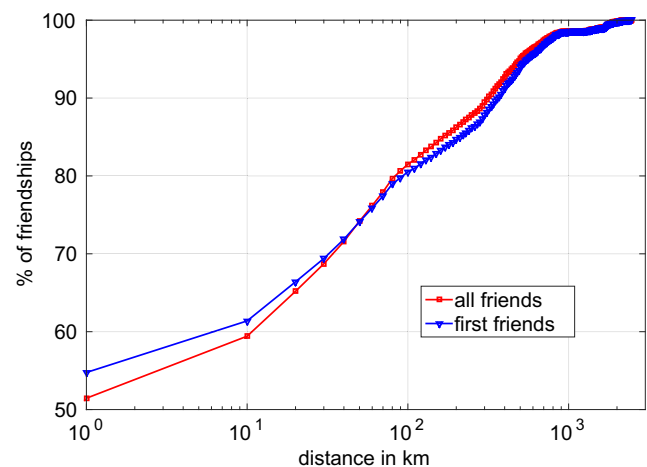
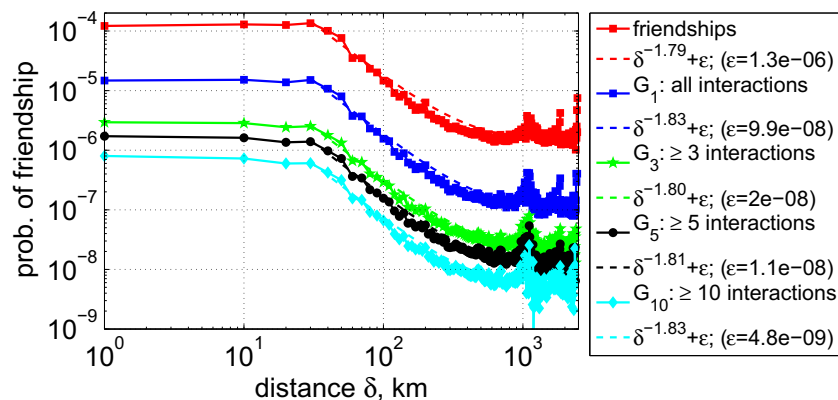


Fig. 7 Cumulative distribution function (cdf) of geographic distance of accepted invitations to join the social network service (blue triangles), compared to all social links (red squares)

Fig. 8 Probability of friendship and of wall interaction between two users as a function of their geographic distance. G_x indicates a minimum of x balanced interactions between a pair of users



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 (Liben-Nowell et al. 2005; Backstrom et al. 2010a).

In Fig. 8 we observe similar shapes for different interaction levels. In order to further understand the relationship between social interactions and spatial distance we compute a different property: the probability that a message is exchanged over an existing friendship link as a function of geographic distance. More formally we calculate $|E_{G_k(d)}|/|E_{G(d)}|$, where $|E_{G(d)}|$ and $|E_{G_k(d)}|$ are the cardinalities of the set of all edges in G which have distance d and the set of all edges in G_k which have distance d . 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 Fig. 9, the probability of interaction ranges between 0.35

and 0.15 even when geographic distances increases from 0 to 1000 km for $k = 1$. If we consider larger thresholds 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*.

5 Age Factors

As we have already discussed in Section 4.1, user age may have an impact on the geographic properties of online social networking. Recall that we study here a social networking platform which is very popular among teenagers but has also a significant number of users in older population layers. To complete our analysis we now investigate whether the conclusions drawn in the previous section hold when the age of the users is considered. As discussed in Section 3, we cannot consider self-reported user age to be reliable in all cases; in particular, we recall that users younger than 14 are likely to report a slightly higher age in order to be able to register to the service, so among the youngest users in our data there may be also users whose real age is below 14. We believe that, despite this caveat, the analysis proposed in this section can be helpful to assess the impact of distance on users of different age classes.

We first explore how the probability of friendship varies for individuals of different age groups. In Fig. 10 we compute the probability that two users are connected as a function of their spatial distance under assumption that at least one of these users is within the selected age group. We can observe that the probability of connection $P(d)$ as a function of geographic distance d exhibits a decay $d^{-\alpha} + \epsilon$ with very similar values of α for the different age groups. This behavior is also observed for the data of the entire social network (black curve in Fig. 10, which is the same as the red curve in Fig. 8). The main effect of age seems

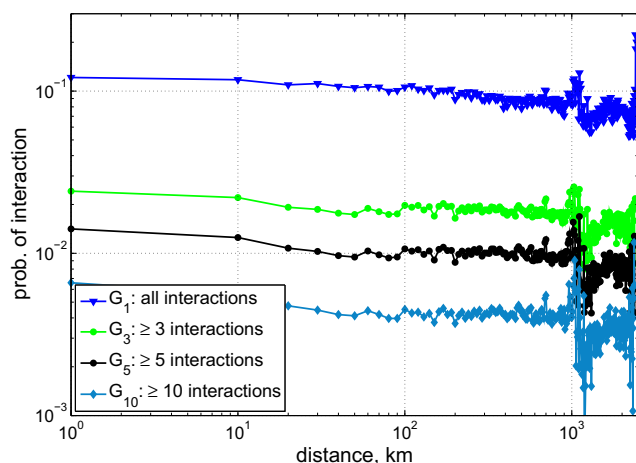
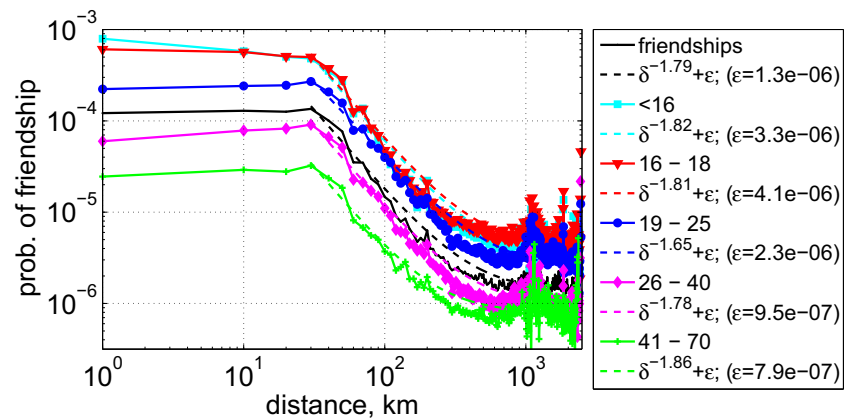


Fig. 9 Probability of interaction with a friend as a function of geographic distance for the weighted wall network and for the thresholded networks

Fig. 10 Probability of friendship in different age groups as a function of geographic distance



to be a parallel vertical shift of the curve of the friendship probability for the entire network reflected in different values for the constant background probability ϵ . Older users have the smallest and younger users the largest probabilities of friendship. Minor differences from this general tendency appear for the youngest users who present higher probability of connections at their immediate neighborhood (distance ≤ 1 km), while older age groups show slightly lower values in this range compared to a shift of the overall friendship probability (black curve).

In summary, *age introduces only a slight bias in the probability of friendship connection* and it becomes significant only at short distances. Thus, distances below 10km seem to constrain teens younger than 16, i.e. users with more limited mobility, while older users, probably due to commuting or migration processes caused by job or family bonds, are more likely to have connections at a distances of 10 to 40km than in their immediate geographic neighborhood.

Finally, we study how age and distance interplay with the interaction activity among friends. As found in other online social platforms (Leskovec and Horvitz 2008), younger Tuenti users tend to be more active than older ones. Figure 11 shows the probability that at least one message is exchanged in both directions over an existing friendship link as a function of geographic distance given that at least one of the friends is within the selected age group. We observe a very similar behavior as in Fig. 9. This becomes even more evident when comparing the data for the different age groups with the complete interaction probability in G_1 (black curve in Fig. 11, or blue in Fig. 9).

We find that teenagers are the most active group, thus, the probability to write a wall message for users younger than 18 ranges between 0.1 and 0.15 while for users older than 26 between 0.04 and 0.08. Again we observe that the youngest users are slightly constrained by distance: the probability of interaction decreases as distance increases. For older users we see, however, a small increase for distances close to 1 000 km. It is possible that these users are interested in reconnecting with old friends scattered all around Spain: for

instance, it was common practice in Spain to be assigned on army service (now abolished) far away from the birthplace, mixing up with people from different parts of the country.

To generalize we conclude that the findings presented in the previous section are even more robust for the users aged between 19 and 40. At the same time we suggest that the slight geographic constraints on online interactions reported in Section 4.4 are caused by the large presence of younger users. These constraints seems to disappear for users older than 19.

6 Distance and Social Interactions

In the previous section we found that the probability of interactions among friends does not seem to depend on geographic distance between them. Here we further explore the factors which influence the social interaction strength by analyzing the threefold relationship between spatial properties, social overlap, and balanced interaction weight of a social tie. Some of the results presented in this section have been reported in our previous work (Volkovich et al. 2012).

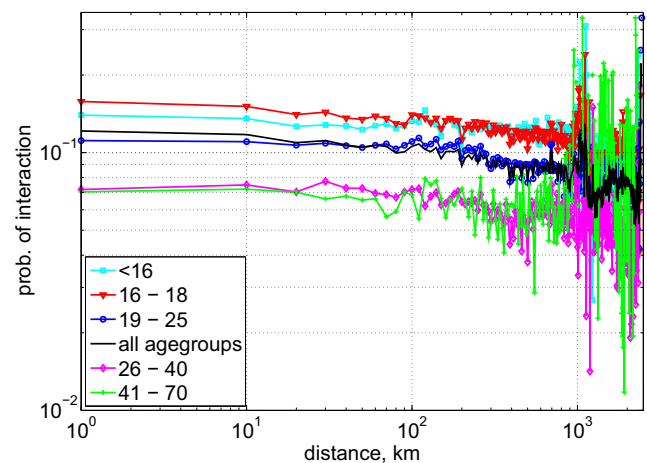
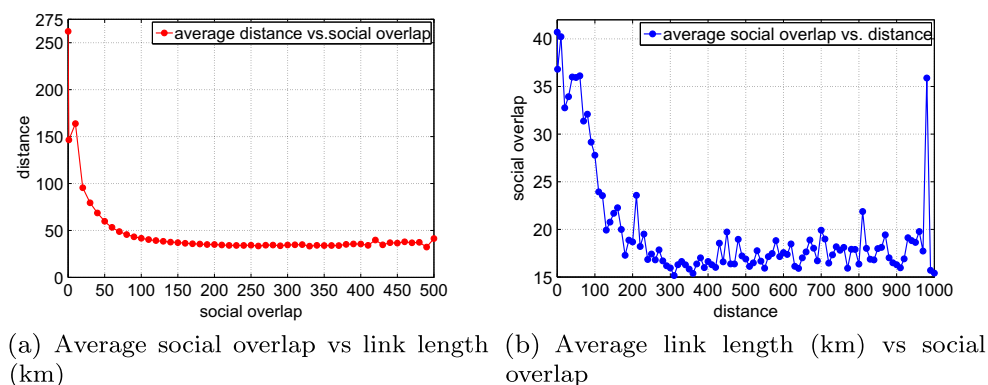


Fig. 11 Probability of interaction in different age groups as a function of geographic distance

Fig. 12 Relationship between link length and social overlap

6.1 Distance and Social Overlap

The structure of social networks tends to reflect that individuals usually belong to social groups: that is, friends usually know each other, creating clusters of individuals that are mutually connected to one another. In particular, when two connected users have several friends in common, it is an indication that their social link is situated inside a particularly well connected social community. We define the **social overlap** of an edge e_{ij} as $o_{ij} = |\Gamma_i \cap \Gamma_j|$, where Γ_i is the set of users connected to user i .

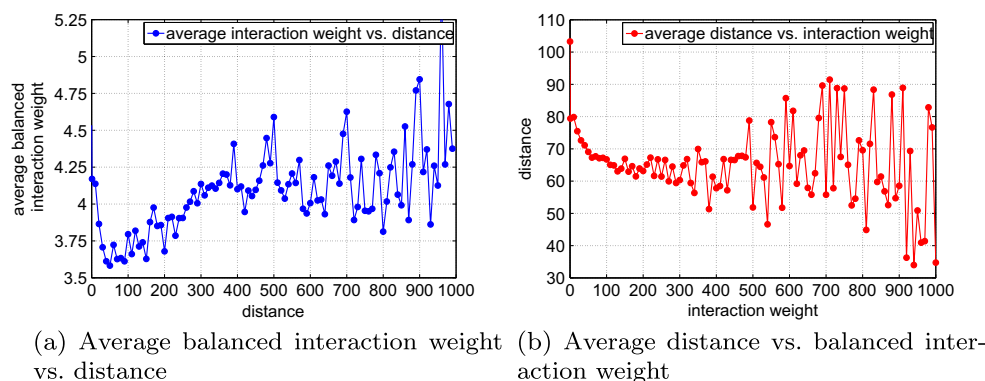
We analyze the relation between social overlap and distance of a social tie in Fig. 12. We see in Fig. 12a that the average geographic distance between two connected users decreases as they share more and more friends, i.e. they are probably inside a particularly well connected community. At the same time, in Fig. 12b, when looking at social links of increasing geographic length, we observe that social connections which span less than 60km exhibit higher average values of social overlap. The social overlap of longer links decays then and stabilizes at distances greater than 100km where it oscillates between 15 and 25 shared friends on average. This indicates how social links can be divided into *short-range* and *long-range* connections, with the separation distance being between 60 and 100km marking the sizes of the zones of attraction of major cities.

The presented analysis indicates how the division of the social network into communities does not take place independently of spatial distance. Instead, social ties belonging to denser connected groups tend to arise at shorter spatial distances than social ties established between members belonging to different groups. Thus, there are Tuenti users who seek connections at longer distances outside their local social circle: this would signal that either they have no potential connection available at short distance, as they may be located in a scarcely populated area, or they are more willing to connect to individuals far away. In general words, we see evidence of *a bridging behavior of spatially longer social links, connecting together diverse portions of the network, while shorter links are tightly integrated inside social groups*.

6.2 Social Interactions

The strength of the interaction between two members is an important facet of social relationships that we capture in our dataset since not all the social links are equally likely to be used for interactions. Here we examine how the social network carries and spreads information by studying whether the balanced interaction weight of a social tie depends on its social position and spatial length.

As previously discussed in Section 4.4 the probability of connection between individuals seems to be affected by

Fig. 13 Relationship between link length and interaction weight

geographic distance in the same way across different levels of user interaction. Furthermore, 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 Fig. 13a. The average balanced interaction weights are unrelated to users' geographic distances and, at the same time, the average distance between two individuals is only slightly related to the number of messages they exchange. In Fig. 13b we observe that there is a slight decay from an average distance of around 80km for a lower number of interactions to 60km if the users interact up to 100 times. Friends who never interacted or interacted just once but not mutually live on average further away, i.e. around 100 from each other.

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 their social connections, spatial factors are not important anymore. Thus, even though the likelihood that two individuals are connected is heavily dependent on distance, when considering how much friends interact geographic distance is not a limiting factor.

Finally, we look at the threefold relationship between interaction weight, structural properties and spatial distance. To this end, we plot the heat maps for the logarithm of balanced interaction weight as a function of distance and social overlap of a social tie. From Fig. 14 we conclude that interaction strength between online friends strongly depends on how many friends they have in common, i.e. if they are inside a particularly well connected community, but very little depends on how far from each other they are. We can therefore conclude that social overlap is a determinant factor for strong ties between two users and the effect of geographic distance is negligible in comparison to that.

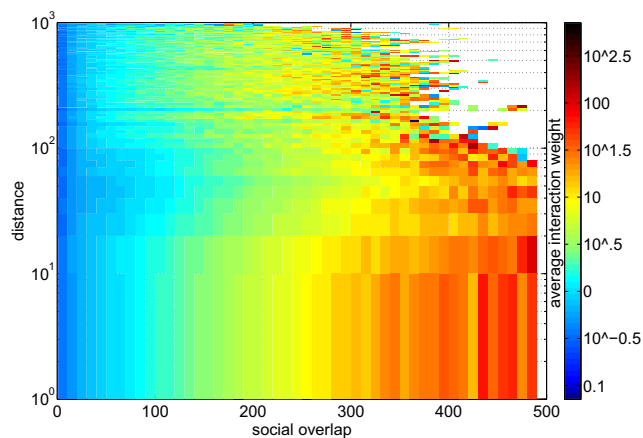


Fig. 14 Threefold relationship between social overlap, distance and balanced interaction weight

7 Conclusions and Outlook

Our work was primarily inspired by the idea that individuals are affected by spatial proximity and geographic factors in their online social interactions. Service providers gather data about where individuals are located and where they go, together with information about their social interactions. This exposes the spatial properties of social connections arising on the Web, that can be studied both to understand the dynamics proper of the online sphere, and to investigate the underlying social connections as they get reflected in online services. In this context, online social networks like Facebook or Tuenti, where so-called friendship connections mostly reflect offline ties, offer an exceptional opportunity for inspecting human social networks at a large scale, pushing forward sociological research in the line of Milgram's small world experiment (Travers and Milgram 1967).

While previous works on the geographic properties of interactions in online social networks are mostly limited to the US, and rely on surveys or samples of users (Backstrom et al. 2010a; Spiro et al. 2016), in this study we offer an unprecedented analysis of the entire social graph from a large online social network representing an important part of the population of a European country. We analyse the interactions between about 10 million Tuenti users; given the low average age of the users, results for older users may be less representative, while with respect to young people our dataset includes the majority of the Spanish population. Indeed, according to user surveys nearly 70% of Spanish people between 14 and 19 had a Tuenti account in 2011 (Telefónica 2012).

The lack of comparably rich datasets and studies from other countries (with the partial exception of the US) makes it difficult to assess to what extent our results might depend on the specificities of Spanish territory and society, such as the low internal mobility compared to other countries (Bell et al. 2015), and to which extent they could be generalized to other European or non-European countries. In absence of adequate terms of comparison, we believe this study offers a landmark for the understanding of the geographic properties of online social interactions, and may be of general interest well beyond the specific focus on the Spanish case.

We found evidence of a marked preference to connect to spatially closer users; also the number of shared friends is higher for shorter distance connections. Furthermore, we have observed that geographical distance has an impact on technology adoption patterns, with accepted invitations to join the service coming in most cases from friends at very short distances. However, once users have signed up to the service and chosen their friendship connections, our results show that spatial factors are not important anymore. This finding is robust when analyzing different age groups separately, and is even stronger when considering only users

aged over 19. Teenagers under 19, on the other hand, appear to be slightly constrained to shorter spatial distance in their interactions, and especially in their friendship connections. This can be explained by their generally lower geographic mobility.

Assuming that interaction strength captures how likely a social tie is to be used to spread information, we have further investigated communication patterns between Tuenti users. We have found that extremely high levels of interaction mainly take place between users with many shared friends. The network community structure and the number of shared friends, much more than geographic factors, appear to be good indicators for a strong friendship tie.

These findings have implications on the understanding of propagation phenomena in social networks over larger distances, such as news spreading, rumor spreading, or the emergence of nation-wide social movements (Conover et al. 2013; Bastos et al. 2014; Borge-Holthoefer et al. 2011). They are relevant also to the delicate debate on the *echo chamber* effect and the tendency to polarization in social media where, due to homophily, users tend to be mostly exposed to content spread by like-minded people (Colleoni et al. 2014).

In this sense, on one hand the positive dependency between spatial distance and social overlap in Tuenti confirms the prevalently local nature of cohesive communities. This was also observed in the communication dynamics of grassroot social movements in Spain on Twitter (Borge-Holthoefer et al. 2011) - a case in which one might expect geographic distance to be less relevant, given the country-wide nature of the movement, and the largely topic oriented nature of Twitter conversations (Kwak et al. 2010b).

On the other hand, our findings point out the importance of *bridge ties* that, despite spanning longer geographic distances, and exhibiting in general smaller social overlap, are as likely to convey information and communication as shorter ties falling into cohesive communities. This confirms the critical role of such social links in connecting geographically (and socially) distant communities, which can ultimately be determinant for enhancing dialogue and contributing to build a more cohesive society.

By increasing our ability to understand, model and predict the geographic properties of online social interactions, the results presented here may help to improve the design of real-life online applications such as recommender systems, news filtering or link prediction. Our study indicates many possibilities for future work, ranging from models that can reproduce the properties that we have observed in empirical data to new algorithms that can help to partition users better into communities with social and spatial focus. An inspection of the geographic composition of clusters resulting from community detection in the social network would help to shed further light on the local dimension of

cohesive communities in different areas, and to inspect the existence of composite communities that cross regional borders. A more detailed analysis of the users' ego-networks could furthermore help to study phenomena such as internal migration.

References

- Adamic, L.A., Buyukkokten, O., & Adar, E. (2003). A social network caught in the web. *First Monday*, 8(6).
- Ahn, Y.Y., Han, S., Kwak, H., Moon, S., & Jeong, H. (2007). Analysis of topological characteristics of huge online social networking services. In *Proceedings of WWW' 07* (pp. 835–844). New York: ACM.
- Backstrom, L., Sun, E., & Marlow, C. (2010a). Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of WWW '10* (pp. 61–70).
- Bakshy, E., Rosenn, I., Marlow, C.A., & Adamic, L.A. (2012). The role of social networks in information diffusion. In *Proceedings of the 21st world wide web conference (WWW 2012)* Lyon.
- Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 499, 1–101.
- Bastos, M.T., da Cunha Recuero, R., & da Silva Zago, G. (2014). Taking tweets to the streets: a spatial analysis of the vinegar protests in Brazil. *First Monday*, 19(3).
- Bell, M., Charles-Edwards, E., Ueffing, P., Stillwell, J., Kupiszewska, M., & Kupiszewska, D. (2015). Internal migration and development: comparing migration intensities around the world. *Population and Development Review*, 41(1), 33–58.
- Borge-Holthoefer, J., Rivero, A., García, I., Cauhé, E., Ferrer, A., Ferrer, D., Francos, D., Iniguez, D., Pérez, M.P., Ruiz, G., & et al (2011). Structural and dynamical patterns on online social networks: the Spanish May 15th movement as a case study. *PloS one*, 6(8), e23–883.
- Cairncross, F. (2001). *The death of distance: how the communications revolution is changing our lives*. Cambridge: Harvard Business School Press.
- Castells, M. (2008). The new public sphere: global civil society, communication networks, and global governance. *The ANNALS of the American Academy of Political and Social Science*, 616(1), 78–93.
- Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K.P. (2010). Measuring user influence in twitter: the million follower fallacy. In *Proceedings of ICWSM '10*. <http://aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1538/0>.
- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data. *Journal of Communication*, 64(2), 317–332.
- Conover, M.D., Davis, C., Ferrara, E., McKelvey, K., Menczer, F., & Flammini, A. (2013). The geospatial characteristics of a social movement communication network. *PloS one*, 8(3), e55–957.
- Dindia, K., & Canary, D.J. (1993). Definitions and theoretical perspectives on maintaining relationships. *Journal of Social and Personal Relationships*, 10(2), 163–173. <http://doi.org/10.1177/026540759301000201>. <http://spr.sagepub.com/content/10/2/163.short>, <http://spr.sagepub.com/content/10/2/163.full.pdf+html>.
- Donath, J., & Boyd, D. (2004). Public Displays of Connection. *BT Technology Journal*, 22(4), 71–82. <http://doi.org/10.1023/B:BTJT.0000047585.06264.cc>.
- Dunbar, R. (1998). *Grooming, Gossip, and the Evolution of Language*. Harvard University Press.
- Evans, B.M., & Chi, E.H. (2008). Towards a model of understanding social search. In *Proceedings of the 11th ACM conference on*

- computer supported cooperative work (CSCW 2008) (pp. 485–494). San Diego: ACM.
- Expert, P., Evans, T.S., Blondel, V.D., & Lambiotte, R. (2011). Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences*, 108(19), 7663–7668.
- Friedkin, N. (1980). A test of structural features of granovetter's strength of weak ties theory. *Social networks*, 2(4), 411–422.
- Garriss, S., Kaminsky, M., Freedman, M.J., Karp, B., Mazières, D., & Yu, H. (2006). RE: Reliable Email. In *Proceedings of the third Symposium on Networked Systems Design and Implementation (NSDI '06)* (pp. 297–310).
- Gilbert, E., & Karahalios, K. (2009). Predicting tie strength with social media. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 211–220). ACM.
- Golbeck, J. (2008). Weaving a web of trust. *Science*, 321(5896), 1640–1641.
- Goldenberg, J., & Levy, M. (2009). Distance Is Not Dead: Social Interaction and Geographical Distance in the Internet Era. arXiv:0906.3202.
- Granovetter, M.S. (1973). The strength of weak ties. *The American Journal of Sociology*, 78(6), 1360–1380. <https://doi.org/10.2307/2776392>.
- Hecht, B., Hong, L., Suh, B., & Chi, E.H. (2011). Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. In *Proceedings of CHI '11*.
- Horowitz, D., & Kamvar, S.D. (2010). The anatomy of a large-scale social search engine. In *Proceedings of the 19th world wide web conference (WWW 2010)* ACM, Raleigh North Carolina.
- Jiang, J., Wilson, C., Wang, X., Huang, P., Sha, W., Dai, Y., & Zhao, B.Y. (2010). Understanding latent interactions in online social networks. In *Proceedings of IMC '10* (pp. 369–382). New York: ACM. <https://doi.org/10.1145/1879141.1879190>.
- Jurgens, D., McCorriston, J., Xu, Y.T., & Ruths, D. (2015). Geolocation prediction in twitter using social networks: A critical analysis and review of current practice. In *Proceedings of the 9th international AAAI conference on weblogs and social media (ICWSM)*.
- Kaltenbrunner, A., Gonzalez, G., Ruiz De Querol, R., & Volkovich, Y. (2011). Comparative analysis of articulated and behavioural social networks in a social news sharing website. *New Review of Hypermedia and Multimedia*, 17(3), 243–266. <https://doi.org/10.1080/13614568.2011.598192>.
- Kaltenbrunner, A., Scellato, S., Volkovich, Y., Laniado, D., Currie, D., Jutemar, E.J., & Mascolo, C. (2012). Far from the eyes, close on the web: impact of geographic distance on online social interactions. In *Proceedings of the 2012 ACM workshop on online social networks* (pp. 19–24). ACM.
- Kleinberg, J.M. (2000). Navigation in a small world. *Nature*, 406(6798), 845. <https://doi.org/10.1038/35022643>.
- Krackhardt, D., & Kilduff, M. (1999). Whether close or far: Social distance effects on perceived balance in friendship networks. *Journal of personality and social psychology*, 76(5), 770.
- Kumar, R., Novak, J., & Tomkins, A. (2006). Structure and evolution of online social networks. In *Proceedings of KDD '06* (pp. 611–617). New York: ACM. <https://doi.org/10.1145/1150402.1150476>.
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010b). What is Twitter, a social network or a news media? In *Proceedings of the 19th international World Wide Web conference* (pp. 591–600). ACM.
- Lambiotte, R., Blondel, V., Dekerchove, C., Huens, E., Prieur, C., Smoreda, Z., & Vandooren, P. (2008). Geographical dispersal of mobile communication networks. *Physica A*, 387(21), 5317–5325. <https://doi.org/10.1016/j.physa.2008.05.014>.
- Laniado, D., Volkovich, Y., Kappler, K., & Kaltenbrunner, A. (2016). Gender homophily in online dyadic and triadic relationships. *EPJ Data Science*, 5(1), 1.
- Leskovec, J., & 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., & Tomkins, A. (2005). Geographic routing in social networks. *PNAS*, 102(33), 11,623–11,628. <https://doi.org/10.1073/pnas.0503018102>.
- Livingstone, S., Ólafsson, K., & Staksrud, E. (2013). Risky social networking practices among underage users: lessons for evidence-based policy. *Journal of Computer-Mediated Communication*, 18(3), 303–320. <https://doi.org/10.1111/jcc4.12012>.
- Mascolo, C. (2010). The power of mobile computing in a social era. *Internet Computing*, 14(6), 76–79.
- McGee, J., Caverlee, J., & Cheng, Z. (2013). Location prediction in social media based on tie strength. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management* (pp. 459–468). ACM.
- Milgram, S. (1977). *The familiar stranger: an aspect of urban anonymity*. Cambridge: Addison-Wesley.
- Miritello, G., Moro, E., & Lara, R. (2011). Dynamical strength of social ties in information spreading. *Physical Review E*, 83, 045–102. <https://doi.org/10.1103/PhysRevE.83.045102>.
- Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In *Proceedings of IMC '07* (pp. 29–42). New York: ACM. <https://doi.org/10.1145/1298306.1298311>. <http://portal.acm.org/citation.cfm?id=1298306.1298311>.
- Mok, D., Wellman, B., & Carrasco, J.A. (2009). Does distance still matter in the age of the Internet? *Urban Studies*, 46(13), 2747–2783.
- Newman, M., Barabasi, A.L., & Watts, D.J. (2006). *The Structure and Dynamics of Networks*, 1st edn. Princeton Studies in Complexity: Princeton University Press.
- Onnela, J.P., Arbesman, S., González, M.C., Barabási, A.L., & Christakis, N.A. (2011). Geographic constraints on social network groups. *PLoS ONE*, 6(4), e16,939. <https://doi.org/10.1371/journal.pone.0016939>.
- Rogers, E. (1995). *Diffusion of innovations*. New York: Free Press.
- Rout, D., Bontcheva, K., Preoiuc-Pietro, D., & Cohn, T. (2013). Where's@ wally?: a classification approach to geolocating users based on their social ties. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media* (pp. 11–20). ACM.
- Sadilek, A., Kautz, H., & Bigham, J.P. (2012). Finding your friends and following them to where you are. In *Proceedings of the fifth ACM international conference on Web search and data mining* (pp. 723–732). ACM.
- Scellato, S., Mascolo, C., Musolesi, M., & Latora, V. (2010). Distance matters: geo-social metrics for online social networks. In *Proceedings of WOSN'10*.
- Scellato, S., Mascolo, C., Musolesi, M., & Crowcroft, J. (2011a). Track globally, deliver locally: improving content delivery networks by tracking geographic social cascades. In *Proceedings of the 20th world wide web conference (WWW'11)* Hyderabad, India.
- Scellato, S., Noulas, A., Lambiotte, R., & Mascolo, C. (2011b). Socio-Spatial properties of online location-based social networks. In *Proceedings of ICWSM'11*.
- Spiro, E.S., Almquist, Z.W., & Butts, C.T. (2016). The persistence of division: geography, institutions, and online friendship ties. *Socius: Sociological Research for a Dynamic World*, 2, 2378023116634,340.
- Telefónica, F. (2012). La Sociedad de la Información en España 2011. Fundación Telefónica.
- Tobler, W.R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46, 234–240. <https://doi.org/10.2307/143141>.
- Travers, J., & Milgram, S. (1967). The small world problem. *Psychology Today*, 1, 61–67.

- Traverso, S., Huguenin, K., Erramilli, T.V.I., Laoutaris, N., & Pagiannaki, K. (2012). Tailgate: handling long-tail content with a little help from friends. In *Proceedings of the 21st world wide web conference (WWW 2012)*, Lyon, France.
- Ugander, J., Karrer, B., Backstrom, L., & Marlow, C. (2011). The anatomy of the facebook social graph. arXiv:1114503.
- Volkovich, Y., Scellato, S., Laniado, D., Mascolo, C., & Kaltenbrunner, A. (2012). The length of bridge ties: structural and geographic properties of online social interactions. In *International AAAI conference on weblogs and social media (ICWSM-12)*.
- Wilson, C., Boe, B., Sala, A., Puttaswamy, K.P.N., & Zhao, B.Y. (2009). User interactions in social networks and their implications. In *Proceedings of eurosys '09* (pp. 205–218). New York: ACM.
- Zipf, G.K. (1948). *Human behaviour and the principle of least effort*. Cambridge, MA: Addison-Wesley.

David Laniado is senior researcher in the Digital Humanities department at Eurecat - Technology Centre of Catalonia. He received from Politecnico di Milano his master degree in Computer Science in 2007 and his Ph.D in Information Engineering in 2012, with a thesis on “Social production of knowledge by online communities”. Since 2010 at Barcelona Media, and since 2015 at Eurecat, he works in the field of computational social science, developing quantitative approaches to analyse and characterize online communities. His research focuses mainly on the study of online interaction patterns, and in particular on online conversation and discussion, deliberation processes and collaboration dynamics. He has co-authored over forty academic publications in this area.

Yana Volkovich is a Senior Data Scientist in AppNexus, a technology company whose cloud-based software platform enables and optimizes programmatic online advertising. Before joining AppNexus Yana was a Senior Scientist in Eurecat, one of the largest Technology Centers in Spain. In 2014 she was visiting Cornell Tech NYC as a Marie Curie Tecniospring Research fellow. Her interests were to explore intersections between data-driven and network-based approaches grounded in social theory, urban research, and geo-social studies. In 2009 she received her PhD in Applied Probability from University of Twente (the Netherlands). Her thesis was entitled “Stochastic Analysis of Web Page Ranking” and focused on probabilistic analysis of power-law graphs.

Salvatore Scellato is a Staff Software Engineer at Google, where he leads mobile projects related to location inference and context-awareness, contributing to products with hundreds of millions of users. Salvatore completed his PhD in 2012 in the Network and Operating Systems group at the Computer Laboratory in the University of Cambridge. His main research interests include online social networks, location-based services, human mobility modelling and related software systems. During his PhD he has published in the fields of data mining, online social systems and user behaviour modelling. Salvatore was also a research intern at Google, and he joined the company soon after completing his PhD.

Cecilia Mascolo is Professor of Mobile Systems in the Computer Laboratory, University of Cambridge. Prior joining Cambridge in 2008, she has been a faculty member in the Department of Computer Science at University College London. She holds a PhD and an MSc degree from the University of Bologna. She specializes on mobile systems and mobile data analytics. Her wider research interests are in human mobility modelling, mobile and sensor systems and networking and spatio-temporal data analysis. She has published in a number of top tier conferences and journals in the area and her investigator experience spans more than twenty projects funded by industry and research councils. She has served as organizing and programme committee member of many mobile, sensor systems and analytics conferences and workshops. She sits on the editorial boards of IEEE Pervasive Computing, ACM Transactions on Sensor Networks and ACM Proceedings on Interactive, Mobile, Wearable and Ubiquitous Technologies.

Andreas Kaltenbrunner is the head of the Digital Humanities Department at Eurecat. His research is centered on social media and social network analysis. Andreas uses methods from computer science and the study of complex systems to resolve sociological research questions. He obtained his PhD in Computer Science and Digital Communication in 2008 from the Universitat Pompeu Fabra with a thesis about stochastic effects in human and neural communication patterns. Afterwards he joined the Information Retrieval Group of the technology center Barcelona Media. There he analyzed structural, temporal and behavioral patterns in human interactions in social media websites. He later co-founded in October 2010 as senior researcher the Social Media research line within Barcelona Media and led it from May 2013 onwards. Since June 2015 he leads the Digital Humanities Research Unit at the newly formed technology center Eurecat.