CrossMark

Calling, texting, and moving: multidimensional interactions of mobile phone users

Matteo Zignani[*], Christian Quadri, Sabrina Gaito and Gian Paolo Rossi

*Correspondence:
matteo.zignani@unimi.it
Department of Computer Science,
University of Milan, Via Comelico
39/41, Milan, Italy

Abstract

The communication networks obtained by using mobile phone datasets have drawn increasing attention in recent years. Studies have led to important advances in understanding the behavior of mobile users although they have just considered text message (short message service (SMS)), call data, and spatial proximity, separately. However, there is a growing awareness that human sociality is expressed simultaneously on multiple layers, each corresponding to a specific way an individual has to communicate. In fact, besides the common real life encounters, a mobile phone user has at least two further communication media to exploit, SMSs and voice calls. This is advocating a multidimensional approach if we are seeking a compound description of the human mobile social behavior.

In this context, we perform the first study of the multiplex mobile network, gathered from the records of both call and text message activities, along with relevant geographical information, of millions of users of a large mobile phone operator over a period of 12 weeks. By computing a set of complex network metrics, at different scales, onto the three single layers given by calls, SMSs and spatial proximity, and their extensions onto a three-level network, we provide a comprehensive study of the global multi-layered network which arises from both the overall on-the-phone communications performed by mobile users and their spatial propinquity.

Keywords: Multiplex network; Mobile phone graph; Social network analysis; Co-location graph; Voice call; Text message; Communication network

Introduction

In recent years, we witnessed the growing awareness of the fact that human communications and social interactions are built on a stratified structure [1]. Today, a variety of techno-communication channels—including online social networks, mobile phone calls, short message services (SMSs), and e-mails—provides an intricate bundle of interactions that is overlaid on real life relationships enabled by individuals' spatial proximity. Among all, communication networks constructed on top of mobile phone interactions have attracted increasing research activities in recent years, becoming a relevant topic in the computational social science [2]. Results have led to important advances in understanding the communication behaviors of mobile users [3, 4] at different scales. For instance, the structural properties of the mobile phone graphs have been investigated

by Nanavati et al. [5] and by Onnela et al. [3], whose studies represent the first attempt to analyze large social networks as they emerge from mobile communications. On the other side, a great effort has been devoted to study the local properties of mobile phone graphs. Many researchers proposed measures to characterize the properties of the links: the burstiness level [6, 7] and the link persistence [4] are often used to describe link dynamics, whereas the link overlap [3] captures the role of the link with respect to (w.r.t.) dense/sparse structures. Nonetheless, most of these studies limit their analysis to one or two ways of communication, while an all-around vision is still missing. Mainly, studies consider the set of voice call data, while, for instance, text messages (either instant messages or SMSs) have been considered rarely and separately from phone calls [5, 8], when analyzing human communications and spatial proximity.

Mobile phone data also provide information about users' mobility, leading to studies which combine human movement and communication patterns. A few of them have stressed the interplay between users who call each other and their geographical proximity. For instance, the analysis of Phithakkitnukoon et al. [9] reveals that most of the places visited by a person are close to their friends' positions, while Calabrese et al. [10] and Wang et al. [11] show that the frequency of encounters between users is highly correlated with their frequency of calls. The above works mainly focus on link and geographical proximity; however, more complex structures like communities have been related to the geographic position. At a country level, Caughlin et al. [12] studied a Dominican Republic mobile phone communication network to determine whether the geographic context can explain the community membership, while Expert et al. [13] have been able to divide the Flemish and the French communities by adapting the modularity function to deal with spatial networks. Finally, Onnela et al. [14] have shown that small social groups are geographically very tight but become much more clumped when the group size exceeds about 30 members.

In this paper, we take the first step in the direction of a multiple layer approach by performing a combined analysis of the networks obtained from SMSs, voice phone calls, and spatial proximity at a metropolitan scale. This study is based on the multiplex network gathered from a large anonymized dataset of call detail records (CDRs) containing voice call and SMS activities and related spatial information of nearly one million mobile subscribers over a period lasting 12 weeks in 2012. Data have been structured like a network of three networks [15] and formally described by a directed multigraph. By computing a set of complex network metrics, at different scales, onto the three single networks and their extensions onto the multiplex network [16], we contribute to some findings on human behavior in the different dimensions captured by mobile phone data.

First, we show that the two single layers describing on-phone interactions, SMSs and calls, are macroscopically similar as far as regards the connected components, but they are microscopically different. In fact, the two single networks do not perfectly overlap, nor one is included in the other, while they rather partially overlap, since many users use a communication medium only (call or SMS). User ego-networks perceivably enlarge in the multiplex network, confirming that both communication media are needed to get a complete vision of the users behavior captured by a mobile phone dataset. Besides, as far as regards in-degree and out-degree distributions, it turns out that the SMS graph behaves more similarly to online social networks, while the call graph is more similar to Web graph. Second, we introduce the notion of multidimensional link reciprocity into the

Zignani *et al. Computational Social Networks* (2015) 2:13

Page 3 of 24

set of metrics for multiplex networks [17]. We show that interactions by mobile phone are much more reciprocal and thus social [18] than what could be speculated if only calls were considered. Nevertheless, reciprocity is much lower than observed in online social networks and in the Web. As third contribution, we add a third layer given by the spatial proximity, obtaining a denser graph than on-phone communication graphs and we find that people communicating by phone are more likely to be in spatial proximity w.r.t. individuals who do not interact through any mobile medium. In particular, interactions by SMSs are more predictive of spatial proximity than calls. However, a correlation analysis between the degree of co-location (proximity) and the strength of the communications between co-located pairs let emerge a novel result: in a metropolitan area like Milan, people who are strongly spatially close do not need to frequently communicate, as observed in other studies [10, 11]. At most, the frequency of communications increases when people not frequently share some locations. Fourth, we study the correlation between the different centralities of the mobile phone users in each layer. The results confirm that the network which merges on-the-phone communications and spatial proximity is made by loosely coupled layers in terms of degree and strength. Finally, we investigate the impact of multiplexity also at mesoscopic scale by performing a community detection analysis on each layer and on the multiplex networks. It comes out that communities at different layers do not match. Among them, the SMS communities are more representative of groups of people sharing the same interest than the call ones, being call-based communities weaker. This finding also explains why we observed that the communities extracted from the multiplex network are mainly pivoted on SMS communities. In general, SMSs are used by pairs and groups of people with closer relationships that lead them also to meet.

The paper is organized as follows: in the "Dataset" section, we describe the mobile phone dataset; in the "Network definitions" section, we introduce the notation and the definitions to cope with multidimensionality; in the "Networks characterization" section, we microscopically and macroscopically characterize each network separately in terms of node sets, connected pair sets, degree distribution, multireciprocity, and connected components; then, we measure the interplay among the different layers by correlating node centralities and the weights associated to the links. Finally, in the "Community" section, we analyze the overlapping between the communities extracted from the different networks we take into account.

Dataset

This paper is based on a large anonymized dataset of call detail records (CDRs) concerning voice calls and short text messages (SMS) of about one million subscribers of an international mobile operator in the metropolitan area of Milan and collected from March 26 to May 31, 2012. The dataset contains an overall amount of more than 63 million phone-call records and 20 million SMS records.

Unlike [19], where we took into account the CDR information about call and SMS interactions only, here, we also add the spatio-temporal proximity dimension induced by the human mobility. To include the spatio-temporal information, each entry of the CDR is described by the 5-ple $t_{CDR} = \langle s, r, t_{start}, d, loc \rangle$, where s and r, respectively, represent the sender and the receiver of the call/sms, t_{start} is the initial time of the activity (when the call starts or a SMS is sent), d is the duration, and loc is the serving cell the user s

is attached to. We consider the cell tower *loc* as a proxy for the physical position of the user *s*. [1] SMS duration is zero, while nearly 40 % of calls have duration equal to 0. Besides missed or unanswered calls, such a large amount of rings is reckoning with a common practice in Italy to use rings for meaning 'call me back soon" or "I have just arrived", for instance, to get synchronized at a meeting. Due to the difficulty in detecting 0-duration calls which express significant interactions, we remove these records from the dataset. Furthermore, according to the literature on mobile phone cleansing [2, 6, 18], we filtered out calls involving other mobile operators to mitigate the bias between operators. Moreover, the positions of the users not supplied by the mobile operator are incomplete as well as the evaluation of the interplay between mobility and the communication activities.

On the basis of the CDR dataset, we construct two preliminary on-phone communication networks, one for each channel, that are then processed to extract only relevant interactions. To the purpose, in the call graph, we consider the pairs of users whose sum of the call durations exceeds the minute and whose total number of interactions is greater than three. This way, we discard pairs with on average one interaction per month [18] and who exchange a small quantity of information. For the same reasons, in the SMS graph, the only relevant pairs are those with a total number of interactions greater than three. After the data cleansing, the analysis is performed on a whole population of about 420,000 people generating almost seven millions calls, 317,000 h of conversations, and four millions SMSs. Furthermore, as shown in the "Networks characterization" section, we obtain maximal degrees similar to reciprocated networks, keeping at the same the directionality of the links.

Previous data are exploited to evaluate and study human interactions facilitated by mobile phones and implicitly to measure the topological closeness of the mobile phone users in a techno-communication dimension. Meanwhile, mobility data allows us to capture some degree of closeness between two individuals in a different dimension: the physical space. The degree of physical proximity that can be extracted from a mobile phone dataset does not equal to measure face-to-face contacts nor to quantify physical interactions. However, the higher spatial granularity of cellular towers in a metropolitan area[2], w.r.t. other mobile phone datasets [9, 11, 14, 20], allows us to obtain more precise results and levels of proximity closer to physical interactions.

To obtain a reasonable value which expresses the closeness of a pair of users in the physical space, we leverage the *co-location rate (CoL)* as defined in [11]. Given $n(u)$, the set of CDR tuples such that u is equal to the sender s, we define *CoL* as:

$$CoL(u,v) = \frac{\sum_{i \in n(u)} \sum_{j \in n(v)} \Theta(T - |t_{start}(i) - t_{start}(j)|)\sigma(loc(i), loc(j))}{\sum_{i \in n(u)} \sum_{j \in n(v)} \Theta(T - |t_{start}(i) - t_{start}(j)|)} \qquad (1)$$

where $\Theta(t)$ is the Heaviside function, σ is the indicator function, and T is a time window. To obtain a proximity measure closer to face-to-face interactions, we set $T = 30$ min (lower threshold w.r.t. literature [11]). The co-location rate takes into account the spatio-temporal proximity given by the simultaneous presence of the pair (u, v) at the same cell tower, normalized by the number of times users u and v are both observed during the same time window. Moreover, to avoid the bias given by a low numerator and denominator in Eq. 1, we discard *CoL* values which correspond to a numerator less than 10. This way, we obtain more than two million pairs with $CoLoc > 0$, from the initial 170 millions.

Network definitions

Interactions among individuals take place through a variety of communication channels and can be described by as much different networks, thus leading to a multiplex graph. Although our dataset provides information about only three levels, in this section, we introduce a general definition of edge-labeled multigraph which covers any multi-layered setting. To represent the directional nature of the communications, we consider only directed networks without any labels on vertices. The same holds for the physical proximity graph, since it is possible to represent an undirected graph as a digraph.

Definition 1. *An edge-labeled directed multigraph is a tuple $\mathcal{D} = (V, E, D, l)$ where V is the set of vertices, $E \subseteq V \times V \times D$, with D, the set of dimensions or layers, is the set of labeled directed edges, and $l : E \rightarrow S$ is a mapping assigning an element $s \in S$ to an edge $(u, v, d) \in E$.*

Likewise, the single layer case, we can extract from a directed multigraph its undirected graph by removing the direction in each layer d and by introducing a function which merges the labels of the edges whenever a link is bidirectional.

Given an edge-labeled directed multigraph, we may need to extract only a specific layer or consider them separately. This occurs, for instance, when comparing the properties of different networks or evaluating the importance of a vertex in a specific layer. To this purpose, we provide the definition of d-network layer.

Definition 2. *Given an edge-labeled directed multigraph $\mathcal{D} = (V, E, D, l)$ and $d \in D$, we define the d-network layer D_d as the graph $D_d = (V_d, E_d)$ where $E_d = \{(u, v) \in V \times V \mid (u, v, d) \in E\}$ and $V_d = \{u, v \in V \mid (u, v) \in E_d\}$. A similar definition holds for undirected multigraphs.*

The multigraph and the network layer definitions model multi-layered settings. To adapt these definitions to the three-layer case given by the voice call, the SMS, and the co-location layers, we set $D = \{c, s, loc\}$, where c, s, and loc, respectively stand for call, SMS, and co-location. In particular, we denote D_c as *call graph* and D_s as *SMS graph*. The co-location network, instead, is described by the undirected graph G_{loc} whose nodes are contained in $V_{loc} = V_c \cup V_s$ and an edge (u, v) exists if $CoL(u, v) > 0$.

We finally define a mapping function l which enables the modeling of the strength of the on-phone interactions and the level of co-location. l is defined as follows:

Definition 3. *Given an ordered pair $< f^c(u, v), \delta(u, v) > \in \mathbb{R}^2$, $f^s(u, v) \in \mathbb{R}$, and $CoL(u, v) \in [0, 1]$,*

$$
l(u, v, d) = \begin{cases} < f^c(u, v), \delta(u, v) > & d = c \\ f^s(u, v) & d = s \\ CoL(u, v) & d = loc \end{cases}
$$

where $f^c(u, v)$ and $f^s(u, v)$ are the number of calls and SMSs from u to v, respectively, and $\delta(u, v)$ is the aggregated duration of the conversations when u calls v.

Zignani *et al. Computational Social Networks* (2015) 2:13

Page 6 of 24

Although the previous mapping definition captures both duration and frequency, in the following, we mainly consider frequency only. This way, weight variables on the different communication dimensions, call and SMS, are comparable. To summarize the objects we are dealing with, in the following sections, besides D_c, D_s, and G_{loc}, we take into account two edge-labeled directed multigraph $\mathcal{D}_{cs}$ and $\mathcal{D}_{csloc}$. The former models the overall on-phone interactions since it includes the call and the SMS layers. The latter represents the overall multiplex network captured by the mobile phone dataset since $\mathcal{D}_{cs}$ is merged with the co-location network.

Networks characterization

Before dealing with many layers simultaneously, we analyze each property in each dimension, separately. This approach is intended to highlight the importance of all dimensions in describing the human behavior in a mobile phone dataset. In this section, we characterize mobile users in terms of communication media usage and physical presence in the city of Milan by measuring how the three networks overlap. That allows us to say which interactions between SMS or call are more predictive of the spatial proximity. The analysis of the layers overlapping stresses the importance of a multilayer approach since we observe that if we focus our attention just on a single dimension, we lose a few interactions. These findings also impact on the degree centrality. We ask whether active users, for example hard texters, are as active in the other communication layer or whether users with a high degree of proximity with many people are equally important into the techno-communication dimension, and we find that the layers are loosely coupled. Finally, we find that at the macroscopical level, the networks are similar despite their microscopical differences.

Networks size and order

In Table 1, we report the basic properties of D_c, D_s, G_{loc}, $\mathcal{D}_{cs}$, and $\mathcal{D}_{csloc}$. The order and the size of the SMS and the call graphs indicate that the on-phone communication layers are very sparse. As for the co-location network, it is completely different from the previous graphs since nodes are less, but the number of links is much higher, i.e., the co-location network is denser, and many nodes are not physically co-located, as it will be discussed afterwards.

To compare our results with the literature on phone calls, we also report the order and the size of the giant weakly connected components. Generally, the percentages of nodes in the giant components are comparable with other mobile phone graphs [5], and the SMS

Table 1 Basic properties of the D_c, D_s, G_{loc}, $\mathcal{D}_{cs}$, $\mathcal{D}_{csloc}$, and of their giant weakly connected components. $|V_{gwcc}|$ and $|E_{gwcc}|$ represent the number of nodes and edges of the giant weakly connected component (*gwcc*), respectively. *med(k)* indicates the median degree, and *med(f)* denotes the median strength. The column *Perc.* reports the ratio between the number of nodes in the giant component and in the whole network

Type	$|V|$	$|E|$	*med(k)*	*med(f)*	$|V_{gwcc}|$	$|E_{gwcc}|$	Perc. (%)	*med(k)*	*med(f)*
D_c	394,834	1,098,774	2	9	356,895	1,070,576	90	3	10
D_s	272,310	575,555	2	7	221,210	537,600	81	3	8
G_{loc}	197,216	2,477,564	5	–	194,238	2,476,041	98	5	–
$\mathcal{D}_{cs}$	417,728	1,674,329	3	12	383,659	1,643,518	91	5	14
$\mathcal{D}_{csloc}$	417,728	4,151,893	3	–	395,040	6,657,428	95	4	–

Zignani *et al. Computational Social Networks* (2015) 2:13

Page 7 of 24

graph is less connected w.r.t. the call graph. Observing the overall number of nodes (first column in Table 1), the multilayer graph $\mathcal{D}_{cs}$ includes more users w.r.t. the single layers. For instance, the social network built on voice calls loses about 5 % of users in the mobile network up to 7 % of nodes when considering the giant component. The phenomenon is further amplified if we consider $\mathcal{D}_{csloc}$. Here, the loss in the voice call graph is 9 %, and the introduction of the spatio-temporal proximity links increases the number of nodes weakly connected (more 3 % of nodes w.r.t $\mathcal{D}_{cs}$).

Node sets

The analysis of the node set in D_s, D_c, and G_{loc}, respectively, allows us to highlight the users' habits in terms of usage and preference of the communication media. Do all users adopt both media or someone prefers to communicate by call or SMS only?

The number of nodes in the different networks in Table 1 suggests that V_c and V_s do not perfectly overlap, nor one is included in the other, rather they partially overlap. In fact, we find that 35 % of active users adopt only the call as interaction media, while the exclusive use of text messages involves 5 % of active users. In general, we observe that many users perceive the two communication media as different. About 40 % of active users prefers only an exclusive communication medium (call or SMS), while the remaining ones prefer integrating the peculiarities of the two communication channels. In fact, calls are more instinctive and similar to face-to-face conversations, whereas text messages are more intimate and allow a greater level of reflection [1].

We extend the overlapping evaluation to the set of people who are co-located (V_{loc}). V_{loc} represents a proper subset of the mobile phone users by construction (see the "Network definitions" section) and contains 47 % of mobile phone users. The remaining 53 %, who interact through the cellular network infrastructure in Milan, do not live mostly in the city or are passing by. Consequently, we refine the above analysis on the exclusive usage of the communication media by focusing only on the users who share locations in Milan. In this case, 40 % of the users share with at least another user a different location in Milan and use both SMSs and calls, whereas 0.6 and 6 % adopt exclusively SMSs or voice calls, respectively.

Link sets

About the analysis of the link sets, in the following, we consider the connected pairs, i.e., (u, v) such that at least a link between u and v or v and u exists. By abusing of notation, we denote as E_c, E_s, and E_{loc} the set of the connected pairs in each layer, respectively. In this case, we do not take into account the direction of the links because we focus only on the interplay between the spatio-temporal proximity and the communication media adopted to interact.

The first quantity we investigate is the number of connected pairs in the techno-communication directed multigraph $\mathcal{D}_{cs}$ that happen to be co-located, i.e., $|(E_c \cup E_s) \cap E_{loc}|$ (see Table 2). We find that the probability of interacting by media when in physical proximity is equal to 0.06. This low value is justified by the spatial granularity of the cellular cells and by the overcrowding of some locations, in particular gathering places where the trajectories of a large amount of people intercept one another. Similarly, given a pair interacting by SMSs or voice calls, we evaluate the probability of being co-located equals 0.14^3, a value much higher than the probability of being co-located when no interactions

Table 2 Overlapping among different connected pair sets. The subsets are disjoint to make their combinations additive

	$(E_c \cap E_{loc} \cap E_s) = I$	$(E_c \cap E_s) \setminus I$	$E_c \setminus (E_s \cup E_{loc})$	$E_s \setminus (E_c \cup E_{loc})$
$\|E\|$	110557	553621	162458	131665
%	3	16	4	3
	$E_c \cup E_s \cup E_{loc}$	$(E_s \cap E_{loc}) \setminus I$	$E_{loc} \setminus (E_c \cup E_s)$	$(E_c \cap E_{loc}) \setminus I$
$\|E\|$	3325308	8485	2331671	26851
%	100	0.2	70	0.8

occur [4]. Consequently, people communicating by phone are more likely to be in spatial proximity w.r.t. individuals who do not interact through any mobile channel.

Since the multigraph $\mathcal{D}_{csloc}$ takes into account both SMSs and calls, we can estimate, for each given pair of interacting users, whether the probability of being in proximity given a SMS communication is greater than the same probability given an interaction mediated by a call. We find that when interactions occur via SMS, this probability equals 0.28[5], and it is greater than the probability obtained when voice calls are used (0.16)[6], i.e., the interactions via SMS are more predictive of a possible spatio-temporal proximity w.r.t. voice calls.

To go further in the analysis, we introduce the weights associated to the different types of link with the purpose of verifying whether interacting pairs happen to be more likely in proximity than pairs that do not communicate but have a positive co-location rate. First, we investigate the distribution of the co-location rate *CoL* of the connected pairs which belong to $E_{loc} \cap (E_c \cup E_s)$ and compare it with the same distribution computed on the co-located pairs that do not communicate. The results, reported in Fig. 1, do not show significant differences in the distributions. However, the latter is slightly above the former, i.e., spatially close people that also use media to communicate are characterized by a proximity index higher than people that do not interact through mobile phone. We also investigate the probability that pairs, simultaneously sharing the same location, communicate by using a given medium (call or SMS separately). To this aim, in Fig. 2, we compare different subsets of the union of the connected pair sets: pairs connected by calls, by
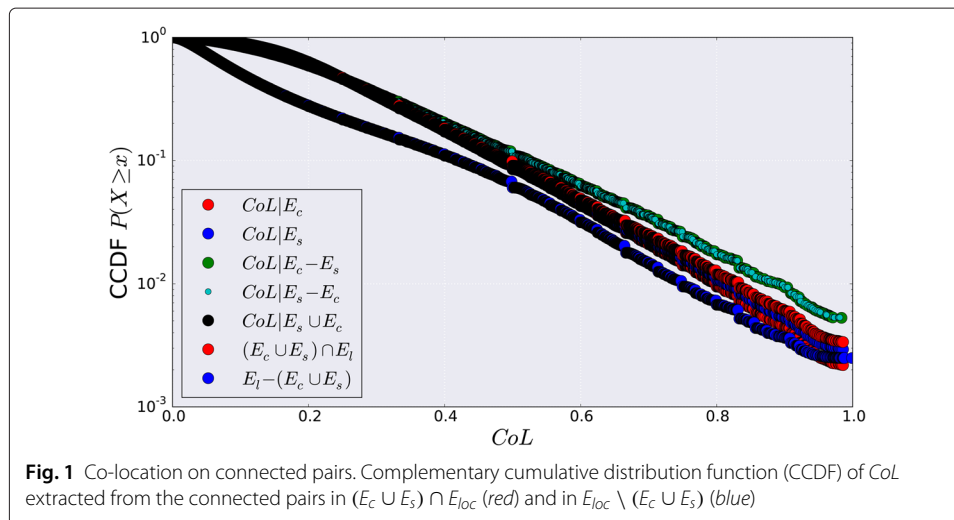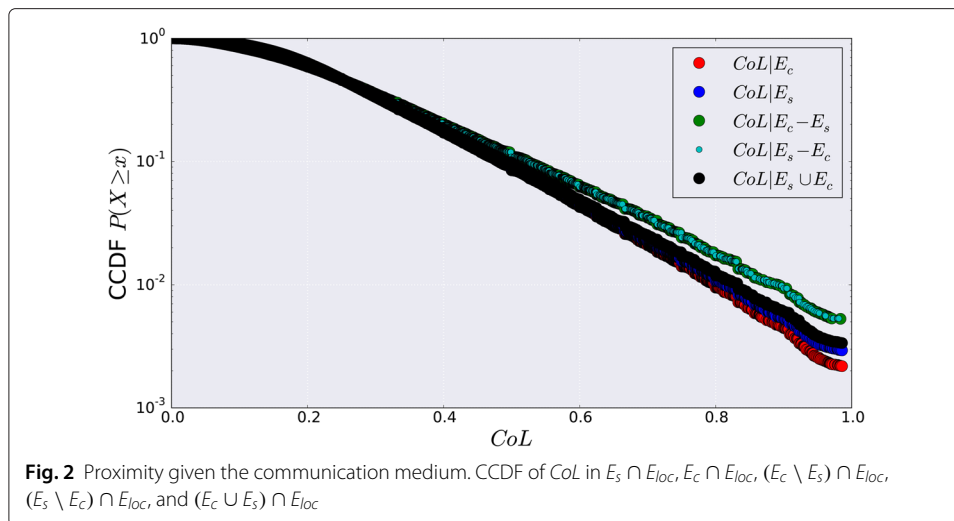

Fig. 1 Co-location on connected pairs. Complementary cumulative distribution function (CCDF) of *CoL* extracted from the connected pairs in $(E_c \cup E_s) \cap E_{loc}$ (*red*) and in $E_{loc} \setminus (E_c \cup E_s)$ (*blue*)

Zignani *et al. Computational Social Networks* (2015) 2:13

Page 9 of 24

Fig. 2 Proximity given the communication medium. CCDF of *CoL* in $E_s \cap E_{loc}$, $E_c \cap E_{loc}$, $(E_c \setminus E_s) \cap E_{loc}$, $(E_s \setminus E_c) \cap E_{loc}$, and $(E_c \cup E_s) \cap E_{loc}$
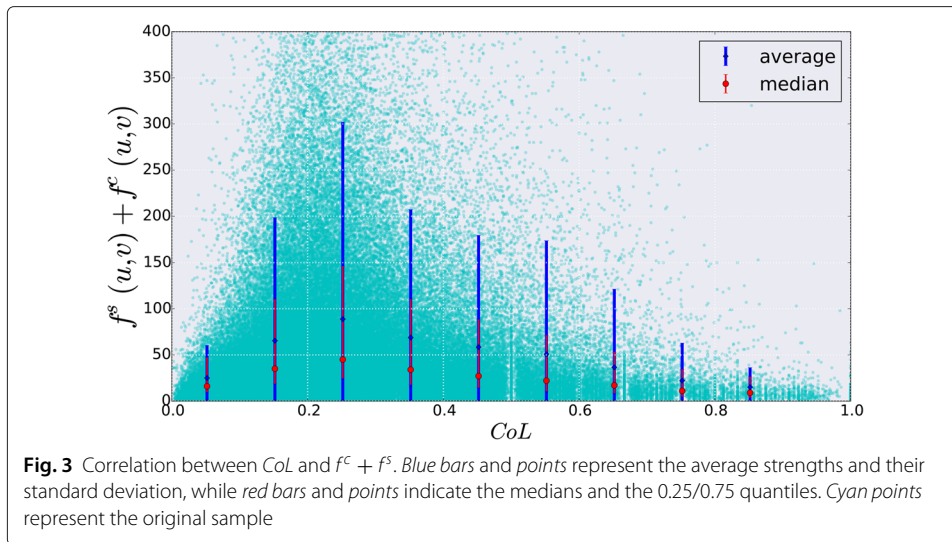
SMSs, by calls or SMSs exclusively, and by calls or SMSs ($E_c \cup E_s$). We note that the distributions are not different. That suggests proximity strength does not rely on the channel people are using to communicate and interact.
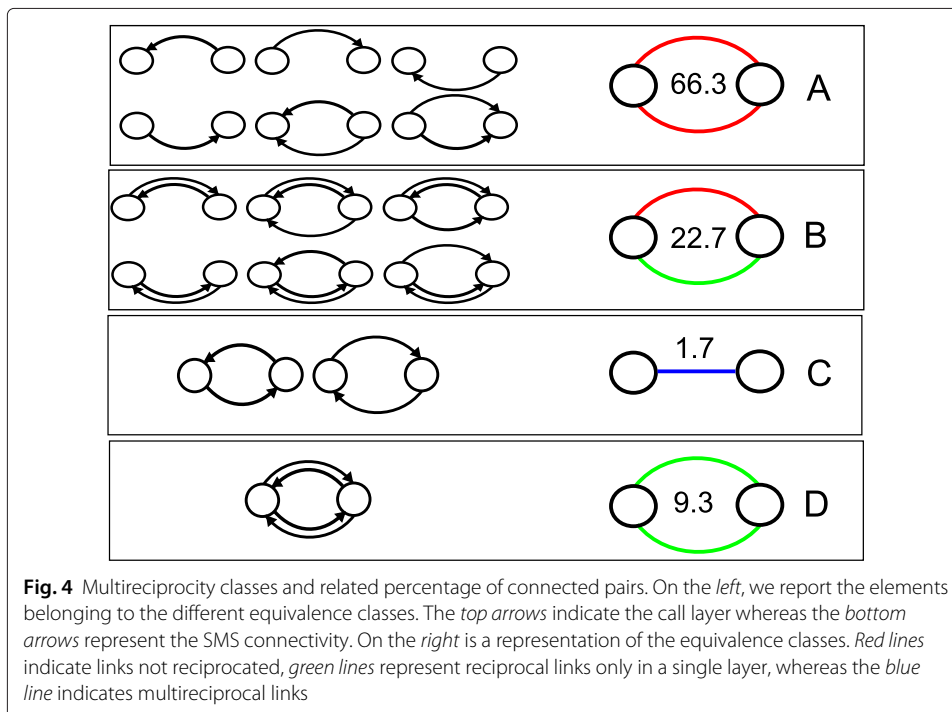
Some recent studies have found a positive correlation among spatio-temporal proximity, the presence of links in the related communication networks and their strengths; so people in proximity are more likely to be connected in the social network and to have intense direct interactions [10, 11]. We observe that both studies rely on mobile phone datasets covering wide geographical areas and consequently are characterized by a coarse spatial granularity. As a consequence, we wonder if the above results could be confirmed by our dataset or if the size of the region and the spatial granularity influence the outcomes. To replicate and compare the results, we apply the same methodology in [11], i.e., we study the correlation between *CoL* and the strength of the on-phone communication links. In Fig. 3, we show the average and median values of the link weights (frequency of the interactions) as a function of the co-location rate. The trend is completely different from the aforementioned results. We observe that from *CoL* = 0.2, the link strength decreases while *CoL* is increasing. That implies people who are very close in the spatio-temporal dimension do not frequently interact, while people who are sometime co-located communicate more frequently. This phenomenon could be explained in the light of the limited geographical area under study. In a city, people sharing simultaneously many places do not need a mobile phone to communicate since they exploit face-to-face interactions. Otherwise, people who are scarcely in proximity complement the face-to-face interactions with the mobile phone communications. We argue that, in a city area, mobile phones perform their original purpose, i.e., make the communication between distant people easier.

Reciprocity and multireciprocity

Any social interaction can be established through interleaved use of different communication channels and the one-dimensional notion of mutual relationship, i.e., the reciprocity or mutuality needs to be extended.

Fig. 3 Correlation between *CoL* and $f^c + f^s$. *Blue bars* and *points* represent the average strengths and their standard deviation, while *red bars* and *points* indicate the medians and the 0.25/0.75 quantiles. *Cyan points* represent the original sample

Among the two possible definitions of mutuality, we consider the dyad census adopting the approach presented in Wassermann and Faust [21]. As requested by the census approach, in Fig. 4, we enumerate all the ways a pair of nodes has to establish a relationship and then we gather them into equivalence classes which express similar behaviors. According to Fig. 4, class A includes links definitely not reciprocal, neither on single layers nor on the multiplex network. Class B refers to links which are reciprocal in one layer only. Class C refers to the true multidimensional reciprocity where links in the opposite directions belong to different layers. Note that classes B and C reckon with the importance of considering both the two communication media when the overall reciprocity on mobile phone has to be evaluated. Finally, class D contains very social edges as reciprocity

Fig. 4 Multireciprocity classes and related percentage of connected pairs. On the *left*, we report the elements belonging to the different equivalence classes. The *top arrows* indicate the call layer whereas the *bottom arrows* represent the SMS connectivity. On the *right* is a representation of the equivalence classes. *Red lines* indicate links not reciprocated, *green lines* represent reciprocal links only in a single layer, whereas the *blue line* indicates multireciprocal links

Zignani *et al. Computational Social Networks* (2015) 2:13

Page 11 of 24

exists in both layers. Note that the extension to more dimensions is easily achieved by defining appropriate equivalence classes. In Fig. 4, we report the percentage of connected pairs which belong to the different reciprocity classes. We observe a high number of non-reciprocated pairs accounting for more than 60 %. As for classes B, C, and D, the overall value of 0.34 is far from the reciprocity value that characterizes online social networks, rather it is more similar to values observed in the Web graph. The dyad census allows us to compute the reciprocity of each network layer, as class B contains reciprocated edges in a single layer. We find the reciprocity of the call graph to be 0.28, while the SMS graph has reciprocity 0.4. If we assume that reciprocity is a measure of the social importance of a tie, these values indicate that there is a greater fraction of social edges in SMS graph than in call graphs. Moreover, the resulting reciprocity is lower than in online social networks and in the Web, where it came out to be around 0.7.

We find that connected pairs are poorly reciprocal, mainly due to the low reciprocity of the call graph even if SMSs, as communication channel, are more suitable for bidirectional relationships. Moreover, only 9 % of pairs take advantage of the multidimensionality offered by the phone media to maintain relationships. As a matter of fact, the reciprocity typical of class C is negligible.

Degree distribution

In this section, we examine the macroscopic structure of the mobile phone network by considering the degree (k) and the strength (s) distributions. Degree and strength distributions give information about the level of interaction of a mobile user on the basis of the number of people contacting him/her, the number of people s/he contacts, and how often. As we are dealing with directed networks, besides the degree and strength distributions, we also analyze the in-degree (k^-), the out-degree (k^+), the in-strength (s^-), and the out-strength (s^+) distributions, where s^- and s^+ are obtained adopting as weight the frequency of the interaction $f \cdot (u, v)$. As for the multidigraph $\mathcal{D}$, we define the degree of a node as $k_{\mathcal{D}}(u) = |\Gamma_{\mathcal{D}}^+(u) \cup \Gamma_{\mathcal{D}}^-(u)|$, where $\Gamma_{\mathcal{D}}^+(u)$ and $\Gamma_{\mathcal{D}}^-(u)$ represent the out-going and in-going neighborhood on $\mathcal{D}$, respectively. An analogous definition holds for the strength distribution. These two variables account for the overall on-phone communication activity of a user whereas $k_{\mathcal{D}_{csloc}}$ sums up phone communications and potential interactions in the physical space.

Regarding the degree, mobile phone networks exhibit heavy-tail distributions as shown in Fig. 5. In particular, the in-degree distribution obeys to a power-law with the exponent $\alpha = 5.12$ as a result in applying the likelihood-based method presented in [22], while the out-degree and degree distributions exhibit a particular behavior due to presence of more nodes with high degree than expected in a heavy-tail distribution. The obtained results on degrees agree with those of socio-technological networks including online social network and the Web. Nonetheless, in online social networks, the distribution of outgoing links is similar to that of incoming links, while in the Web, the incoming links are significantly more concentrated on a few high-degree nodes than the outgoing links [23]. In our dataset, we observed a hybrid behavior where SMS degree distribution is more similar to online social networks, while the degree distribution in the call graph is more similar to the Web graph case. So, as suggested by the distribution in Fig. 5a, people use voice calls to maintain more relationships w.r.t. text messages. A further insight involve the degree
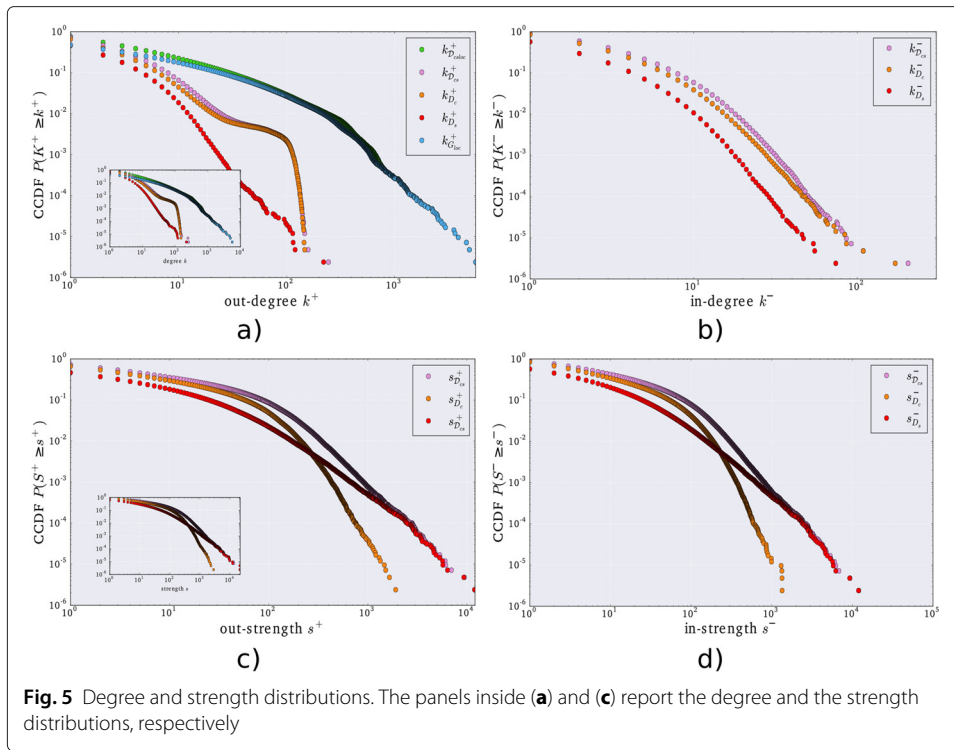
Zignani *et al. Computational Social Networks* (2015) 2:13

Page 12 of 24

Fig. 5 Degree and strength distributions. The panels inside (**a**) and (**c**) report the degree and the strength distributions, respectively

distributions of the multigraphs $\mathcal{D}_{cs}$ and $\mathcal{D}_{csloc}$. In Fig. 5a, b, we observe that the probability of connecting (or being connected) with more than k people in $\mathcal{D}_{cs}$ is ever greater than in the single layers. This way, the multiplex mobile network captures much more relationships than when considering only a single communication channel. For instance, the number of relationships captured by the multigraph is on average almost twice as much a single layer. This observation is further stressed if we take into account the degree distribution in $\mathcal{D}_{csloc}$. As reported in Fig. 5a, the introduction of the co-location network increases the number of neighbors of a node. Specifically, the co-location impacts on the tail of the distribution, i.e., the probability of having hundreds of neighbors is much more higher when we add the information about spatio-temporal proximity w.r.t. the on-phone interactions only.

Analogous observations hold for the strength distributions reported in Fig. 5c, d. Here, the values of the shape parameter are comparable within the same layer, so in- and out-strength distributions are similar. Also, strength values almost double when considering SMS and call, thus reinforcing that the multidimensional view on mobile phone datasets is mandatory to really understand the communication attitudes of phone users.

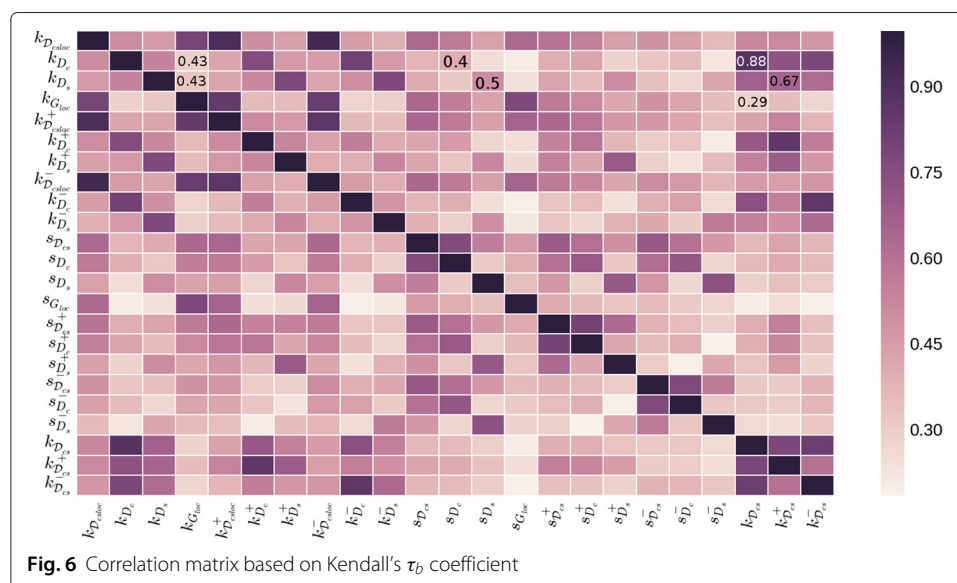**Degree and strength correlations across dimensions**

In the light of the above results, we ask whether a user maintains her/his level of activity across the diverse dimensions or conversely a user who is active through voice calls is co-located with few people or is not disposed to communicate through SMSs. That corresponds to verify if statistically significant correlations between the degrees and the strengths of the same group of users in different layers exist. We can evaluate the degree of correlation between pairwise layers by adopting different methods [24]. To get an overall picture of the pairwise degree and strength correlations, we apply a rank correlation

Zignani *et al. Computational Social Networks* (2015) 2:13

Page 13 of 24

analysis on the different pairwise quantities. As a rank correlation method, we compute the Kendall's rank correlation coefficient $\tau_b{}^7$ on the ranking induced by the degrees and the strengths. In Fig. 6, we visualize the rank correlation matrix, where each row (column) corresponds to a centrality measure computed on the relative network.

Before comparing the interplay among the different layers, we analyze the correlation between the degree and the strength in each single layer. In fact, by correlating the degree and the strength, we can say if users who communicate or is contacted by many people adopt heavily and frequently their mobile device. As for the call graph D_c, we observe in Fig. 6 a positive correlation between the degree k_{D_c} and the strength s_{D_c}. However, $\tau_b = 0.4$ indicates that it is not always true that a high number of called people corresponds to a proportional call activity, i.e., some individuals communicate with a lot of people but not so frequently. The same observation holds for the SMS graph, where $\tau_b(k_{D_s}, s_{D_s}) = 0.5$. Otherwise as regards the combined usage of the communication media ($\mathcal{D}_{cs}$), by comparing $\tau_b(k_{\mathcal{D}_{cs}}, k_{D_s})$ and $\tau_b(k_{\mathcal{D}_{cs}}, k_{D_c})$, we find that the addition of the calling activity to the texting layer could change the importance of the users in D_s, i.e., mobile phone users with a high rank in the SMS graph could drop off their ranking if one also adopt the call activity to measure the users' importance.

As for the correlation between the call/SMS layers and the co-location network, results in Fig. 6 confirm the loosely coupled interplay observed in the connected pair analysis. Here, we compare the degree distribution in G_{loc} that captures how many potential face-to-face interactions a user could have and the degree distributions in the call and SMS graphs, respectively. Also, in this case, the degree in G_{loc} is poorly correlated with the centrality in the call or in the SMS networks. That means mobile phone users who are spatially close to many individuals but communicate with few people exist. The same observation is still true whereas we consider the global propensity to communicate expressed by $\mathcal{D}_{cs}$. In this case, we measure $\tau_b(k_{\mathcal{D}_{cs}}, k_{G_{loc}}) = 0.29$.

In general, we find that the activity and the popularity is not straightforwardly maintained across the on-phone communications and the physical space; so, the different dimensions are poorly coupled in terms of degree and strength.

Fig. 6 Correlation matrix based on Kendall's τ_b coefficient

Zignani *et al. Computational Social Networks* (2015) 2:13

Page 14 of 24

Neighborhood overlapping

Regarding the comparison of the voice call and text message networks, the previous results show that there is not a perfect overlapping between people calling and messaging and highlight the importance of treating multidimensionality to better describe the social relationships. However, the overlapping of the node sets V_c and V_s is a general and sharp characteristic, which does not account for the different behavior of nodes. For instance, a user who mainly calls and sporadically texts belongs to $V_c \cap V_s$, even if she/he mainly is a caller. The same observation applies to the degree distribution on $\mathcal{D}_{cs}$, where we lose which, among the layers, mostly contributes to the expansion of the node's neighborhood.

To obtain further insights about nodes overlapping and neighborhood distribution across different dimensions, we now adopt a user-centric view by focusing on the evaluation of the in/out-neighborhood overlapping of a node u. To the purpose, we complement the Jaccard index of the neighbors set by defining the *exclusivity* ϕ of a layer d_i for a node u. This measure takes into account the different set of the out-going (in-going) neighbors on the two dimensions as follows:

$$\phi_{\{d_1,d_2\}}^{(.)}(u) = \frac{\Gamma_{d_1}^{(.)}(u) - \Gamma_{d_2}^{(.)}(u)}{\Gamma_{d_1}^{(.)}(u) \cup \Gamma_{d_2}^{(.)}(u)}, \tag{2}$$
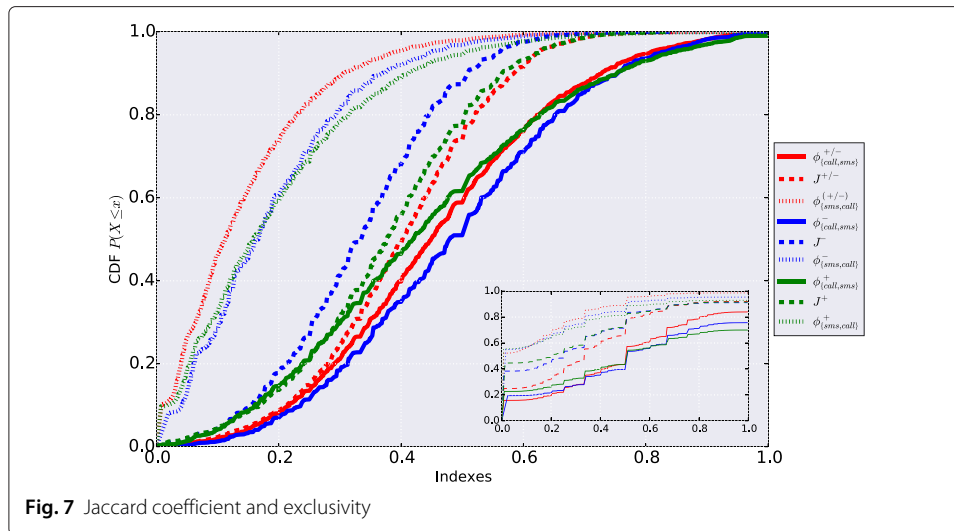
where $\Gamma_d^+(u) = \{v | (u, v, d) \in E\}$, $\Gamma_d^-(u) = \{v | (v, u, d) \in E\}$, and $\Gamma_d^{+/-} = \Gamma_d^+(u) \cup \Gamma_d^-(u)$ are, respectively, the out-going, the in-going, and the general neighborhood of the node u on the dimension d. For instance, a user who adopts only calls to maintain social relationships with most of her/his contacts will have a $\phi_{\{c,s\}}^{(+)}(u)$ close to 1. By adopting exclusivity, a mobile phone user is characterized by the call and text message exclusivity values and by the Jaccard index J that captures the intersection between call and SMS neighbors.

In Fig. 7, we report the distributions of the different indexes computed on nodes with degree (in-degree, out-degree) greater than 10, while the in-set figure shows the distributions on all nodes. Comparing the call exclusivity in both figures, we observe that many nodes with degree less than 10 exclusively adopt voice calls. In fact, the probability of observing an exclusive caller is about 0.3 in the in-set figure, while is close to 0 when nodes are filtered. Further, the Jaccard index distribution (dashed line) in Fig. 7 shows that users who communicate with more than 50 % of their neighbors are about 25 % of the population, while people are more likely to engage in relationships via calls than via text messages. In fact, the distributions of $\phi_{\{c,s\}}^{(+)}(u)$ and $\phi_{\{c,s\}}^{(-)}(u)$ show that about 20 % of users almost exclusively call their friends and contacts. Text messages, on the other hand, are less widespread as the main medium to exclusively relate with most of their contacts.

Connected components

In this section, we exploit reachability analysis to examine the macroscopic shape of our networks with the aim of evaluating their structural similarity and comparing them with other mobile phone networks or with other networks, such as, for instance, the Web graph.

We firstly proceed by extracting weakly connected components in order to detect group of users which are marginal w.r.t. other users. In Table 1 and in Fig. 8, we report the order, the size of the giant weakly connected component *GWCC*, and the distribution of the component size for each network. The large majority of the users belongs to the giant component, while the rest of nodes form small components whose size does not exceed 30

Zignani *et al. Computational Social Networks* (2015) 2:13

Page 15 of 24

Fig. 7 Jaccard coefficient and exclusivity

elements. The order of the $\mathcal{D}_{cs}$'s giant component w.r.t D_c's one highlights the role of SMS links. Through these links, the number of reachable nodes increases 9 % w.r.t nodes only reachable by calls. Moreover, SMS users (users who exclusively use SMS) aggregate other minor connected components of the call graph into the giant component. This suggests that in spreading and diffusing information, the mobile operator and third parties clients should take into account both communication channels to reach more phone users.

The link direction allows us to extract strongly connected components (SCC). Figure 8 shows the distribution of the SCCs size. Both strongly and the weakly connected components exhibit a heavy-tail distribution, and this behavior applies to both mono- and bi-dimensional networks. The considered networks are similar in terms of size distribution, and the obtained results are in line with those of the Web graph and of other online social networks. Shortest path and reachability analysis allows the identification of a giant strongly connected component and of the regions connected to it. This way, we can compare the macroscopic structure in terms of the Bow-tie model. The Bow-tie

Fig. 8 SCC size distributions

Zignani *et al. Computational Social Networks* (2015) 2:13

Page 16 of 24
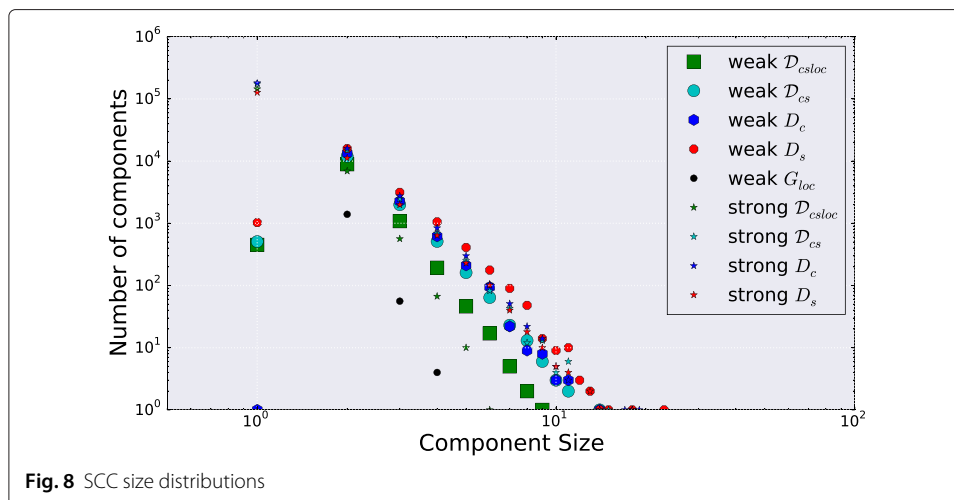
model, introduced by Broder et al. [25], is characterized by a giant SCC, an IN region containing components from which SCC can be reached, and an OUT region reachable from the SCC. We assign each node to the proper region and compute the relative sizes of the regions. Results, shown in Table 3, indicate that the three networks are structurally similar, as about 40 % of users belongs to the SCC, 6 % forms the IN region, and 20 % are in the OUT region. Moreover the reported percentages confirm the macroscopic structural difference of the telecom graphs w.r.t. the WWW graph observed in [5] in terms of the Bow-tie model.

Community

We have shown that the three considered layers are macroscopically similar, but nodes, links, and consequently ego-networks are very different when considering call or SMS on their own. We now wonder what happens at the intermediate level where people aggregate in communities. Do SMS and call communities match? Otherwise, are multidimensional communities [8] pivoted on SMS or call communities? What are the interplays between SMS, call, and multidimensional communities with the co-location communities? To answer these questions, we first detect communities onto the networks by means of three different detection algorithms. Then, we introduce a covering approach along with a few metrics to measure the degree of similarity between communities. Finally, we apply them both to compare SMS, call, and co-location communities and to deepen the structure of the multidimensional communities.

Community detection methods

The cluster analysis described in this section relies on the following community detection algorithms: the well known Louvain method (LM) [26] and label propagation (LP), a very powerful yet simple algorithm [27]. Besides, as LP might happen to move a label towards nodes far from the origin, thus leading to a very poor community structure, we also perform label propagation with hop attenuation (LPHA) [28] that overcomes this drawback.

Label propagation (LP) [27]: The algorithm requires that each node $v \in V$ determines its community by choosing the most frequent label shared by its neighbors. Initially, every node belongs to a different community. After some iterations, groups of nodes quickly reach a consensus on their label and they begin to contend those nodes that lay between groups. Here, we use the asynchronous update version, since it solves some problem of label fluctuations in particular graph structures. We choose LP due to its scalability since it has a linear time complexity and, unlike many others algorithms, it is parameter free.

Label propagation with hop attenuation (LPHA): LP is a very fast, scalable, and performing algorithm; however, in some contexts, a label can get to nodes of the network very far from the original one, thus leading to a very poor community structure. To curb this problem, Leung et al. [28] proposed some improvements and adjustments to the original

Table 3 Portion of nodes in the different elements of the bow-tie model

Type	IN	SCC	OUT	Tendrils	Tubes
$\mathcal{D}$	6.1	48.1	22.0	13.5	0.6
D_{call}	7.0	42.4	21.9	16.3	0.7
D_{sms}	5.9	41.2	26.5	5.8	0.1

Zignani *et al. Computational Social Networks* (2015) 2:13

Page 17 of 24

algorithm. We introduce in our LP implementation the hop attenuation improvement. The idea is to assign a *time to live* score to labels in order to prevent them from reaching too far nodes.

Every node updates its label choosing the one with maximum score shared by its neighbors. When the label is chosen, the hop score is updated subtracting a $\delta \in (0, 1)$:

$$s_i'(\mathcal{L}_u) = \left(\max_{u' \in \Gamma(u)(\mathcal{L}_u')} s_i(\mathcal{L}_{u'}) \right) - \delta$$

where $\mathcal{L}_u$ is the node u label, $s_u(\mathcal{L}) \in [0, 1]$ is the score of label $\mathcal{L}$, and $\Gamma(u)$ is the set of neighbors for node u.

Louvain: The Louvain method (LM) was introduced by Blondel et al. [26] and is one of the most popular greedy algorithms for *modularity* optimization. The method is very fast and can produce very high-quality communities. The algorithm is iterative. Each iteration consists of two steps. In the first one, every node is initially set to a new community. Then, for every node i and its neighbors j, the algorithm calculates the gain in modularity moving i from its community to j's community. The node i is then moved to the community with maximum gain. If there is no positive gain, it remains in its community. The first phase finishes when there are no more movements. The second step is to group together all the nodes in the same community in a macro node. A new graph is built, in which macro nodes are linked by an edge if there is an edge between two nodes belonging to the two different macro nodes. After the new graph is built, a new iteration starts and a new phase one is initiated.

We run the three community detection algorithms on the weighted undirected networks G_c, G_s, G_{loc}, and $\mathcal{G}_{cs}$. To estimate the weight of the undirected link, we sum the strengths of the directed links connecting the pair. Figure 9 shows the size distribution of the communities detected by LM, LP, and LPHA. Louvain and LP detect a giant community, while LPHA splits it into smaller ones. The community size distributions of LP and LPHA are power-law, as found in other networks. Label propagation algorithms find small-medium size communities, while Louvain favors small and medium-large communities. The above results motivate the choice of three different community detection
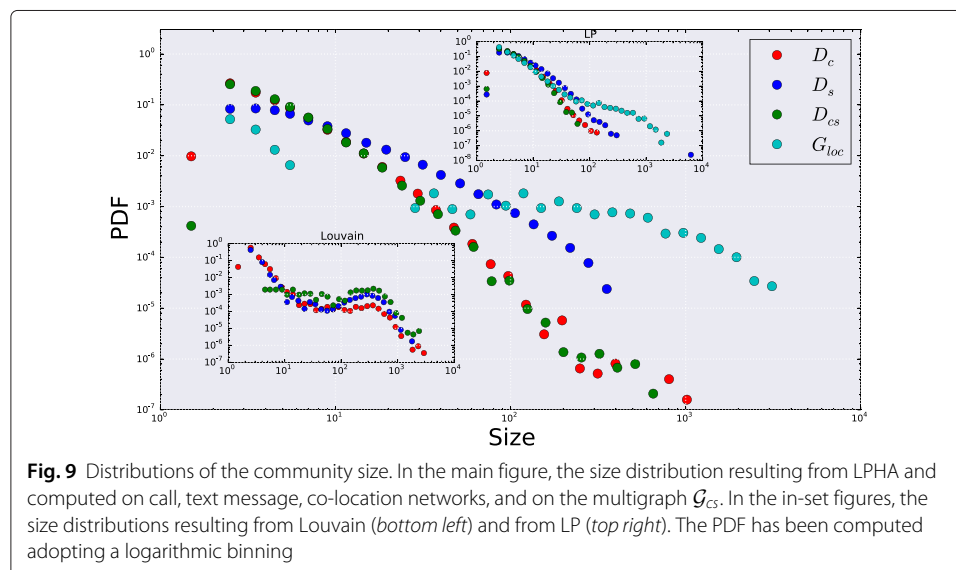

Fig. 9 Distributions of the community size. In the main figure, the size distribution resulting from LPHA and computed on call, text message, co-location networks, and on the multigraph $\mathcal{G}_{cs}$. In the in-set figures, the size distributions resulting from Louvain (*bottom left*) and from LP (*top right*). The PDF has been computed adopting a logarithmic binning

Zignani *et al. Computational Social Networks* (2015) 2:13

Page 18 of 24

algorithms. An outcome that holds for three different partitions is more independent from the chosen community detection method.
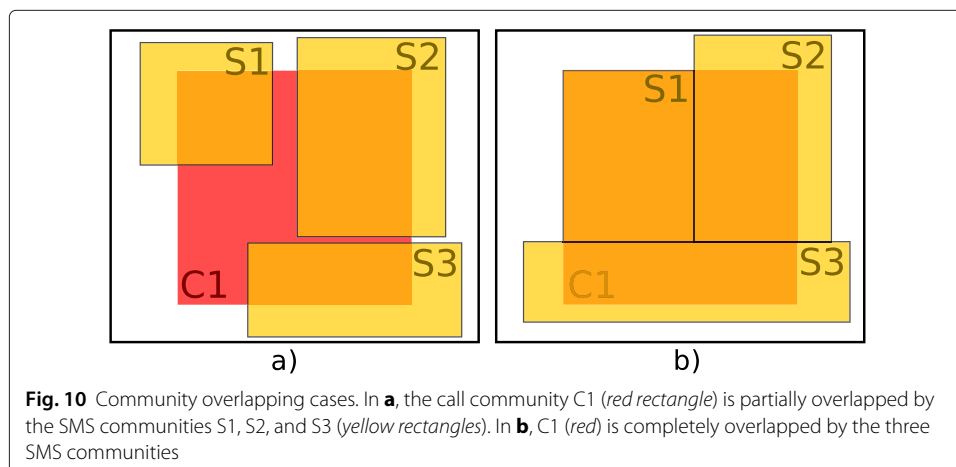
Communities match method

Although people aggregate in communities in both call and SMS layers, not necessary those communities coincide across the layers. As a consequence, we aim to investigate the degree of matching of the communities detected on the four networks. The evaluation of the overlapping between communities has some common aspects with standard point measures of cluster similarity [29]. The latter relies on finding the best matching between communities, while we are interested in evaluating the degree of their overlapping, and they assume the same node set, while we need to evaluate partitions on different sets as described in the "Networks characterization" section.

This way, we adopt a covering-like approach similar to the one used in [30]. Given a partition P_i on V_i and a community $c_i \in P_i$, we compute how it is covered by the communities of another partition P_j on V_j. We consider the maximal covering, i.e., the set of communities performing the maximum coverage of c_i:

$$MT(c_i, P_j) = argmax_{X \subseteq P_j, \forall c_j \in X, c_i \cap c_j \neq \emptyset} \sum_{c_j \in X} |c_i \cap c_j| \qquad (3)$$

In Fig. 10, we report the meaningful covering cases. The community c_i can be partially tiled or covered by one or more communities c_js of the other partition P_j, as in Fig. 10a. Otherwise, c_i can be totally covered by one or more communities as in Fig. 10b. As for total covering, we separately analyze some specific cases: (a) the perfect matching indicating communities that coincide, (b) communities which are properly subsets, and (c) communities that are covered by multiple communities of the other layer.

The partial covering by many sets includes different situations where the degree of the overlapping and the importance of the covering vary. For instance, many communities could marginally overlap to a single community, resulting in a low overlapping level or the region outside the covered community (yellow area in Fig. 10b) changes, indicating a more or less strong similarity between c_i and MT. We use three metrics to evaluate the degree and the goodness of the covering, besides the population percentages of the classes described above:

Fig. 10 Community overlapping cases. In **a**, the call community C1 (*red rectangle*) is partially overlapped by the SMS communities S1, S2, and S3 (*yellow rectangles*). In **b**, C1 (*red*) is completely overlapped by the three SMS communities

Zignani *et al. Computational Social Networks* (2015) 2:13

Page 19 of 24

- The cardinality of $MT(c_i, P_j)$, denoted as ν. Communities with a high ν act as bridges across the communities in the other layer.
- The coverage ratio CR as the fraction of nodes of c_i covered by MT, i.e., $\sum_{X_j \in MT} |X_j \cap c_i| / |c_i|$. The closer to 1 the coverage ratio, the greater the accuracy of the covering. This quantity is meaningful for the partial overlapping case.
- The coverage precision CP defines as $\sum_{X_j \in MT} |X_j \cap c_i| / |\bigcup_{X_i \in MT} X_i|$. It accounts for that portion of nodes which are outside the covered set. Values of CP close to 1 indicate that MT almost coincides with c_i in the total covering case or is almost a proper subset of c_i in the partial overlapping.

Communities across call, SMS, and co-location

First, we examine the communities overlapping focusing and comparing the call and the SMS layers, separately. Although, at macroscopic level, the size distributions of the communities detected by the three algorithms exhibit some differences, results about community matching are very similar. Thus, we will show results on the partitions [9] of the call layer, P_{call}, the SMS layer, P_{sms}, the co-location layer, P_{loc}, and the multidimensional graph, P, induced by the communities detected by LPHA, only.

We ask whether people form similar groups despite the communication channel they adopt or vice versa different media capture different groups. Furthermore, we wonder which communication layer is more representative of the groups given by the spatial proximity. The first difference emerges just examining the probability distribution function of the community size in Fig. 9. SMS communities are generally smaller than call communities, independently of the detection algorithm. These outcomes support the idea that SMS is the communication channel usually adopted by groups to build and maintain close social relationships or by socially anxious people [1].

More insightful results have been obtained by analyzing how the communities of the two layers overlap. In the comparison, we only consider the communities with size greater or equal to 10. As for communities equality, they are never identical and only a few are proper subset. Whereas the lack of identical communities could be expected due to the diversity of the node sets, the scarcity of proper subsets suggests that communities express different groups in the two layers. The almost inclusion of the SMS node set into the call set rebounds on the number of totally covered SMS communities. In fact, we find that most of the SMS communities are totally covered by 7.5 call communities on average or they are partially covered with a high coverage ratio, but the coverage precision is very low (0.8 quantile equals to 0.05). This indicates that neither the union of some call communities can precisely express a single SMS community. A similar behavior concerns the covering of the call communities with the SMS ones. In this case, SMS communities are a little more precise in the covering, but they only partially overlap with the call communities.
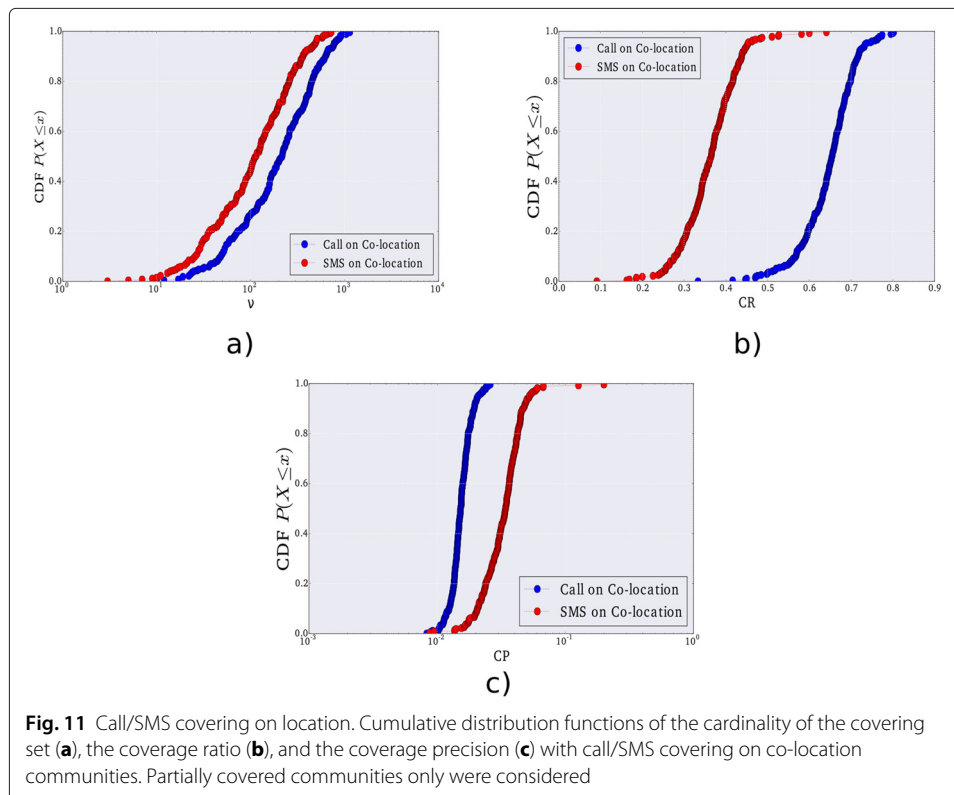
The analysis of the call/SMS overlapping highlights that these communication channels are used by different groups to establish interactions and relationships. Communities dictated by voice call are totally different from SMS communities, nor a community in a layer captures a specific subset in a community on the other layer.

We extend to the co-location network our analysis on the overlapping between the communication media. In particular, we focus on the pairwise interplay between the co-location graph and the interactions on the call and SMS networks separately, leaving the

discussion about the multidimensional case to the next section. As for the SMS communities, we observe that a single co-location community is covered by many SMS groups. This phenomenon is even more pronounced if we observe call communities as shown in Fig. 11a. These high cardinality values could not necessarily imply that the call and SMS networks are completely different from the co-location at a mesoscopic scale. Indeed, we could have big location communities containing many small SMS or call groups. That is not our case, as reported in Fig. 11b, c. In fact, despite the high coverage ratio of calls, the coverage precision is very low for both layers. That means call and SMS communities are entirely different from the co-location communities. These results are quite expected, especially in light of the lack of an increasing relation between *CoL* and the strength of the communications.
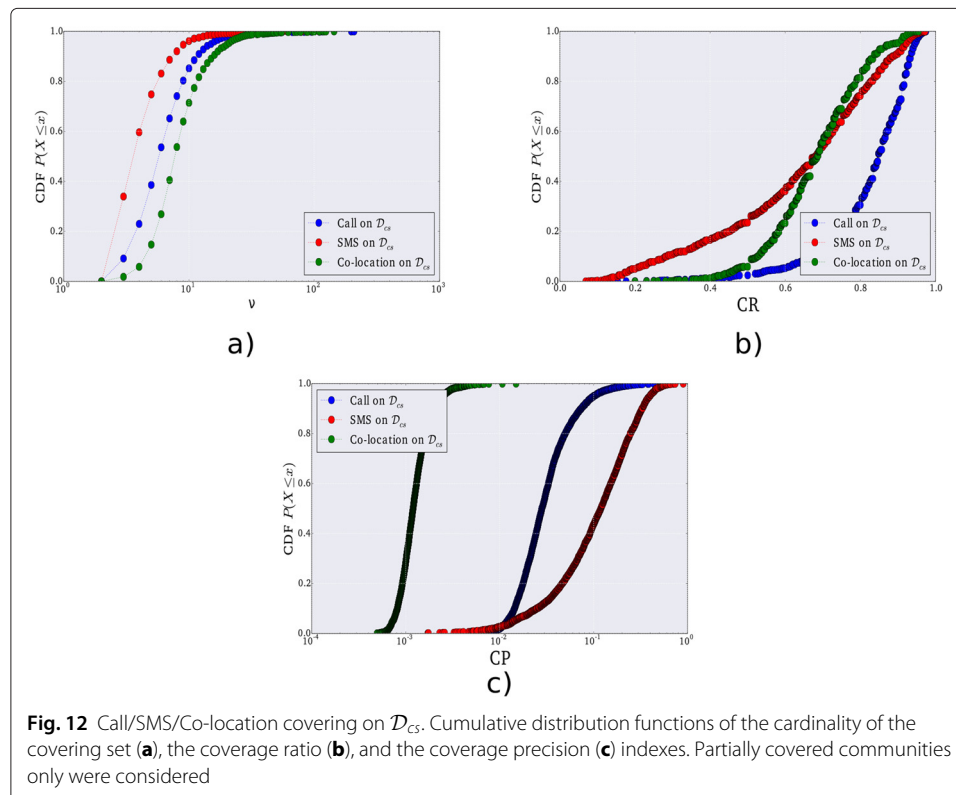
Multidimensional communities

As the two layers of the networks have completely different structures with respect to communities, we ask ourselves the following question: Is the multidimensional network related to the single layers? And is it driven by SMS or call communities? The community detection in multidimensional networks is still an open issue [31]. Here, we follow the approach presented in [32]. We proceed to extract multidimensional communities by introducing a mapping which transforms $\mathcal{G}_{cs}$ into a monodimensional weighted network W. Precisely, we assign as weight of an undirected link in W the total number of interactions (SMSs and calls) between the connected users.

Fig. 11 Call/SMS covering on location. Cumulative distribution functions of the cardinality of the covering set (**a**), the coverage ratio (**b**), and the coverage precision (**c**) with call/SMS covering on co-location communities. Partially covered communities only were considered
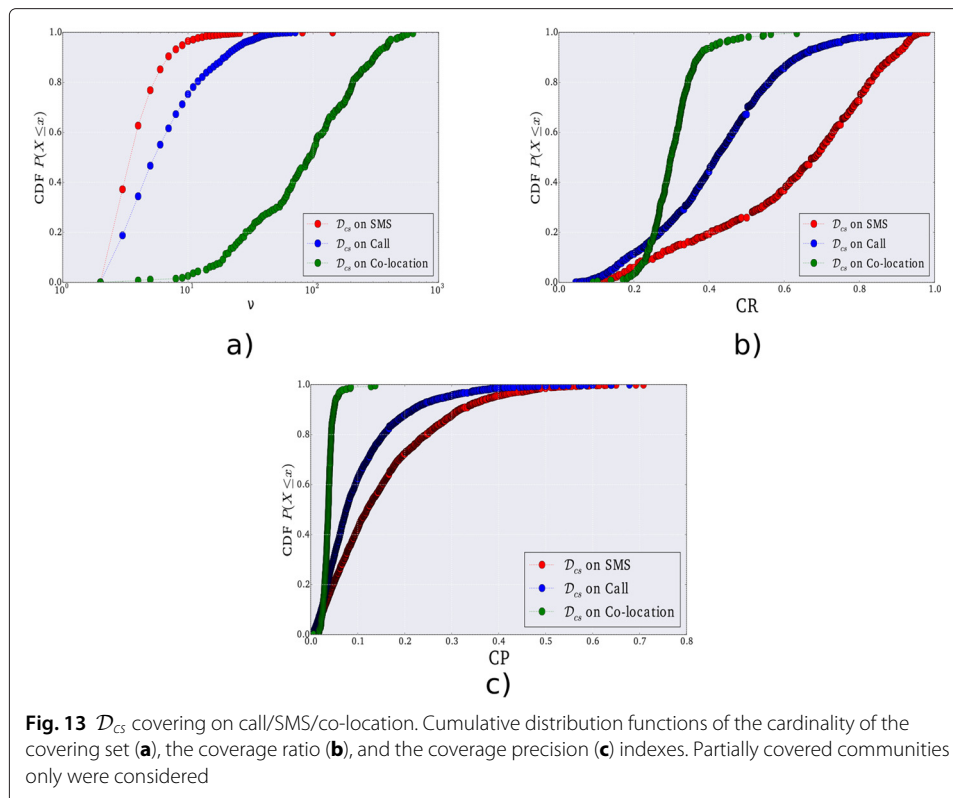
First, we look at how call and SMS communities cover multidimensional communities. Specifically, we evaluate how multidimensional groups aggregate or split layer communities, and we implicitly highlight the value of including different dimensions.

As highlighted by Fig. 12a, call communities perform a higher coverage of a multidimensional community than SMS, but many of them are needed as given by the covering set cardinality in Fig. 12b and with a smaller coverage precision (see Fig. 12c). Note that, as the call graph is greater than the SMS one, a great number of call links exist also in the multidimensional communities. This way, the multidimensional communities can be covered by using a large number of call communities, but the goodness of the covering is less than that performed by SMS communities, the latter being more precise and with a better matching. These results suggest that call communities are split into different multidimensional communities as a consequence of the introduction of the SMS weights. The same holds for SMS communities but the phenomenon is less evident.

In order to get an insight into the mechanism underlying the structure of multidimensional communities and understanding whether they are driven by SMS or call communities, we perform the opposite covering procedure, i.e., we tile the call and SMS communities with the multidimensional ones. The analysis of the covering explicitly shows how the communities are broken and split by the multidimensional partitioning. As for SMS, the distributions of v, CR, and CP reported in Fig. 13a–c remark that SMS communities are split into fewer multidimensional communities than call ones. Moreover, the partial covering is more coating and precise in the SMS layer than in the call layer. So, call communities are less similar and influent to the multidimensional one w.r.t. SMS communities.

Fig. 12 Call/SMS/Co-location covering on $\mathcal{D}_{cs}$. Cumulative distribution functions of the cardinality of the covering set (**a**), the coverage ratio (**b**), and the coverage precision (**c**) indexes. Partially covered communities only were considered

Zignani *et al. Computational Social Networks* (2015) 2:13

Page 22 of 24

Fig. 13 $\mathcal{D}_{cs}$ covering on call/SMS/co-location. Cumulative distribution functions of the cardinality of the covering set (**a**), the coverage ratio (**b**), and the coverage precision (**c**) indexes. Partially covered communities only were considered

All these results lead to conclude that multidimensional communities are pivoted by SMS communities which are merged in greater multidimensional communities by means of call links, whereas SMS links bring portions of call communities into different multidimensional communities.

For further information in Fig. 13, we report the overlapping results between the multidimensional and the co-location communities. However, the same observations in the previous analysis are still true and mainly highlight the weak interplay between co-location and on-phone interaction communities.

Conclusion

This paper makes a contribution in showing that the study of single social network offers a very partial description of user's interactions and that only multiplex studies can offer a vision closer to reality. With this aim, we have studied the multiplex social network build on call and SMS communications and we added a further dimension given by the spatio-temporal proximity. We have found that, despite the macroscopic similarities of the two networks, the on-phone communication behaviors are very different at the microscopic and the mesoscopic scales. In particular, the two single networks partially overlap, since many users adopt exclusively a communication channel (call or SMS). This diversity results in an enlargement of the users' ego-networks in the multiplex network. As a consequence, the studies of the interactions expressed only through calls are incomplete and, according to [8], biased by the generational difference in the use of a particular media. For this reason, modeling mobile phone data by a network of communication networks could

Zignani *et al. Computational Social Networks* (2015) 2:13

Page 23 of 24

represent an advance in the computational social science to offer a more complete vision of user's interactions.

Interactions and communications have many connections and relations with how people move and visit common locations. In fact, we confirm that people interacting by mobile devices are more likely to be in spatial proximity w.r.t. individuals who do not interact through any mobile medium. In particular, interactions by SMSs are more predictive of spatial proximity than calls. This result could be applied to solutions for the next-location prediction problem which include features from the communication layer.

The joint study of mobility and communication networks has also highlighted that their interplay depends on the spatial granularity of the mobility traces and on the covered area. In a metropolitan area, people sharing simultaneously many places do not need to communicate since they exploit the co-presence. Otherwise, people who are scarcely in proximity complement the face-to-face interactions with the mobile phone communications. The last finding is in contrast with the results in [10] and [11] so we are going to analyze the same dependency by coarsening the spatial granularity.

Finally, we highlighted how the diversity of the layers impacts on the mesoscopic structure given by the communities which form and emerge due to the call and SMS interactions and the spatial closeness of the users. Specifically, we observe that communities at different layers either do not match or they barely preserve from a layer to another. The above results further highlight that at a metropolitan scale, the interplay between co-location and communication is not well defined. However, the chance to associate a spatio-temporal information to SMS and call communities may offer many applications. Besides locating in the urban space to the communication communities when possible, it may be applied to improve home-work locations according to the communication patterns or to filter out "non-social" interactions.

Endnotes

[1] For the purpose of ensuring customer anonymity, each subscriber is identified by a surrogate key.

[2] In our metropolitan environment, the average radius of a cellular cell is about 45 m.

[3] $|(E_c \cup E_s) \cap E_{loc}|/|E_c \cup E_s|$

[4] $2|E_{loc} \setminus (E_c \cup E_s)|/(|V_c \cup V_s|(|V_c \cup V_s| - 1) - 2|E_c \cup E_s|) = 2 \times 10^{-5}$

[5] $|E_s \cap E_{loc}|/|E_s|$

[6] $|E_c \cap E_{loc}|/|E_c|$

[7] τ_b takes into account ties.

[8] By multidimensional communities, we mean the communities computed onto $\mathcal{D}_{cs}$

[9] The above algorithms extract not-overlapped communities.

Zignani *et al. Computational Social Networks*   (2015) 2:13

Page 24 of 24

References

1. Reid, D, Reid, FM: Mobile World. In: Hamill, L, Lasen, A, Diaper, D (eds.) Computer Supported Cooperative Work, pp. 105–118. Springer, London, (2005)
2. Blondel, VD, Decuyper, A, Krings, G: A survey of results on mobile phone datasets analysis (2015). arXiv preprint arXiv:1502.03406
3. Onnela, J-P, Saramäki, J, Hyvönen, J, Szabó, G, Lazer, D, Kaski, K, Kertész, J, Barabási, A-L: Structure and tie strengths in mobile communication networks. Proc. Nat. Acad. Sci. **104**(18), 7332–7336 (2007)
4. Hidalgo, CA, Rodriguez-Sickert, C: The dynamics of a mobile phone network. Physica A: Stat. Mech. Appl. **387**(12), 3017–3024 (2008)
5. Nanavati, AA, Singh, R, Chakraborty, D, Dasgupta, K, Mukherjea, S, Das, G, Gurumurthy, S, Joshi, A: Analyzing the structure and evolution of massive telecom graphs. Knowl. Data Eng. IEEE Trans. **20**(5), 703–718 (2008)
6. Karsai, M, Kaski, K, Barabási, A-L, Kertész, J: Universal features of correlated bursty behaviour. Scientific Reports. **2**, 397 (2012)
7. Quadri, C, Zignani, M, Capra, L, Gaito, S, Rossi, GP: Multidimensional human dynamics in mobile phone communications. PloS One. **9**(7), 103183 (2014)
8. Ling, R, Bertel, TF, Sundsøy, PR: The socio-demographics of texting: an analysis of traffic data. New Media Soc. **14**(2), 281–298 (2012)
9. Phithakkitnukoon, S, Smoreda, Z, Olivier, P: Socio-geography of human mobility: a study using longitudinal mobile phone data. PloS One. **7**(6), 39253 (2012)
10. Calabrese, F, Smoreda, Z, Blondel, VD, Ratti, C: Interplay between telecommunications and face-to-face interactions: a study using mobile phone data. PloS One. **6**(7), 20814 (2011)
11. Wang, D, Pedreschi, D, Song, C, Giannotti, F, Barabasi, A-L: Human mobility, social ties, and link prediction. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining. KDD '11. New York, NY, (2011)
12. Caughlin, TT, Ruktanonchai, N, Acevedo, MA, Lopiano, KK, Prosper, O, Eagle, N, Tatem, AJ: Place-based attributes predict community membership in a mobile phone communication network. PloS One. **8**(2), 56057 (2013)
13. Expert, P, Evans, TS, Blondel, VD, Lambiotte, R: Uncovering space-independent communities in spatial networks. Proc. Nat. Acad. Sci. **108**(19), 7663–7668 (2011)
14. Onnela, J-P, Arbesman, S, González, MC, Barabási, A-L, Christakis, NA: Geographic constraints on social network groups. PLoS ONE. **6**(4), e16939 (2011)
15. D'Agostino, G, Scala, A: Networks of Networks: The last frontier of complexity. Understanding complex systems. Springer (2014). doi:10.1007/978-3-319-03518-5
16. Berlingerio, M, Coscia, M, Giannotti, F, Monreale, A, Pedreschi, D: Multidimensional networks: foundations of structural analysis. World Wide Web. **16**, 1–27 (2012)
17. De Domenico, M, Solé-Ribalta, A, Cozzo, E, Kivelä, M, Moreno, Y, Porter, MA, Gómez, S, Arenas, A: Mathematical formulation of multilayer networks. Phys. Rev. X. **3**, 041022 (2013)
18. Lambiotte, R, Blondel, VD, de Kerchove, C, Huens, E, Prieur, C, Smoreda, Z, Van Dooren, P: Geographical dispersal of mobile communication networks. Physica A: Stat. Mech. Appl. **387**(21), 5317–5325 (2008)
19. Zignani, M, Quadri, C, Bernadinello, S, Gaito, S, Rossi, GP: Calling and texting: social interactions in a multidimensional telecom graph. In: Proceedings of the complex networks 2014 workshop on complex networks and their applications. Complex Networks '14. IEEE, Marrakech, (2014)
20. Gonzalez, MC, Hidalgo, CA, Barabasi, A-L: Understanding individual human mobility patterns. Nature. **453**, 779-782 (2008)
21. Wasserman, S, Faust, K: Social network analysis: methods and applications. Cambridge University Press (1994)
22. Clauset, A, Shalizi, CR, Newman, MEJ: Power-law distributions in empirical data. SIAM Review. **51**(4) (2009). arxiv.org/pdf/0706.1062
23. Mislove, A, Marcon, M, Gummadi, KP, Druschel, P, Bhattacharjee, B: Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, pp. 29–42. ACM, (2007). conferences.sigcomm.org/imc/2007/papers/imc170.pdf
24. Boccaletti, S, Bianconi, G, Criado, R, del Genio, CI, Gómez-Gardeñes, J, Romance, M, Sendiña-Nadal, I, Wang, Z, Zanin, M: The structure and dynamics of multilayer networks. Phys. Rep. **544**(1), 1–122 (2014)
25. Broder, A, Kumar, R, Maghoul, F, Raghavan, P, Rajagopalan, S, Stata, R, Tomkins, A, Wiener, J: Graph structure in the web. Comput. Netw. **33**(1), 309–320 (2000)
26. Blondel, VD, Guillaume, J-L, Lambiotte, R, Lefebvre, E: Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. **10**, p10008 (2008)
27. Raghavan, UN, Albert, R, Kumara, S: Near linear time algorithm to detect community structures in large-scale networks. Phys. Rev. E. **76**(3), 036106 (2007)
28. Leung, I. XY, Hui, P, Liò, P, Crowcroft, J: Towards real-time community detection in large networks. Phys. Rev. E. **79**, 066107 (2009)
29. Meilă, M: Comparing clusterings—an information-based distance. J. Multivariate Anal. **98**(5), 873–895 (2007)
30. Tibély, G, Kovanen, L, Karsai, M, Kaski, K, Kertész, J, Saramäki, J: Communities and beyond: mesoscopic analysis of a large social network with complementary methods. Phys. Rev. E. **83**(5), 056125 (2011)
31. Mucha, PJ, Richardson, T, Macon, K, Porter, MA, Onnela, J-P: Community structure in time-dependent, multiscale, and multiplex networks. Science. **328**(5980), 876–878 (2010)
32. Berlingerio, M, Coscia, M, Giannotti, F: Finding redundant and complementary communities in multidimensional networks. In: Proceedings of the 20th ACM international conference on information and knowledge management. CIKM '11. ACM, NY, USA, (2011)