Uncovering the Spatio-Temporal Structure of Social Networks using Cell Phone Records

Luis G. Moyano*, Oscar R. Moll Thomae†, Enrique Frias-Martinez*
*Telefonica Research
Madrid, Spain; Email: moyano,efm@tid.es
†MIT, Boston, MA, USA; Email: rimoll@mit.edu

***Abstract*—Although research in the areas of human mobility and social networks is extensive, our knowledge of the relationship between the mobility and the social network of an individual is very limited, mainly due to the complexity of accessing adequate data to be able to capture both mobility and social interactions. In this paper we present and characterize some of the spatio-temporal features of social networks extracted from a large-scale dataset of cell phone records. Our goal is to measure to which extent individual mobility shapes the characteristics of a social network. Our results show a non-trivial dependence between social network structure and the spatial distribution of its elements. Additionally, we quantify with detail the probability of a contact to be at a certain distance, and find that it may be described in the framework of gravity models, with different decaying rates for urban and interurban scales.**

***Keywords*-Human Mobility; Social Networks; CDR; Gravity Model;**

I. INTRO

Social networks have a fundamental role in our everyday lives. Our social connections determine to a great extent our daily activities, ranging from family to work, from leisure to travel. Understanding of social networks, interepreted broadly as social ties and not only in terms of network applications, has become essential due to the profound implications that their structure can have in human activities [1]. The amount, structure and type of social connections people have depend on a number of external factors such as gender, socio-economic factors, and many others. An important aspect of social networks is that they are geographically embedded [2], a fact that affects and is affected by the structure of the network. The exact mechanisms of this interplay between social structure and the geographical characteristics of its units have been addressed from different points of view [3], [4], [5], [6]. However, the lack of empirical data has been a limiting factor for the validation of any attempt to describe or explain this relationship.

In recent years, there has been a remarkable improvement in the deployment of pervasive infrastructures, such as mobile phones, GPS, online social networks, etc. In this context, and due to their wide coverage and high penetration, cell phones have become one of the main sensors of human behavior. As such, cell phone networks can capture both the social network of an individual (captured as cell phone calls between users) and the spatial characteristics of that individual (captured using the location of each cell phone antenna when phone calls are made). The high penetration of cell phones implies that they can capture a large amount of spatio-temporal relationships at a scale not available to other pervasive infrastructures. This opens the door to characterize how the structure of a social network is related to the mobility of the individuals that defined those social interactions.

Cell phone call records are generated by telecommunication operators for invoice purposes and may be gathered in datasets called Call Detailed Records (CDRs). A considerable amount of research based on CDR analysis have mainly focused on human mobility [6], [7], [8], where the variable under study in these cases is the position of the user, with no information about the user's social contacts. However, to address the relationship between the spatial distribution of a user and her social structure, one should focus in understanding how this structure changes in space. To this end, a central question is to uncover the probability of having a contact located at a certain distance d. This question is well suited to be explored by analyzing cell phone datasets, where the distance d between users is captured at the moment an interaction takes place, d being defined by the two cell phone towers used to deliver the call. Cell phone records contain multiple calls between a given pair of users, and each one of these calls may have associated a potentially different distance.

In other types of datasets, such distance at the time of the interaction is general not available. As a result, the most frequent approach to deal with this multiplicity of distances between two individuals is to select a unique quantity to represent the distance between two users, which may be the distance between homes [5], [9], the distance between zip codes[10], the most frequently used towers [11], and other equivalent measures of average position. As a result, these studies do not consider the full extent of spatio-temporal information but a coarse-grained description, as they assume that the distance between two individuals is constant, which is not the general case. To avoid this limitation, other studies have made use of location-based social networks, where individuals self-publish a location through a given service (such as Foursquare, or any equivalent application) [12], [13], [14], which can be used to characterize spatio-temporal relations. In this case, a user may have several associated positions, but in general these are not more frequent than phone calls, leading to less accurate results.

Previous analysis of spatially-embedded social networks have shown that the probability of two contacts being at a given distance may be well described by the so-called *gravity models*. These models propose that there is an inverse power-law dependence with the distance for certain variables of interest, and in some cases the exponent found has been 2, mimicking the distance dependence of the law of gravity. For instance, the probability of contacts living at distance d [10] or the communication intensity between cities [15] have been found to observe an inverse square law dependence with d. It should be noted that these studies assign a fixed distance between agents (contacts or cities), in some sense providing a static characterization of the spatial distribution of social agents. In this work we focus on a more dynamical point of view, taking into account the actual distance between users in every call. To our knowledge, there is no previous work tackling in a systematic fashion the study of the structure a social network and the actual distances involved in each link at different times. This allows us to quantify this relationship with unprecedented detail reaching unexplored scales. Additionally, we gain information about the statistical structure of the distance improving the understanding of the underlying mechanisms of human mobility and social ties.

II. Related work

Related work in the area of spatio-temporal analysis of social networks is scarce due to the complexity of capturing data that reflects both interactions and mobility. The best part of studies use cell phones as sensors, and generally have two main different approaches: (1) cell phone records are used for the study or (2) data is collected from volunteers, directly from the phone or from the activity of a location-based social network application.

An example of the first approach is, for instance, the work by Wang et al. [11], where the authors study to what extent mobility patterns shape social networks, finding that mobile homophily, network proximity and tie strength strongly correlate with each other. In [10], Lambiotte *et al.* analyze cell phone calls and, using zip codes as a general measure for the location of the user, find that the probability of having a call with a user at distance d is inversely proportional to the square of the distance $P_d \sim d^{-2}$. In [15], Krings *et al.* found an equivalent result for the aggregated call length between two cities using information aggregated at a city level. The previous three approaches have one characteristic in common, they all consider fixed mobility information to characterize a social tie (call), i.e. each individual has a fixed location (which in each case is the most used antenna, zip code or city) when calculating the distance for each interaction. In [5], Cho *et al.* analyze a combination of location-based services and phone calls. Their main result focuses on the distance between users' homes and, even though they compute the distribution of distance at the time of the call, there is no analysis of it, and it is only used for motivation purposes to justify their particular theoretical mobility model. In the present work, we do a full description of the probability of a call being at a given distance.

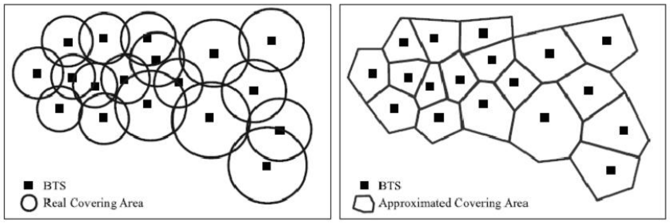
Examples of the second approach can be found in the results presented by Cranshaw *et al.* [16] and Backstrom *et al.* [9]. Cranshaw *et al.* [16] traced 500 individuals and analyzed the location entropy for predicting friendship finding a positive correlation between entropy of locations and common friends. Backstrom *et al.* [9] used information collected from Facebook and user-supplied addresses to predict the location of an individual. In any case this approach has two main limitations: (1) the reduced number of individuals considered for the study and (2) the validity of the results only for the location-based social-network used for the study.

III. Data preliminaries

The data captured by cellular infrastructures provides a key source for investigating large-scale social interactions and human mobility. Cell phone networks are built using a set of base transceiver stations (BTS) that are in charge of connecting cell phone devices with the network. Each BTS tower has a geographical location typically expressed by its latitude and longitude. The area covered by a BTS tower is called a cell. Each cell is typically divided in three sectors, each one covering 120 degrees. At any given moment, one or more BTSs can give coverage to a cell phone. Whenever an individual makes a phone call, the call is routed through a BTS in the area of coverage. The BTS is assigned depending on the network traffic and on the geographic position of the individual.

The geographical area covered by a BTS ranges from less than 1 km^2 in dense urban areas to more than 3 km^2 in rural areas. For simplicity, we assume that the cell of each BTS tower can be approximated with a 2-dimensional non-overlapping region computed using Voronoi tessellation. Figure 1 (left) shows a set of BTSs with the original coverage of each cell, and Figure 1 (right) presents its approximated coverage computed using Voronoi.

CDR databases entries are generated when a mobile phone connected to the network makes or receives a phone call or uses a service (e.g., SMS, MMS, etc.). In the process, and for invoice and legal purposes, the information regarding the time and the BTS tower where the user was located when the call was initiated is logged, which gives an indication of the geographical position of a user at a given moment in time. Note that no information about the exact position of



(left) (right)

Figure 1: (Left) Example of a set of BTSs and their coverage and (Right) Approximated coverage obtained applying Voronoi Tesselation.

a user in a cell is known. Also, no information about cell phone location is known or stored if no interaction is taking place.

From all the data contained in a CDR, our study uses the encrypted originating number, the encrypted destination number, the time and date of the call, the duration of the call, and the latitude and longitude of the BTS tower used by the originating cell phone number and the destination phone number when the interaction happened. Data was sampled for a period of 6 complete weeks from a single telecommunications operator on a national level from a European country. In order to preserve privacy, all the information presented is aggregated and original records are encrypted. No contract or demographic data was considered nor available for this study.

The dataset was pre-processed in order to capture only regular calls between users. In this way, we filtered out calls containing irregular features that may introduce spurious effects. We only considered calls between mobile phones, discarding special numbers and land lines. Since there is missing information about users that don't belong to the operator, we focus our analysis in the graph formed only by users belonging to the operator. Additionally, since we are interested in the social significance of each call we took into account only users that have at least one reciprocal call during the time window spanned by the dataset. Results do not change qualitatively if the number of reciprocal calls are of the order of one. We are interested in the characteristics of the social graph formed by the social ties, so we define a link (and only one link) to be present if there are any number of calls between two users. We call such a link a *contact*, to distinguish it from a *call*, which can occur several times. The number of calls between them is analized separately as a link weight. In other words, there are not multiple links between two users.

Finally, we compute the giant connected component (GC). The largest of the remaining clusters has a number of nodes no larger than 0.002% relative to the nodes in the GC. We discard these, and only consider nodes and links that belong to the GC, to avoid any bias product of the other clusters small relative size. After this preprocessing, the resulting dataset has approximately $69.8M$ number of links, $14.6M$ number of nodes and $404M$ number of total calls.

IV. METHODOLOGY

We provide a detailed exploration of the statistical properties of the different relevant variables to understand the relation between social network properties and the geographical properties of the nodes at the time of a call.

To this end, we compute the social graph from the CDR data by identifying the unique user identification codes with nodes, and pairs of users who have called each other (once or more) with links. Nodes form the set V, and links the set E, so the graph $G = G(V, E)$ represents the social graph of the CDR. Note that a given pair of users, characterized by the link e_{ij}, must have made at least a call in each direction, so we consider that links are undirected. The degree k_i of node i is defined as the number of links of that node, i.e. the number of different contacts user i has.

The total number of calls n_{ij} between a given pair of users i and j can be interpreted as a discrete scalar weight of link e_{ij}, defined in the interval $[1, n_{max}]$, being n_{max} the maximum number of calls recorded in the dataset. Two variables, the distance associated to the call s between users i and j, d_{ijs}, and its duration, t_{ijs}, are associated to a particular link e_{ij}, so there are n_{ij} values of (d_{ij}, n_{ij}) for each link e_{ij}.

As we are interested in comparing variables with different dimensionality, we compute also the average of the distance $\bar{d}_{ij}$ and the average of call duration $\bar{t}_{ij}$ between users i and j:

$$\bar{d}_{ij} = \frac{1}{n_{ij}} \sum_{s=1}^{n_{ij}} d_{ijs} \quad (1)$$

$$\bar{t}_{ij} = \frac{1}{n_{ij}} \sum_{s=1}^{n_{ij}} t_{ijs}, \quad (2)$$

where s goes from 1 to n_{ij}. Additionally, we look at the weighted mean distance $\bar{d}_i$:

$$\bar{d}_i = \frac{1}{k_i} \sum_{j=1}^{k_i} \sum_{s=1}^{n_{ij}} n'_{ij} d_{ijs}, \quad (3)$$

where $n'_{ij} = n_{ij}/N_i$ and $N_i = \sum_{j=1}^{k_i} n_{ij}$ is the total number of calls of user i.

In this context we will present the statistical characteristics of the geographical distance d_{ijs} associated to the s call between users i and j, as well as the associated call duration t_{ijs}. It is useful to note that the distance d_{ijs} is in fact the distance between the BTS towers associated to users i and j when the call s is taken place. Because of this, the error in the distance is very small, of the order of a meter. This error should not be confused with the uncertainty in the actual distance between users, which depends among other factors on the density of BTS towers, and which will be analized in future work. Moreover, the error in the duration of the call, which is always measured in seconds, is very low as well, being of the order of the second. In order to compare call variables d_{ijs}, t_{ijs} with link variables n_{ij}, we compute the mean of those quantities $\bar{d}_{ij}$, $\bar{t}_{ij}$. In the same way, to compare link variables (n_{ij} and the above mentioned d_{ijs} and t_{ijs}) with node variables k_i we compute weighted means $\bar{n}_i$, $\bar{d}_i$, $\bar{t}_i$, where the bar denotes that it is an averaged quantity, and the sub-index determines the nature of the average, which reflects appropriately its central tendency.

To understand the structure of the underlying social graph of the CDR dataset, we measured some of its most relevant characteristics, mainly the different probability distribution for the variables of interest. Furthermore, we focus in the relationship between the social network and the distance properties of the users, turning our attention to the conditional probability distributions that allow for a direct visualization of the relationwhip between the variables. Importantly, we observe two different regimes in the distribution of call distance. We discuss the validity of gravity models in this level of description and provide insights as to what the underlying causes of these two regimes could represent.

V. Results

A. General characterization of the social network

Let us recall that the number of different contacts of user i corresponds to the degree k_i of node i in the social graph. In Fig. 2 we show the complementary cumulative probability distribution (CCDF) of the degree for the complete social graph. We observe that it decays slower than exponentially, a sign of non-trivial graph structure, although it is not clear that the tail is a power-law. As in many other instances of social networks, the heterogeneity of the distribution shows that a few users present a large number of different contacts, while a large set of users show just a few contacts in the time window spanned by our data. This dependence of the degree is consistent with other results for cell phone calls [13] or even other types of social networks [14].

Fig. 3 shows the distribution of the number of calls between any two users i, j, $P(n_{ij})$. We see that it also presents signs of non-trivial structure, with a decay slower than exponential, pointing to the fact that some users have a remarkably larger number of calls than others. Interestingly, the number of calls of every user in the social network span three orders of magnitude, from just a few calls during the 6 week window, to almost 1000 calls, nearly an average of 1 call per hour. Even though the precise shape of these CCDFs is out of the scope of this work, it is nevertheless an important feature that can provide deeper insight into the social network structure.

To assess how these two variables are related it is instructive to show the probability of the mean number of calls given that the node has a degree in some interval of interest. In particular we choose two set of nodes, one corresponding to those with lowest values of the degree ($k \leq 3$), the other one corresponding with high degree ($k \geq 45$). We have chosen extreme values for the sets to represent different social structure characteristics, while maintaining enough statistics to have well defined distributions. In solid dots, the probability distribution of the number of calls for low degrees has a distinct heavy-tail, i.e. most low-connected users make very few calls, some making a large number of calls. On the other hand, the most connected users present a completely different functional form, with a precise mode

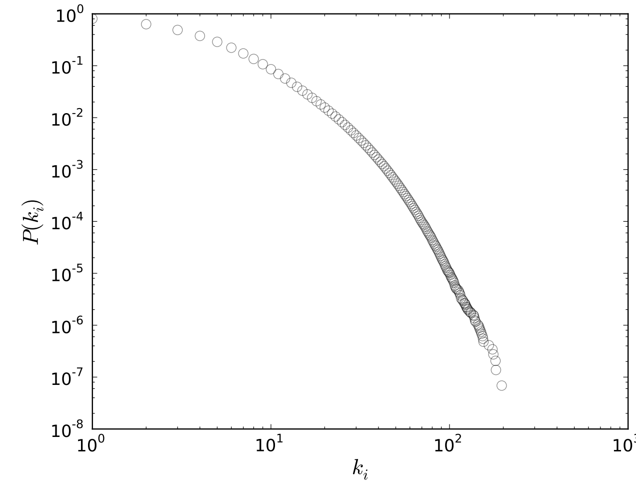


Figure 2: Complementary cumulative probability distribution of degree k_i. The distribution does not have an exponential decay, a characteristic of social networks.

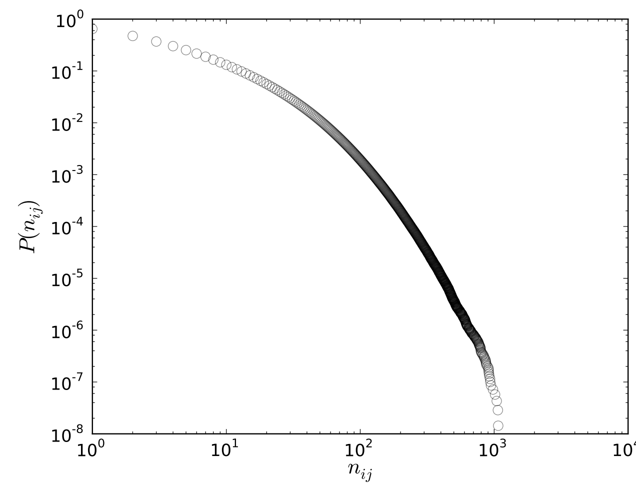


Figure 3: Complementary cumulative probability distribution of the number of calls n_{ij}.

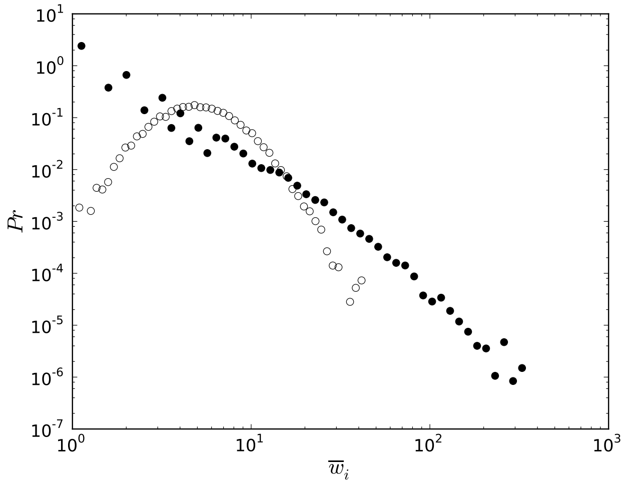


Figure 4: Probability distribution of the mean number of calls $\bar{n_i}$ given that $k_i \geq 45$ (open dots) and $k_i \leq 3$ (solid dots). See Table I for more information.

Table I: Table of quantiles for Fig. 4

	p20	p40	p50	p60	p80
$P(\bar{n_i}\|k_i \leq 3)$	1	2	2.67	3.5	6.67
$P(\bar{n_i}\|k_i \geq 45)$	3.96	5.18	5.82	6.54	8.62

approximately at 6 calls. Very few of them make few calls, and also very few make a very large number of calls.

As mentioned in Sec. IV, each link e_{ij} has associated two sets of dimension n_{ij}, one with the duration of each call, the other one with the call's distance, which we will discuss later. In Fig. 5 we show the distribution of the duration, in seconds, associated to every call. The distribution is also quite broad and presents a maximum approximately at 14 s. Calls lasting 1 and 2 seconds are possibly over-represented, because in most cases they are not intended but a misplaced call or something of equivalent spurious nature. It is worth highlighting that the presence of a heavy-tail implies a non-negligible probability of calls on a wide range of t_{ijs}, some lasting over 8 hours, spanning almost fiver orders of magnitude.

B. Interplay between geographical distance and social network structure

The fact that there is an interplay between our social network and the way we move spatially should be intuitive. It is understood that most of our movements are directly or indirectly related to our social connections, from going home or work, out for leisure or travel, generally speaking, these movements are connected to some extent with someone else in our network.

A more difficult question is to precisely characterize this relationship. Here we approach this problem from a statistical point of view, and we try to answer the question by analyzing the probability distribution for the distance at the time of a call for different sets of users with common social traits. These conditional probability distributions allow us to understand qualitatively and quantitatively this interdependence.

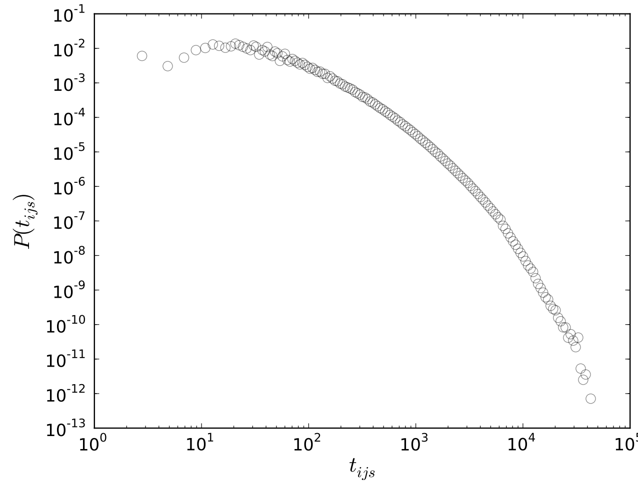


Figure 5: Probability distribution of call duration t_{ijs} in seconds.

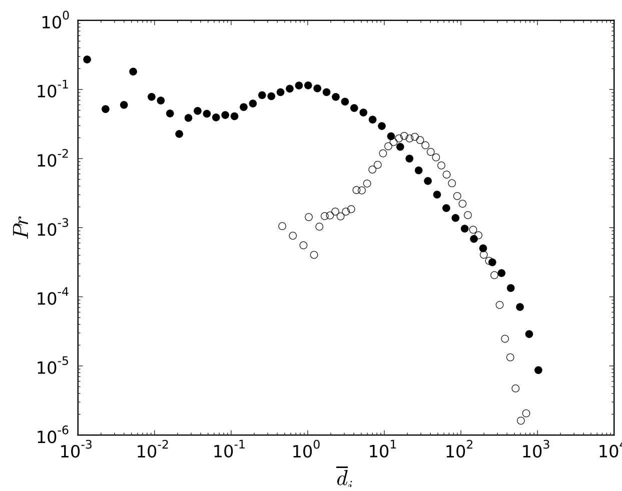


Figure 6: Conditional probability distributions: $P(\bar{d}_i|k_i \leq 3)$ (solid dots) and $P(\bar{d}_i|k_i \geq 45)$ (open dots), where $\bar{d}_i$ is measured in km. See Table II for more information.

One of the most direct ways to characterize where people are relative to each other depending on the structure of the network is to relate distance at the time of the call and node degree k_i. We show in Fig.6 the conditional probability density functions $P(\bar{d}_i|k_i \geq 45)$ (open dots) and $P(\bar{d}_i|k_i \leq 3)$ (solid dots).

We observe a significant difference between both dis-

Table II: Table of quantiles for Fig. 6

	p20	p40	p50	p60	p80
$P(d_i \vert k_i \leq 3)$	1.97	5.99	9.37	14.83	50.24
$P(\bar{d}_i \vert k_i \geq 45)$	17.77	27.84	34.09	42.16	70.79

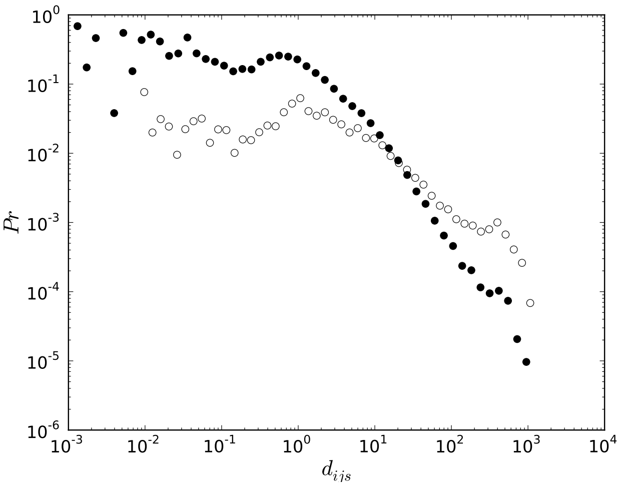


Figure 7: Conditional probability distributions: $P(d_{ijs}|t_{ijs} \leq 10s)$ (solid dots) and $P(d_{ijs}|t_{ijs} \geq 8000s)$ (open dots), where d_{ijs} is measured in km and t_{ijs} is measured in seconds. See Table III for more information.

Table III: Table of quantiles for Fig. 7

	p20	p40	p50	p60	p80
$P(d_{ijs} \vert t_{ijs} \leq 10)$	0	0.34	0.98	2.02	8.65
$P(d_{ijs} \vert t_{ijs} \geq 8000)$	4.12	22.28	51.47	129.21	380.15

tributions, in particular, the modes of the distribution are remarkably apart, being 1 km for nodes with low degree and approximately 20 km for nodes with large degree (hubs), an order of magnitude of difference. Our result shows that users with a large number of contacts have a significantly larger probability of calling someone farther away than users with low number of contacts. This relates to the fact that the geographical scope or influence area of a user is related in a proportional way to the number of contacts that user has. As for the number of calls, the distribution for low degree users is broader than the one the most connected users, for which its standard deviation is much smaller and the probability is concentrated around its maximum, yielding a more representative scale. An effect that it is important to comment is that many users that are nearby when calling use the same tower and are assigned a $d = 0$, which in turn makes $\bar{d}$ take values below the resolution of the towers. This is particularly the case of low connected users.

It is also instructive to compare the relationship between the distance at the time of the call with the duration of that call. In Fig. 7 we show the probability density functions $P(d_{ijs}|t_{ijs} \leq 10s)$ (solid dots) and $P(d_{ijs}|t_{ijs} \geq 8000s)$ (open dots). Again, the two distributions show different characteristics, suggesting that the two variables are strongly correlated. Both distributions decay with a heavy-tail, but the probability of distance for long calls decays less rapidly than the one for short calls. This result is along what could be expected, since a call between distant users is possible to have a conversational nature, and thus the users may spend more time in it, whether a call for nearby users possibly relates to some kind of coordination, where, for instance, a user delivers a short instructive message or notice.

Interestingly, the above results could be interpreted by supposing that calls between distant people are infrequent and this causes them to be longer. To test this hypothesis, we computed the probability distribution of the mean call duration and the total call duration between users, given that the number of calls between them is higher (and lower) than appropriately chosen values. These distributions are shown in Fig. 8, upper and lower panels respectively. As before, solid dots represent the set with low number of calls, while open dots the set with higher number of calls. We find that the pairs of users that present a higher number of calls talk consistently more time than users that call each other sporadically. Mean call duration presents a maximum value at approximately less than 30s for low number of calls, and 70s for large number of calls, while total time duration is slightly below 30s for infrequent callers and near 10000s for frequent callers. To conclude, highly connected users talk more in total and in average than low connected users, and do so with contacts farther away.

C. *Probability of distance at the time of call*

The probability density function of the distances associated to every call is shown in Fig. 9

The distribution is quite broad, presenting a heavy-tail with two apparent different regions or regimes, one approximately between 1 km and 10 km, and another from 10 km onward, both compatible with power-law decay. The first range decreases as $d^{-0.77}$ while the second region decreases as $d^{-1.5}$. The mode of the distribution is approximately at 0.9 km, and the tail extends almost up to 10^3 km. It is worth to mention that the distribution observed here may be the result of the type of structure in Fig. 7, a point that is being currently studied.

These results are in the same line with previous work on equivalent measures of distance between users even though, as discussed in Sec. II, distance is not always defined in the same way and thus may not be directly comparable, a point that remains to be clarified. The authors in [13] report similar findings with $d^{-0.5}$, where in this case d corresponds to distance between user check-ins. The authors in [14] report $d^{-1} + \epsilon$, in this case the distance is associated to a fixed geographical location made public by the user. In [9] similar findings are shown where the dependence decreases as d^{-1},

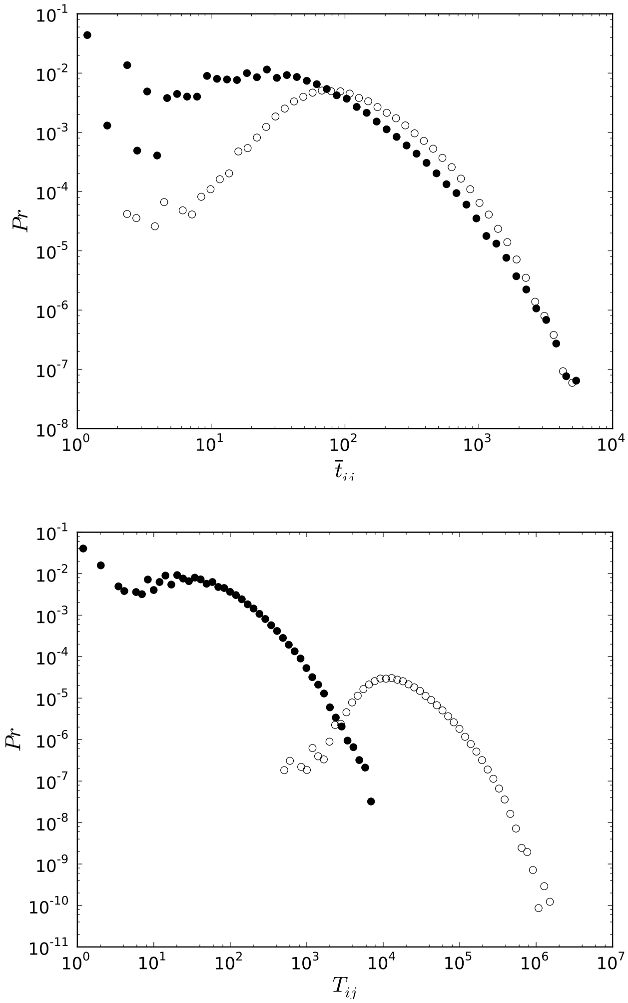


Figure 8: Probability distribution of mean call duration (upper panel) and total call duration (lower panel) given that the number of calls $n_{ij} \leq 2$ and $n_{ij} \geq 125$.

and the distance is a user supplied address. Interestingly, as pointed out before, the authors in [5] do report distance at the moment of call for a CDR dataset, but with no description or quantification of the distribution. Nevertheless, our results appear to be compatible with theirs.

As mentioned in Sec. I, [10] and [15] explore a framework known as gravity models and find that the probability of a call between users that live at a distance d decays as d^{-2}. Our results show that when the variable under study is the actual distance between users when they initiate a call, the dependence can still be interpreted as power-law decay, but in contrast, there are two regimes depending on the scale considered, and with different decay rates. Two or more regimes where also found in [17], where they focus on finding a unified functional form for all the regimes observed. In [18], results from [6] are reinterpreted by proposing different

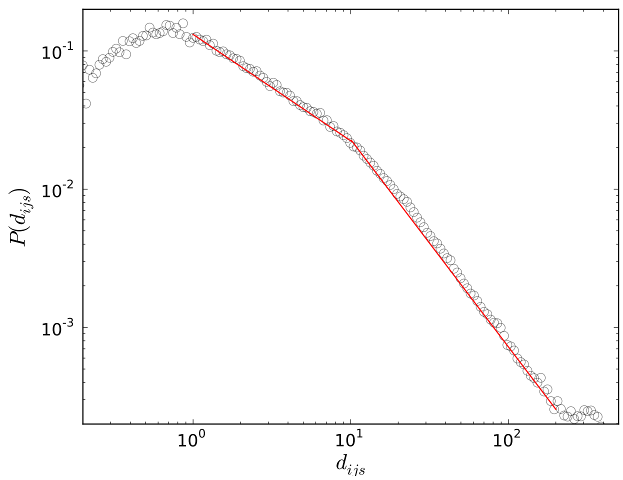


Figure 9: Probability distribution of the distance d_{ijs} (in km) associated to a call. The region approximately between 1 km and 10 km may be described by a power-law decay with exponent $\alpha = -0.77$, while the region from 10 km onward shows a faster decay consistent with an exponent $\alpha = -1.5$.

power-law exponents to different distance scales. However, the model presents three integer exponents, a result that is not found in our observations. We believe that our findings are compatible with urban/interurban distances (being the threshold between the two approximately at 10km), which points toward different mechanisms associated to the two intervals. The study of such mechanisms are subject of ongoing work.

VI. Conclusion

A thorough understanding of human mobility is necessary for a wide range of critical areas, such as city planning, network dimensioning, public health development, prevention of large-scale catastrophes, and many others. Every step forward in uncovering the basic mechanisms of human displacement can have a major impact in these and other important areas. In this work we have presented an initial characterization of a massive cell phone dataset, in order to understand the relationship between social network structure and the dynamical spatial distribution of its constituents.

We focused in the distance at the time of the call, which describes in detail the spatial situation of the users associated to the call, avoiding the use of a static proxy for their location. We provide an in-depth statistical analysis of the social graph configuration underlying the call dataset, along with other relevant variables such as link weights in the form of call duration and number of calls.

Our results describe quantitatively that users with a higher number of connections systematically have these contacts farther away than users connected just to a few others.

Additionally, the mean number of calls is strongly correlated with the degree of the user, showing a well defined heavy-tail in the case of users with low number of connections, while distributed around a clear maximum in the case of highly connected users, which make consistently more calls in average.

Moreover, the distance of a call is strongly correlated with its duration, being short calls nearby much more probable than the distant ones.

Finally, we discuss the fact that the probability of two users being at distance d at the time of a call decreases with the distance approximately as $d^{-0.77}$ for scales compatible with urban distances, and as $d^{-1.5}$ for scales compatible with interurban distances This power-law dependence has been put forward before for related variables, under the name of gravity models, the exact value of the exponent found is not agreed upon and possibly depends on the exact definition of the distance. Our work extends these previous results to a more dynamical framework that implies a finer scale of description. The fact that the statistical behavior of the distance is strongly dependent on the urban/interurban structure underlying human displacement suggests that different microscopic mechanisms may be in play and should be taken into account for an accurate modeling of the probability of two contacts being at a certain distance apart.

Future work aims at including in the analysis the complete time dependence encoded in the timestamp associated to the phone call, in order to properly model the full set of dependencies between spatio-temporal characteristics and social network structure. Such study will help improve mobility prediction models, event detection algorithms, behavioral analysis of agents in urban/interurban environments, and related problems. Furthermore, applications to modeling of geographical influence of users, epidemic spreading analysis, among others, are also possible.

REFERENCES

[1] M. Newman, *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc., 2010.

[2] M. Barthélemy, "Spatial networks," *Physics Reports*, vol. 499, no. 1, pp. 1–101, 2011.

[3] B. Wellman and K. Hampton, "Living networked on and offline," *Contemporary Sociology*, vol. 28, no. 6, pp. 648–654, 1999.

[4] J. Goldenberg and M. Levy, "Distance is not dead: Social interaction and geographical distance in the internet era," *Arxiv preprint arXiv:0906.3202*, 2009.

[5] E. Cho, S. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011, pp. 1082–1090.

[6] M. Gonzalez, C. Hidalgo, and A. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[7] C. Song, Z. Qu, N. Blumm, and A. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[8] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, no. 7075, pp. 462–465, 2006.

[9] L. Backstrom, E. Sun, and C. Marlow, "Find me if you can: improving geographical prediction with social and spatial proximity," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 61–70.

[10] R. Lambiotte, V. Blondel, C. De Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. Van Dooren, "Geographical dispersal of mobile communication networks," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 21, pp. 5317–5325, 2008.

[11] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A. Barabási, "Human mobility, social ties, and link prediction," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011, pp. 1100–1108.

[12] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo, "A tale of many cities: universal patterns in human urban mobility," *PloS one*, vol. 7, no. 5, p. e37027, 2012.

[13] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo, "Socio-spatial properties of online location-based social networks," *Proceedings of ICWSM*, vol. 11, pp. 329–336, 2011.

[14] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins, "Geographic routing in social networks," *Proceedings of the National Academy of Sciences*, vol. 102, no. 33, pp. 11 623–11 628, 2005.

[15] G. Krings, F. Calabrese, C. Ratti, and V. Blondel, "Urban gravity: a model for inter-city telecommunication flows," *J. Stat. Mech*, p. L07003, 2009.

[16] J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh, "Bridging the gap between physical location and online social networks," in *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, 2010, pp. 119–128.

[17] M. Seshadri, S. Machiraju, A. Sridharan, J. Bolot, C. Faloutsos, and J. Leskove, "Mobile call graphs: beyond power-law and lognormal distributions," in *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008, pp. 596–604.

[18] N. Scafetta, "Understanding the complexity of the lévy-walk nature of human mobility with a multi-scale cost/benefit model," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 21, no. 4, pp. 043 106–043 106, 2011.