Zignani et al. *EPJ Data Science* (2019) 8:8
https://doi.org/10.1140/epjds/s13688-019-0187-7

EPJ Data Science
a SpringerOpen Journal

EPJ.org

REGULAR ARTICLE Open Access

Urban groups: behavior and dynamics of social groups in urban space

Matteo Zignani[1*], Christian Quadri[1], Sabrina Gaito[1] and Gian Paolo Rossi[1]

*Correspondence:
matteo.zignani@unimi.it
[1]Department of Computer Science, University of Milan, Milan, Italy

Abstract

The tendency of people to form socially cohesive groups that get together in urban spaces is a fundamental process that drives the formation of the social structure of cities. However, the challenge of collecting and mining large-scale data able to unveil both the social and the mobility patterns of people has left many questions about urban social groups largely unresolved. We leverage an anonymized mobile phone dataset, based on Call Detail Records (CDRs), which integrates the usual voice call data with text message and Internet activity information of one million mobile subscribers in the metropolitan area of Milan to investigate how the members of social groups interact and meet onto the urban space. We unveil the nature of these groups through an extensive analysis, along with proposing a methodology for their identification. The findings of this study concern the social group behavior, their structure (size and membership) and their root in the territory (locations and visit patterns). Specifically, the footprint of urban groups is made up by a few visited locations only; which are regularly visited by the groups. Moreover, the analysis of the interaction patterns shows that urban groups need to combine frequent on-phone interactions with gatherings in such locations. Finally, we investigate how their preferences impact the city of Milan telling us which areas encourage group get-togethers best.

Keywords: Mobile phone graph; Mobile social groups; Quasi-clique; Group points of interest; City's points of interest

1 Introduction

The understanding of tight-knit social groups represents a key factor in the development of services which integrate contextual information from social and mobility data sources [1]. Besides being a fundamental concept driving many sociological studies, the idea of social groups is central in modern social networking services and instant-messaging applications, e.g. WhatsApp, Snapchat and Skype. This is due to people's increasing propensity to share images and videos with a restricted group of close friends built around specific interests [2]. The central role of social groups is further emphasized when these groups move around and/or are easily mappable onto locations in a city [3–5]. It is the basis of a rich offering of targeted applications and services, e.g. content dissemination of location-aware information useful to a group [6, 7] or recommendation of locations fitting a group's interests [8, 9]. So, an understanding of the typical traits of social groups in an urban context

is mandatory for purposes of developing more personalized location-based social applications and solutions.

In this paper we focus on *urban groups*, i.e. cohesive social groups that express their interactions in urban places. Through an extensive analysis, we unveil the nature of these groups and propose a methodology for their identification. Our analysis rests on an anonymized mobile phone dataset based on Call Detail Records (CDRs) over a span of 67 days that integrate the usual voice call data with text message and Internet activity information of one million mobile subscribers in the metropolitan area of Milan. This wealth of data provides us with a unique opportunity to study how social groups interact and meet in an urban space having a large population. In fact, in addition to the reconstruction of more complete social interactions merging call and text communications, the provided mobility information allows us to obtain more detailed user mobility traces. This is a key component for the identification of the co-location of group members.

The contributions of our work can be summarized as follows:

- We propose a procedure for the identification of urban groups. The identification approach is applicable to every context providing a graph that expresses both the interactions/communications among users and the users' mobility traces. Due to its high modularity, the methodology can be employed to discover whatever subgraph expresses the concept of group, as well as to map, when feasible, the group's activities in urban places. In this latter respect, it finds its favorite locations.
- By applying the above procedure, we analyze how urban groups meet and behave within the urban space. We show that these groups meet all the main criteria of what makes for a sociological group, namely: *mutuality* (i.e. groups are highly dense subgraphs where each one interacts with any other); *reachability* (i.e. within a group no one is disconnected); *interactivity* (i.e. urban group members interact with one another frequently, and in large groups they devote much greater efforts to interacting with one another and to maintaining relationships established within the group than they do in small groups).
- We provide a characterization of the urban groups by analyzing their size and membership, and we find similarities with modern instant messaging services (e.g. WhatsApp and WeChat). In addition , we also focus on the preferences of urban groups by investigating the places where they meet and the frequency with which they gather. Specifically, we show that, in strict analogy to human mobility, urban groups are characterized by few visited locations; also they need to combine on-phone interactions with gatherings in such locations, since the visitation patterns of these locations is regular. Finally, we investigate how their preferences impact the city of Milan. This tells us which areas encourage group get-togethers best.
- We also highlight how mobility and interaction information define social roles within urban groups [10]. Specifically, we focus on the identification of leader/follower relations through the visit patterns of the places hosting urban group gatherings. We find a subset of members (the leaders) who take part frequently in the get-togethers, while other members (the followers) play a much more marginal role w.r.t. the urban group activities. The same observation also holds for the frequency of the interactions within a group. In this case, within the largest groups, we identify the presence of a backbone of strong links involving a small subset of group members.

- Generally, we show that cellular network data—CDRs—are a feasible and rich source of data to discover and analyze the behavior of social groups, since they capture both social interactions and a medium-grain mobility needed to identify likely group get-togethers.

The paper is organized as follows. In Sect. 2 we describe the cellular network data by providing an explanation of the social and localization information and then by discussing their advantages/limitations. In Sect. 3 we introduce the procedure for the identification of the urban groups from mobile phone data; as to the details about the single steps and their complexity, we present those in the Appendix. In Sect. 4 we focus on the size and the membership of the urban groups identified by our methodology and report their main spatio-temporal characteristics. Then in Sect. 5 we report our results concerning the preferred locations of the urban groups, the identification of leaders/followers within them, and the presence of strongly interactive relationships within these tight-knit groups. In Sect. 6 we discuss urban groups from the metropolitan viewpoint highlighting the city areas which facilitate and support urban group gatherings. Finally, in Sect. 7 we summarize our contributions.

2 Dataset

We performed our analysis of urban groups by mining a large anonymized dataset of Call Detail Records (CDRs) involving the voice calls, short text messages (SMS) and Internet traffic of about 1 million subscribers of one of Italy's major mobile operators [11]. The information provided in the database covers the metropolitan area of Milan for a period of 67 days, namely from March 26 to May 31, 2013. During this period approximatively 63 million phone calls and 20 million text messages were exchanged, all of which were recorded in the database. The temporal window covered by the dataset is extensive enough to reconstruct most of the on-phone social relationships [12].

2.1 Data description

For billing purposes, cellular network operators trace their customers' activities [13]. So, whenever a user makes a call, sends a text message or accesses the Internet, an entry is recorded in the charging database. Each entry in the CDR is represented by the 6-ple $t_{\text{CDR}} = \langle s, r, t_{\text{start}}, d, loc_{\text{start}}, loc_{\text{end}} \rangle$, where s and r respectively represent the sender's ID and the receiver's ID ,[a] t_{start} is the initial time of the activity (when the call starts or a text is sent or Internet access occurs), d is its duration, and loc_{start} and loc_{end} are the serving cells the user s is attached to when the activity gets started and has ended. Depending on the type of activity that has occurred, the information provided is different, so leading to the following uniquenesses: (i) both SMS (i.e. text message) and Internet activity have null duration d; and (ii) Internet activity has the field receiver r set to null.

2.2 User's localization

CDR-based datasets have been adopted extensively in literature to study human mobility patterns [14–18]. All these research projects derive locations by positioning cell towers in geographical areas where each cell tower may cover a zone as wide as a few kilometers. The dataset we are leveraging reports data about cell towers within a city space, where a dense placement of cells (one or very few hundred meters of coverage radius) has adopted. This feature enables quite an accurate localization of users while they are performing their

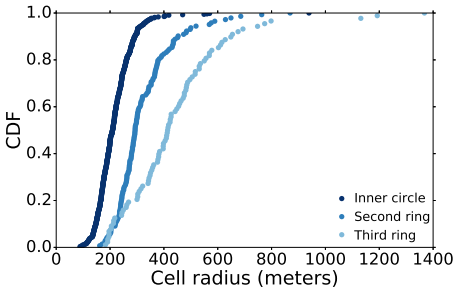

Figure 1 Cell radius. Cumulative distribution functions (CDF) of the radius of the cells in the three city regions

on-phone activity. As we will promptly show, the mean cell radius we consider is about 200 meters, or roughly a city block.

As the dataset contains labels assigned to area names, i.e. zones covered by a group of cells but without information about cell size and precise positioning [19], we adopted the following procedure to estimate the effective cell size distribution and position. We assume that each cell $cell_i$ is a circle with center c_i and radius r_i. To estimate the center of the cells we use the web-service *UnwiredLabs*[b] named `LocationAPI` that provides the cell center along with the estimated error. Currently, we are not using this last data and we assume that the cell center corresponds to the exact position provided by the system. For each cell, $cell_i$, r_i is half the mean of the Euclidean distances between the center of $cell_i$ and the centers of the six closest cells .[c]

As Milan has a radial topology, we consider three city regions: the inner circle of 3 Km radius, corresponding to the city center; a second ring in the range of 3 Km to 4 Km moving outward from the city center; and a third ring, in the range of 4 Km to 5 Km. The inner city circle corresponds to downtown Milan, while the other rings include suburbs. We obtain 538, 143, and 88 cells inside each region, respectively.

Having mapped the cell tower onto the city and computed the cell radius, we analyze the radius as a function of the cell position. The cumulative distribution function (CDF) of the cell radius for each city region is reported in Fig. 1. From the figure it emerges that the radius of the cells increases as we move farther from the city center. In fact, the mean of inner circle, second ring and third ring are 217, 325 and 446 meters, respectively. Given this small coverage radius we are able to provide a good approximation of the mobile users position suitable for the detection of their co-location.

3 Methodology

Social groups are often identified by the notion of cohesive groups, i.e. subsets of individuals among whom there are frequent and relatively strong interactions. Within these groups, beliefs, interests and ideas are often very homogeneous due to the pressure to achieve uniformity and adhere to group standards exerted by intense interactions [20]. Places figure among the interests of a cohesive group. In fact, shared places encourage the formation and consolidation of social relationships; conversely, groups might choose a specific place as conducive to expressing themselves better.

Combining quite a precise positioning of the customers with their on-phone relationships, our mobile phone data enable us to identify and characterize cohesive groups that couple strong on-phone interactions with the attitude to share specific urban places where they co-locate to perform various social activities, e.g. family-, work- and leisure-oriented

ones and/or participatory events. We call them *urban groups*. So, given a graph expressing the relationships among the operator's customers, an urban group is identified by a particular subgraph, a *quasi*-clique [21], whose a subset of members co-locate at least once. The subset cardinality is governed by the parameter η.

Operationally, to identify the social relationships we leverage the communication activities modeled as a graph. Meanwhile, we exploit the customers' localization to discover the aggregation in urban spaces. To this end, we perform three steps, namely: *interaction graph building*, *cohesive group identification* and *co-location filtering*. As our final output we obtain the set of urban groups, along with the information of the aggregation events. Our approach differs from previous works which have studied social groups by mining Bluetooth proximity data [22], since in our dataset the interplay between cohesive groups and physical proximity is not immediate and direct as in the Bluetooth case.

3.1 Interaction graph building

The purpose of the first step is to reconstruct the network structure of the interactions mediated by both voice calls and text messages. Following the standard approach in literature we represent such a complex structure by a graph whose nodes are customers and whose edges connect two customers who communicate by calls or texts [23, 24].

However, the choice of linking two users depends on the purpose of the communication. In fact, all calls and texts do not have the same social value; this is particularly true in the case of advertisements and commercial messages or communications issued by call centers. Moreover, we have to take into account missing links between other operators' subscribers since we have full access to the call/text records of one operator but only partial access to calls to/from subscribers of other operators. To cope with the above issues and obtain a graph which models the relationships between the operator's customers only, we filter out incoming and outgoing communications that involve other mobile operators' customers,[d] according to the literature on mobile phone cleansing [24–27]. This way we eliminate the inter-operator bias.

After applying the filters, we construct two preliminary graphs, one for each communication channel, from which to extract only the interactions with social relevance [28]. To this end, in the weighted call graph $G_c = (V_c, E_c)$, we consider the pairs of users whose sum of call durations exceeds one minute and whose total number of interactions is higher than 3 and we store this last value in the attribute f_c of the link. In the text message graph $G_t = (V_t, E_t)$, rather, the only relevant pairs are those with a total number of interactions higher than 3. This value we store in the attribute f_t. Through the filtering on duration and frequency, we are able to remove accounts/users whose behavior (degree, in/out degree) resembles call centers or customer care services. In the final step we merge G_c and G_t into the *interaction graph* G by taking $G_c \cup G_t$. To keep the information about the number of interactions, for each e in G we sum the attributes f_t and f_c if $e \in E(G_c) \cap E(G_v)$, while we keep the original attribute if e is not in the intersection. We denote the overall number of interactions (strength) in G as w. After the building process, the interaction graph, whose order and size have been reported in Table 1, captures the network among the operator's subscribers and the strength of their interactions which more likely express social relationships. The interaction graph is the input of the next stage which identifies cohesive groups.

Table 1 Summary of the properties of the interaction graph G. The first three columns report the number of nodes, the number of links and the density of the graph, respectively. $\hat{k}$ and $\hat{w}$ indicate the average degree and the average strength. The last two columns report the percentage of nodes in the giant connected component and the average clustering coefficient $\hat{c}$

Nodes	Links	Density	$\hat{k}$	$\hat{w}$	% nodes in GCC	$\hat{c}$
289,448	429,273	$1.02 \cdot 10^{-5}$	3	29	78%	0.12

3.2 Cohesive group identification

Representation of the on-phone communications through an interaction graph allows us to identify cohesive groups of customers, i.e. subsets of users among whom intense, direct and frequent ties do exist. The identification of cohesive groups, which is a central problem in both graph theory and social network analysis, entails different methods—from community detection [25, 29] to enumeration of particular maximal subgraphs [23, 30]. In this work we focus on the latter approach since community detection methods, when applied to this phone graph, have been shown to return loosely connected subgraphs barely interpretable as groups or tight-knit communities [31]. In fact, the communities detected by different algorithms are characterized by an average density which varies from 0.019 (Louvain algorithm [25]) to 0.35 (Leung's algorithm [32]). Such values indicate weak cohesiveness of the members within the communities, whatever the algorithm we used; making the community approach unsuitable for the identification of cohesive groups. Similar conclusions have been reported in [33], where authors claimed that Louvain and InfoMap algorithms applied on phone graphs (weighted or unweighted) yield tree-like communities which do not fit well with the notion of social group.

Among the different formalizations of cohesive groups, we adopt a relaxation of the notion of clique, namely the *quasi-clique*, i.e. a particular dense subgraph. The notion of clique well embodies one of the main properties of a cohesive group, i.e. the mutuality. But the completeness of the subgraph is too strict a constraint. In literature many definitions that weaken the notion of clique have been proposed. They range from n-cliques or n-clubs to k-core [20]. Here we use the notion of quasi-clique or γ-clique, since it allows us to quantify how much we loosen the completeness constraint; meanwhile, at the same time, it ensures the reachability of the group members, a further property of cohesive groups. Formally, given a graph $G = (V, E)$, a γ-clique is subgraph G_S spanned by S, a subset of V, that is connected and γ-dense. G_S is γ-dense if $|E(G_S)| \geq \gamma \binom{V(G_S)}{2}$. In this work we use $\gamma = 0.8$ because it is a good trade-off between imposing too strong constraints on the subgraph density and loosing the idea of cohesive group. Indeed, values below 0.8 lead to a loss of cohesion in case of large groups, whereas values above 0.8 are too restrictive for small groups, because almost all pairs of nodes should be connected. Besides, above the 0.8 threshold, the number of detected quasi-clique significantly drops (–73% for $\gamma = 0.9$) and causes a loss of generalizability. Following our approach, the identification of cohesive groups turns into the enumeration of all quasi-cliques of maximum cardinality. To accomplish this task we adopt the Uno's enumeration algorithm [34] which returns all the locally maximal quasi-clique in a given graph. Then, for each quasi-clique we verify whether it is connected or not, discarding the unconnected ones. This way we identify all the connected locally maximal quasi-cliques, representing the cohesive groups whose members would be verified to be co-located in the last stage.

3.3 Co-location filtering

The high spatial granularity of the data enables us to localize users with a precision of the city block when an on-phone activity is performed. We exploit the location information to detect the co-location of the quasi-clique members. We extract from the CDR 6-tuple the sequence of the recorded locations of each user, along with the temporal annotation. Thus we obtain an array $T_{\mathrm{MOB},u}$ of 2-element sets (loc, t) called the mobility trace of the user u.

The mobility traces of all the users are the starting point of the co-location algorithm. As we are interested in detecting the co-location of the members of the quasi-cliques, the co-location algorithm runs on each quasi-clique separately. Specifically, a quasi-clique experiences a co-location event when a fraction η of its members share a location for a time period. In this work we use $\eta = 0.6$. The output of the algorithm is the list of co-location events, where each co-location event is identified by the triplet $\langle (t_s, t_e), loc, M_e \rangle$, where t_s and t_e are respectively the starting and ending times of the co-location time interval, loc is the location, and $M_e \subseteq M$ is the set of quasi-clique members participating in the co-location event. So, the co-location algorithm checks if a cohesive group is an urban group and identifies when and where an urban group gets together. For more details about the co-location filtering algorithms see the Appendix.

4 Urban group behaviors

In this section we analyze the structural and spatio-temporal characteristics of urban groups, showing that urban groups represent a significant portion of all existing cohesive groups in the interaction graph. From a social viewpoint, urban groups are statistically similar to other groups found in different socio-technological social networks. The number of members in each urban group, i.e. the size, is quite small, very similar to the size of WhatsApp groups [35], and favors the formation of strong relationships. Moreover, the level of overlapping among different groups is lined up with other social networks, expressing the attitude of groups to connect around a particular interest. From a spatio-temporal viewpoint, urban groups also present interesting characteristics. They usually prefer to meet in very few locations and often experience co-location events, i.e. they get together on average every three days.

4.1 Size and membership

A preliminary albeit fundamental aspect of our investigation on urban groups is to measure their relevance within different types of social aggregations, i.e. the number of urban groups related to the overall number of cohesive groups. To this aim, we compare the number of cohesive groups before and after the co-location filtering. We find that *most of the cohesive groups we can capture through on-phone interactions are urban groups*. In particular, we identify more than 28,000 urban groups. They represent 75% of the quasi-cliques with size greater than 4 in the interaction graph, and involve about 23,800 of the operator's subscribers. To assess whether the emergence of the urban groups is not only due to the well-known correlation between the on-phone interactions and co-location which characterizes the reciprocal calls between pairs of users [15, 36], we test if the measured number of groups is significantly higher than the one obtained by a null model, in which a dependency between communications and co-location exists. Specifically, the null model is based on the co-location graph studied in our previous work [31] and on the observation that, given a link between two customers in the co-location graph, the probability

that they communicate by call or text is 0.06. So, for each link in the co-location graph we draw the corresponding link in the interaction one with probability 0.06, then we extract the quasi-cliques. We repeat the model generation 100 times and we measure the significance. We obtain a p-value much lower than 0.001, showing that the aforementioned correlation at a link level alone does not explain the emergence of the measured number of cohesive groups.These findings suggest that (i) the correlation between physical proximity and on-phone interactions, which holds for pairs of users [15, 36], can be extended to groups; and (ii) on-phone social networks are much more accurate than their online counterpart in mirroring people's offline sociality. Meanwhile, they share with them the power to generate high volumes of data traffic.

With the ever growing relevance of social networking sites, the size of a group of persons represents one of the main aspects of a social environment, since it influences the strength of relationships, the intensity of participation in group activities and the consonance of aims [37]. In Fig. 2a we show the probability distribution function of the urban group size. It highlights that small groups ($k = 5, 6$) are predominant in mobile phone networks. Moreover, the short tail of the distribution—its maximum value is 13—indicates a substantial difference w.r.t. community detection approaches. In fact, community detection algorithms identify hundred/thousand-people communities, whereas these groups vanish when we search for highly dense regions in the mobile phone graph. The result supports the findings in [33] showing that community detection algorithms may return loosely connected subgraphs that we can vaguely assimilate to tight-knit groups or communities. Surprisingly, by comparing the group size with the group size measured on WhatsApp [35], we observe that their sizes are very similar.

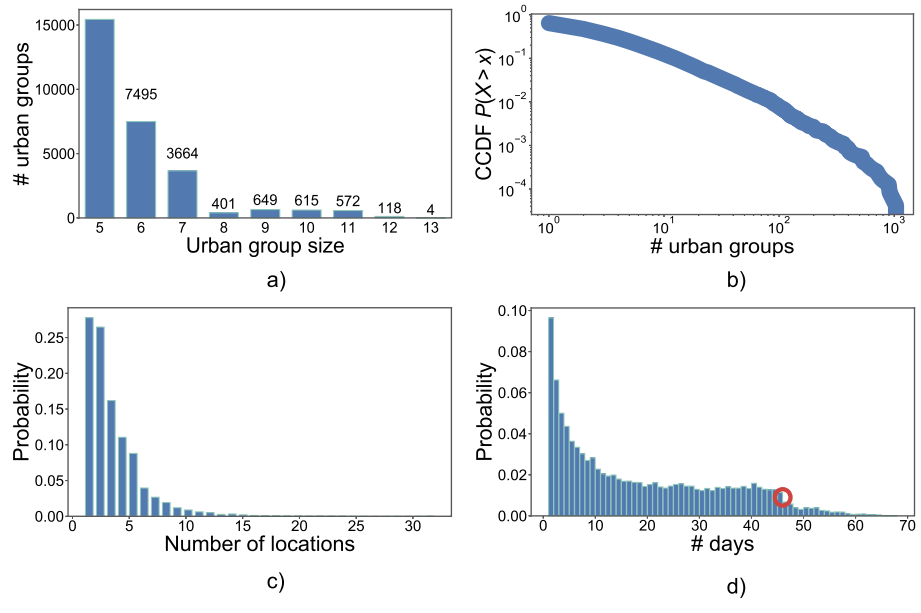

Figure 2 Urban group characteristics. (**a**) The histogram of the urban groups size. (**b**) The distribution of the number of urban groups per user. (**c**) The histogram of number of locations in which an urban group meets up. (**d**) The histogram of the number of days an urban group meets up. The red circle indicates the sharp decrease of the probability after value 46 of the number of days, which corresponds to the number of working days within the observation time period

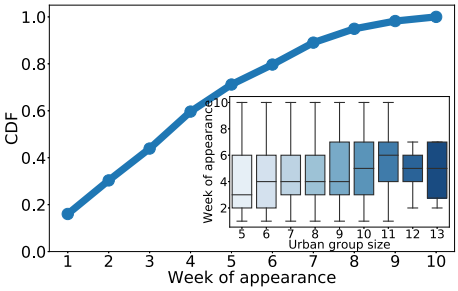

Figure 3 Urban group stability. The cumulative distribution function of the probability that an urban group is detected within a certain number of weeks. In the inset, the boxplot of the probability that an urban group is detected within a certain number of weeks grouped by urban group size

The formation of an urban group depends on the time needed by the subgraph to reach the minimum required density γ. To test how stable the definition of urban groups at different time periods of different length is, for each group, we measure the number of weeks needed so that the subgraph reaches the required density threshold. In Fig. 3 we report the distribution of the probability that an urban group is detected within a certain number of weeks. As we can see, most of the urban groups (around 90%) are detected within 7 weeks. Moreover, we find that this result does not depend on the size of the group, as we can observe in the inset of Fig. 3, where the distribution grouped by group size is shown.

Groups could form around common interests or existing social structures, such as family, workmates, teammates, so an individual may likely participate in different social groups [37, 38]. To verify whether this phenomenon holds also for urban groups, we investigate if the operator's customers belong to a single group or if they participate in different groups, each corresponding to different interests [39]. To this aim, in Fig. 2b, we report the distribution of the number of urban groups a user belongs to. The distribution follows a heavy-tail trait, i.e. most of users belong to few cohesive groups, but people participating in many urban groups do exist. In particular, half of the population share at most 2 urban groups, while the average number of groups per user is 6. Similar results have been observed in other social networks, such as Flickr [38] and LiveJournal [40].

4.2 Locations and visit patterns

Given the strict interplay among groups, interests and places, urban groups are supposed to meet in specific locations, somehow related to the group activity. We identify a group gathering by detecting when its members are co-located in a cell tower. However, cells have different coverage radius according to the distance from the city center (see Fig. 1) and this could affect the characteristics of the urban groups. In particular, the larger the coverage radius the higher the probability of co-location events among group members and this could lead to an overestimation of the size of the urban groups. To investigate how the length of cell radius affects the characteristics of the urban groups, we only consider urban groups that get-together in the cells that belong to the innermost ring and we repeat the analysis we conducted in the previous section. We perform the Kolmogorov–Smirnov test and we obtain the following results: 0.053 (p-value < 0.001) for the distribution of group size and 0.033 (p-value < 0.001) for the distribution of number of groups for each subscriber. Based on these results showing no significant statistical difference between the distributions, we do not make any restriction on where a co-location event takes place.

To investigate the connection between locations and urban groups, we measure the number of locations where each group gets together. In the following, we will use the

notion of location instead of cell to overcome the artifacts introduced by the network load balancing algorithm, which associates mobile users to different cells, according to the current network status, even if the users' position does not change. We exploit a coarse subdivision of the metropolitan area directly provided by the network operator and aggregate neighbor cells in groups of size from 8 to 15. In Fig. 2c we report the histogram of the number of locations where each group co-locates. As we can observe, most of the groups co-locate in very few places; mean, median and standard deviations are 3.16, 2 and 2.54, respectively. This result shows that *urban groups are characterized by few visited locations*, in strict analogy with individuals' mobility [18, 41].

As urban groups are characterized by a tight-knit network of communications and a limited set of preferred locations, a question arises about whether or not groups need to combine frequent encounters with on-phone interactions to express the group sociality. To this aim, we analyze the continuity of the encounters of each urban group by computing the number of days each urban group co-locates. In Fig. 2d we show the histogram of the number of days each group is co-located (we consider a group co-located in a day if at least one co-location event exists on that day). The mean, median and standard deviation of the number of days distribution are 18.20, 14.0 and 15.33, respectively, with more than 70% of groups meeting on more than 5 days. This result shows that the encounters among the members of a group are not sporadic and indicate some regularity. We can argue that, on average, *urban groups need to combine on-phone interactions and get-togethers in a few urban places to fully express and support their activities*.

4.3 Interactivity of the urban groups

Along with mutuality and reachability properties, interactivity—i.e. the frequency of interactions among members—defines a cohesive group. For a group to be cohesive, it is in fact required that the group members maintain frequent interactions with one another. By leveraging the number of interactions between the pairs forming a group, we can evaluate if the interactivity property holds for urban groups and, consequently, measure the effort members devote to maintaining their relationships inside a group. In line with previous works on subgraphs in call graphs [12], we adopt the intensity int of an urban group to assess the effort of maintaining the relationships within an urban group. The intensity of an urban group ug_i is defined as the geometric mean of its link weights:

$$\text{int}(ug_i) = \left(\prod_{(i,j)\in E(ug_i)} w_{i,j} \right)^{1/|E(ug_i)|}, \tag{1}$$

where $E(ug_i)$ denotes the links forming the urban group ug_i. Here the effort, i.e. the link weight, coincides with the number of interactions.

In Fig. 4a, we report the distributions of the urban group intensity grouped by the size of the subgraphs. We observe that for $k = 5, \ldots, 9$ the distributions are very similar, while for bigger groups the distributions shift towards higher intensity values. We make this trait more explicit in the inset figure, where we show the box-plot of the intensity as a function of the group size. Each bar spans the likely range of variation (from first to third quantile), the segment inside the rectangle indicates the median of the distribution and the points below and above the whiskers represent outliers. The figure highlights two important points: first, regardless of the size of the urban group, more than 75% of groups

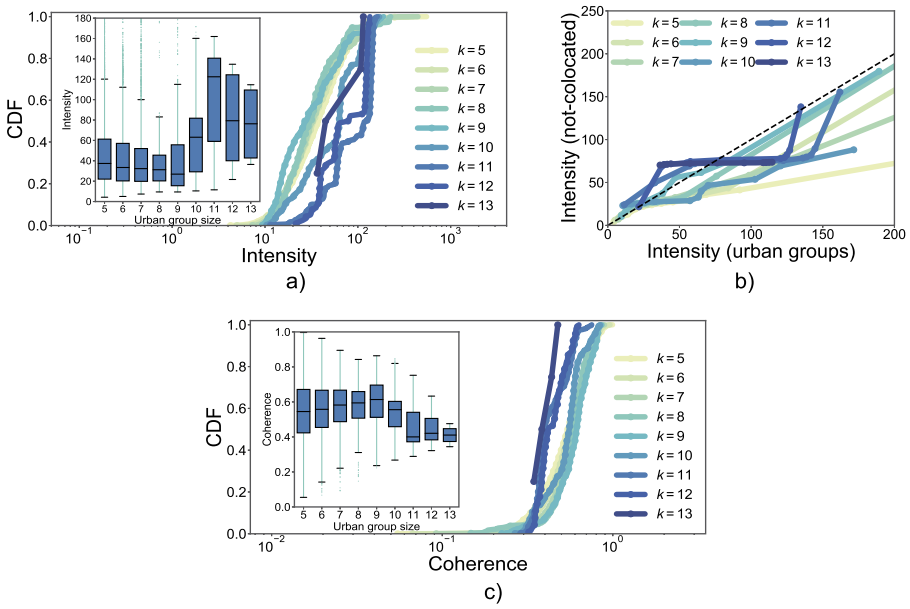

Figure 4 Intensity and coherence. In (**a**) the distribution of the intensity int for $k = 5, \ldots, 13$. The x-axis is logarithmic and the y-axis is linear. In the inset figure, the box and whisker plots of the intensity as a function of the urban group size. Boxes span the 1–3 quartile interval and the whiskers indicate the boundaries after which data are outliers. In (**b**) the qq-plot between the distribution of intensity evaluated on urban groups (x-axis) and not-colocated cohesive groups (y-axis). The dotted black line indicates the theoretical trend in case of equality of the distributions. The points represent 0.1 equally spaced quantiles. In (**c**) the distribution of the coherence coh for $k = 5, \ldots, 13$. The x-axis is logarithmic and the y-axis is linear. In the inset figure, the box and whisker plots of the coherence as a function of the urban group size

reach an intensity higher than or equal to 20. So, *the members within these groups interact more than 20 times with each of the other members.* Secondly, we distinguish two typical behaviors involving groups with $k = 5, \ldots, 9$ and larger groups ($k = 10, \ldots, 13$). Specifically, smaller groups are mainly characterized by an average intensity within 20 and 60, while in bigger groups the intensity intervals range from 50 to 130. This shows that *members of large urban groups devote many efforts in interacting with one another and to maintaining relationships established within the group.*

We have just shown that a high level of interactivity characterizes urban groups. Now we ask whether the physical proximity of the members of the urban group impacts the interactions occurring within the group. That is, we wonder if the co-location property stimulates on-phone interactions within cohesive groups. To this aim, we compare the distributions of the intensity in urban groups and not-colocated cohesive groups. In Fig. 4b we report the comparison by the qq-plot for different sizes. By the qq-plot we are able to verify whether or not two distributions are equal by computing and displaying their quantiles. The 45° line in the figure represents the identity case (black dotted line), while in case of diverse distributions the plot lies below or above the line. In the figure we observe that for smaller groups ($k = 5, \ldots, 8$), the distributions of the intensity are similar only for the first 0.1-quantiles, while urban groups show higher values of intensity for the highest quantile, i.e. with $k = 5, \ldots, 8$ urban groups are more interactive than not-colocated cohesive groups of the same size. For bigger groups, this phenomenon is even more accentuated, since not-colocated groups take higher values for the lowest quantiles than urban groups; by contrast, in urban groups the highest quantiles are much higher than

in not-colocated groups. In general, we find that *urban groups are much more interactive than their not-colocated counterparts.* So, the opportunity of meeting in urban spaces strengthens the relationships expressed by on-phone communications; meanwhile tight-knit groups, whose interactions are strong and frequent, likely co-locate in a few specific locations.

5 Preferences of urban groups

People's aptitude to prefer specific elements is an across-the-board aspect affecting diverse human activities, from online social engagement, where users frequently interact with a strict subset of their online friends [42, 43], to offline activities, where a limit on the number of people with whom an individual establishes stable social relationships has been shown [44–46]. This aptitude also holds for human mobility, whose footprint can be described by very few most visited locations [18]. Here, we investigate if such an aptitude characterizes the behavior of the urban group members; specifically, we ask whether or not a backbone of strongest ties exists within urban groups, if groups have preferred meeting places and whether or not different roles emerge.

5.1 Favorite interactions within urban groups

The previous results about urban group intensity have shown the average effort to maintain relationships within co-located cohesive groups. However, this behavior could be the effect of relationships much more active than others, i.e. heterogeneity, or a homogeneous interactivity involving all the ties forming a group. To assess the homogeneity among interactions in a group, we measure the coherence of an urban group ug_i. Given an urban group ug_i, its coherence $\mathrm{coh}(ug_i)$ [47] is defined as:

$$\mathrm{coh}(ug_i) = \mathrm{int}(ug_i)\big|E(ug_i)\big| / \sum_{(i,j)\in E(ug_i)} w_{i,j}. \tag{2}$$

By AM-GM inequality,[e] the coherence takes values between 0 and 1. The more homogeneous the ties within an urban group, the closer to 1 the coherence. Figure 4c shows the distributions of the coherence for $k = 5, \ldots, 13$ and the inset figure reports the box and whiskers plots of the coherence as a function of the group size. The figure indicates that the distributions for $k = 5, \ldots, 10$ are very similar, while larger groups are characterized by coherence values closer to 0. Regardless of the high variability of smaller groups, the inset figure indicates the same trait. Specifically, urban groups with size $k \in [5, 10]$ have a median value close to 0.6, while larger groups ($k \in [11, 13]$) result in a median close to 0.4. These results indicate that the interactivity of the links within urban groups is more uniformly spread in smaller groups than in larger ones. In fact, in these cohesive groups the relationships between some pairs of members are stronger than others. So, *in larger groups the social effort, i.e. the number of interactions, is more focused on some specific relationships, while in smaller groups the effort is more evenly balanced among all ties.*

5.2 Favorite locations

We have shown that both individuals and groups share the attitude to visit a small set of locations. We know from the literature that individuals are actually very regular in this and have a few favorite locations [48]. Do urban groups behave similarly? We approach the

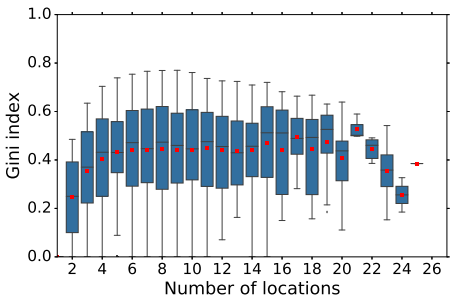

Figure 5 Visited areas. The boxplot of Gini-index of visited areas by urban groups

Table 2 Descriptive statistics of the favorite locations analysis

Minimum days	Mean	Median	Std.	75-pct	90-pct	95-pct
50%	1.09	1.00	0.54	1.00	2.00	2.00
60%	0.89	1.00	0.51	1.00	1.00	2.00
70%	0.78	1.00	0.51	1.00	1.00	1.00
80%	0.68	1.00	0.51	1.00	1.00	1.00
90%	0.54	1.00	0.52	1.00	1.00	1.00
100%	0.40	0.00	0.51	1.00	1.00	1.00

analysis in two ways. First, we compute the Gini index of the number of days each group meets in a particular location. Secondly, for each urban group we extract the number of locations where a group meets at least a given percentage of days. To avoid the bias introduced by groups meeting too infrequently, we restrict the analysis to those urban groups having a number of distinct days greater than 2. In Fig. 5 we report the Gini index distribution grouped by the number of locations visited by urban groups. As we can observe, the values of Gini index are far from 0 (equality condition); mean values range from 0.35 to 0.50, if we consider a number of different locations somewhere between 3 and 22. These results highlight that almost all groups distribute their get-togethers unevenly among the set of locations. Table 2 reports the descriptive statistics of the number of locations which satisfy the condition using different percentage values. From the results it emerges that most of groups have at most one location, and the median value is 1 except for the highest values of the percentage of days. It is worth noting that this characteristic holds both for groups visiting just a few locations and those visiting many locations. So, *urban groups have the tendency to meet in very few favorite locations, disregarding the total number of locations visited by their members. This aspect holds both for individuals and groups.*

5.3 Role discovery: leaders and followers

Today's massive diffusion of instant messaging services is rooted in the advent of on-phone communications that, ever since their introduction, have made interactions within a group of persons easier. This is confirmed by the previous results showing that the on-phone interactions of urban groups are intense and frequent. By contrast, face-to-face interactions require considerable effort to synchronize all members of a group. Thus, it is quite uncommon for individual members of an urban group to participate in all the get-togethers of that group. That is, we wonder if the mutuality and interactivity properties also characterize face-to-face interactions, or, by contrast, if a bias exists among the members.

The relaxation of the urban group's members presence, governed by the η parameter in the co-location filter, allows us to capture, for each co-location event, the presence of

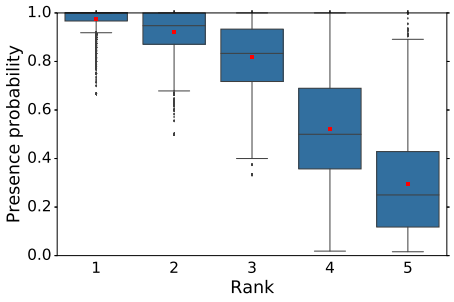

Figure 6 Member presence. The boxplot of the presence probability of each member of all urban groups of size 5

a subset of the group's members. Given this information we are able to measure the degree of participation of each member in the group gatherings. To this aim, we compute the presence probability of each member as the ratio between the number of days the member participates in urban group gatherings and the total number of days in which the group got together. Figure 6 shows the box-plot of the presence probability of each member for all urban groups of size 5. Similar results are observed for the other group sizes. The value of rank indicates the importance of the member in terms of days of presence. Thus, rank equal to 1 refers to the most present member while value equal to 5 refers to the least present member. From the figure we observe that the distribution of the presence probability of all the members having the highest rank is concentrated very close to 1 (mean and standard deviation are 0.98 and 0.05, respectively), meaning that these users participate in almost all gatherings of the group they belong to. By contrast, for rank 5 we observe lower and broader values (mean and standard deviation are 0.39 and 0.29, respectively). Given these results, let us divide urban group members into two main groups: leaders and followers. A leader is a member who frequently participates in urban group gatherings, whereas a follower is a member who sometimes or rarely joins group get-togethers. While it is easy to identify the leader, or leaders in some cases, it is harder to categorize a member as a follower, as we can observe from Fig. 6. In fact, the distributions of the presence probability of members with ranks 4 and 5 are spread. Thus, there are groups where the distinction between leader and follower is more pronounced, and others where it is indefinite. Clearly, this aspect reflects the variety of behaviors within each single urban group. The results show *the existence of a bias among the members taking part in urban group gatherings: there is a subset of members (the leaders) who take part frequently in the get-togethers, while another subset of members (the followers) who are less involved in the group's face-to-face interactions*.

6 The urban groups from the city viewpoint

When we try to map the activities of these cohesive urban groups in an urban space, or a city, we comprehend how their behavior and dynamics greatly influence the design, planning and dimensioning of both online and offline services. For instance, they shape the traffic flows of mobile networks, affect the planning of urban services, inspire the rise of new location-based services, and direct advances in content management and mobile edge computing. In this section we analyze the co-location events through the lens of the city so as to investigate how urban group gatherings are distributed in the city space. In particular, we are interested in finding popular locations and differences between day- and night-time [49]. To perform the analysis we consider all co-location events that occurred

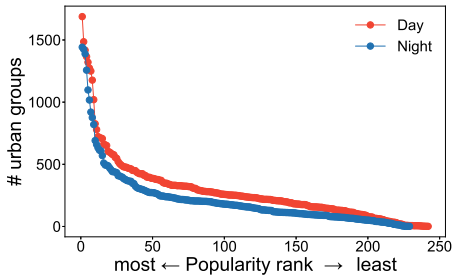

Figure 7 Meetings. Number of urban groups meeting in each city location during the day and the night hours

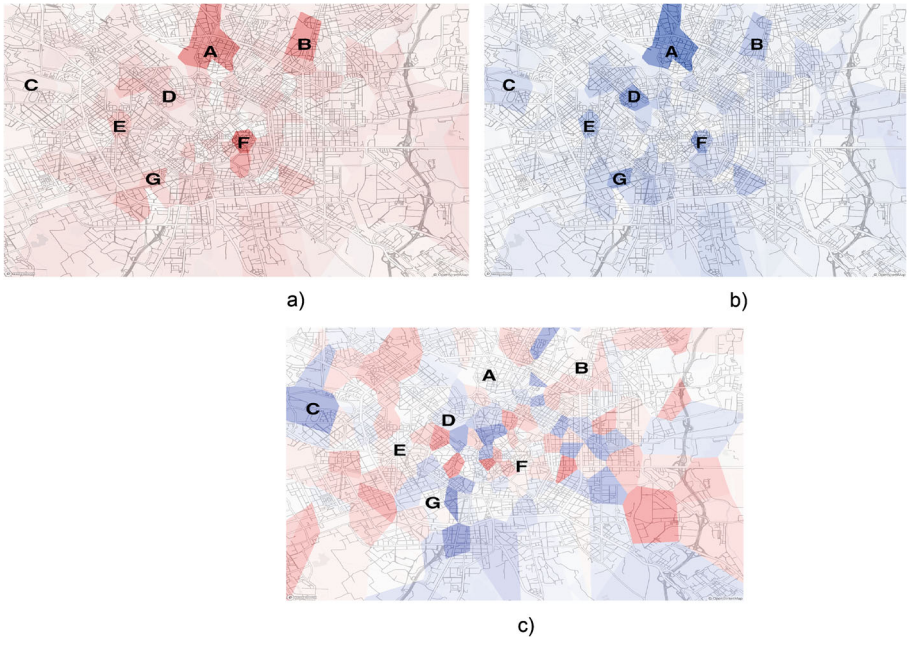

Figure 8 Location popularity. (**a**) The heatmap of the location popularity during the day hours. (**b**) The heatmap of the location popularity during the night hours. (**c**) The heatmap of the rank difference, red means that the location loses popularity between day and night, while blue means that the location gains in popularity, the color intensity reflects the absolute value of the rank difference of the location

during the entire dataset time interval and we divide the events that took place during the day from those that took place at night. We consider a co-location event belonging to the day if it occurred between 8:00 a.m. and 7:00 p.m.; otherwise we consider the event as a nighttime one.

In the Fig. 7 we report the number of distinct urban groups that visit[f] each city location (sorted from the most to the least popular) by distinguishing between day- and nighttime hours. As we can observe from the figure, the popularity of the locations is not uniformly distributed. In fact, a small set of locations have high popularity; only 9 and 6 locations have a number of urban groups higher than 1000 if we consider the day and night hours, respectively.

In Fig. 8a and 8b we report the heatmap of the location popularity during day and night hours, respectively. We can identify 7 city zones denoted by capital letters from *A* to *G*. In the discussion about the difference between day and night, we have to consider that Milan has no a clear division in functional areas, such as educational, business and shopping

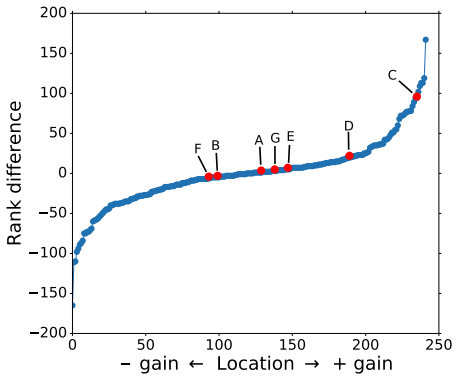

Figure 9 Location ranking. Rank difference between day and night for each location ordered by value of the difference in ascending order

districts. This characteristic clearly emerges from the two heatmaps where we can observe that most of the metropolitan areas are popular during both day and night. In particular, the region *A* is a business district that is also full of pubs and concert clubs, *B* is a residential area with small markets and shops, *C* holds the football stadium and concert arena, *D* is a place full of restaurants and pubs, *E* is a shopping, entertainment and nightlife district, *F* is the downtown area, and *G* is one of the Milan's most famous night life districts.

To deepen the analysis of the differences between day and night behavior we compute the variation of popularity rank. Figure 9 shows the rank difference for each location. Then in Fig. 8c we report the heatmap of the rank differences across the city map (red means that the location loses popularity between day and night, while blue means that the location gains popularity; the color intensity reflects the absolute value of the rank difference of the location). From Fig. 9, we can observe that most locations have a small variation, only 18% of locations have an absolute variation higher than 50, while the percentage decreases to 4% if we consider an absolute variation higher than 100. It is interesting to note that almost all the metropolitan districts considered exhibit a very small variation between day and night. The only exception is zone *C*, where the football stadium and concert arena are found—both of which are used mainly during the night hours. This result is due to the multi-functionalities of those areas, combined with the "happy hour"effect. Another interesting aspect that emerges from Fig. 8c is that some of the highest variations are in proximity of areas *D*, *F* and *G*. We can observe two opposite variations: the locations close to *D* and *G* zones gain in popularity during the night hours, while the locations surrounding *F* lose rank positions. These findings reflect the Milan nightlife, which moves outward the downtown (area *F*) to districts *A*, *C*, *D* and *G*.

7 Conclusions

Mobile social networks are evolving gradually toward serving the needs of small groups of friends who are very close to one another and/or share common interests. This new type of online social services is shifting away from the large communities of friends of former social networks; it is more oriented toward light-hearted amusement, intimacy and intense sharing of specific contents, and less to information and self promotion. A few emerging social networks, such as Snapchat or WeChat, the impressive rise of groups in WhatsApp, as well as the rise of interest-driven social networks, such as Strava, all prove the point: people like to share images and videos with a restricted group of close friends, a social circle where they feel comfortable talking about themselves, even acting goofy, and

not having to suffer the strain of performing in public or thinking hard before publishing a post [2]. The trend echoes Dunbar's social grooming [50] and leads us to envision groups consisting of few persons with strong social ties who interact frequently to informally share information about their daily life and they do that mainly by exchanging geo-localized information and camera-based messages[g] (videos or images).

When we try to map the activities of these groups in an urban space, or a city, we comprehend how their behavior and dynamics greatly influence the design, planning and dimensioning of both online and offline services. For instance, group mobility affects the planning of urban services and inspires the rise of new location-based services, while group interactions shape the traffic flows of mobile networks, and direct advances in content delivery and mobile edge computing.

This paper unveils the real nature of mobile and cohesive social groups, named urban groups, providing a thorough analysis and evidence of their behavior and dynamics, and showing that this achievement can be obtained by mining an anonymized mobile phone dataset based on Call Detail Records (CDRs). The analysis puts in the spotlight some interesting urban group behaviors. For instance: (i) urban groups are chiefly small social groups, whose members are very interactive; (ii) the group members move and keep interacting on the move; (iii) they have periodic gatherings and meet up in favorite city places, revealing that they are rooted in the territory; and (iv) it is easy to identify a group leader and the followers.

Appendix: Co-location filtering algorithm

The algorithm takes as input the set of members (M) of the quasi-clique, the list (T_M) of the mobility traces of all its members and three parameters: the minimum percentage of members required for each co-location event (η), the time threshold (Δ) and the temporal granularity (τ), whose meaning will be explained later. The output of the algorithm is the list of co-location events, where each co-location event is identified by the triplet $\langle (t_s, t_e), loc, M_e \rangle$, where t_s and t_e are the starting and ending time, respectively, of the co-location time interval, loc is the location, and $M_e \subseteq M$ is the set of p-clique members participating in the co-location event.

The pseudo code of the co-location filtering algorithm is depicted in Algorithm 1. The first step (line 2) initializes the set of potential locations, where co-location events could happen, as the union of the locations of each member. Here we need the union operator instead of the intersection because we do not impose that all members have to participate in a co-location event. Then the algorithm iterates over all the potential locations and performs two tasks: (i) temporal filling of the mobility traces of all quasi-clique members (lines 4–9) and; (ii) detection of the co-location (lines 10–13).

A.1 Temporal filling of traces

For each location the preprocessing of the mobility traces performs a transformation to ease the co-location events detection. It is composed by four sequential steps operating performed for each member. First, it sorts the original trace according to the timestamp in ascending order. Second, the procedure `TimestampToInterval` transforms each point of the trace, identified by the timestamp t of the CDR record, in a new point $\langle t_s, t_e \rangle$, where $t_s = t - \Delta$ and $t_e = t + \Delta$, representing the extremes of time interval the user is supposed to be. Here we assume that if the user was in a location at time t she/he remained

Algorithm 1 Quasi-clique co-location

1: **function** Co-location(M, T_M, Δ, τ, η)
2: $L \leftarrow \bigcup_{m \in M} L_m$ ▷ Set of potential locations
3: **for all** $l \in L$ **do**
4: **for all** $m \in M$ **do**
5: Sort T_m^l by timestamp in ascending order ▷ TimestampSort(T_m^l)
6: $I_m^l \leftarrow$ Convert timestamps t_m^l to intervals $(t_m^l - \Delta, t_m^l + \Delta)$ ▷ TimestampToInterval(T_m^l, Δ)
7: Merge time overlapping intervals in I_m^l ▷ MergeOverlapping(I_m^l)
8: $F_m^l \leftarrow$ Convert intervals in arrays of temporal ticks having τ as time granularity ▷ FillInterval(I_m^l, τ)
9: **end for**
10: $F_l \leftarrow$ Concatenate all traces for location l for all $m \in M$ ▷ $F_{m_1}^l \parallel F_{m_2}^l \parallel \ldots \parallel F_{m_n}^l$
11: $O_l \leftarrow$ Count the occurences of each temporal tick ▷ OccurrenceCount(F_l)
12: Remove temporal ticks where the membership cardinality is below the threshold τ ▷ FilterMembershipCardinality(O_l, η)
13: $COL_l \leftarrow$ Construct co-location intervals by merging consecutive temporal ticks ▷ OccurrenceToInterval(O_l)
14: **end for**
15: **return** all $COL_l \neq \emptyset$
16: **end function**

Notations

Inputs	
M	Set of members of the quasi-clique
T_M	Timestamped mobility traces of all members of the quasi-clique
Δ	Time threshold
τ	Temporal granularity of detection
η	Minimum percentage of quasi-clique members required to be co-located

Variables	
L_m	Set of locations visited by member m
T_m^l	Timestamped mobility trace of member $m \in M$ in the location l
I_m^l	Intervals mobility trace of member $m \in M$ in the location l
F_m^l	Filled intervals mobility trace of member $m \in M$ in the location l
F_l	Filled intervals mobility traces of all members in location l
O_l	List of co-location event occurrences in location l
COL_l	List of co-location event intervals in location l

in that location from $t - \Delta$ until $t + \Delta$, in line with [15] we use $\Delta = 30$ minutes. Third, the procedure `MergeOverlapping` takes all the time intervals and merges the ones that overlap. Given the above sorting, two consecutive intervals i_1, i_2 overlap if $t_s^2 \leq t_e^1$. In the last step, the procedure `FillInterval` converts each time interval in I_m^l into an array of temporal ticks according to the temporal granularity parameter τ ($\tau = 1$ minute). For instance, the produced array is $\langle t_s, t_s + \tau, t_s + 2\tau, \ldots, t_e - \tau, t_e \rangle$. By construction, the temporal ticks produced are unique over the dataset time frame, simplifying the co-location detection. The output of the task is a list of temporal ticks for each member of the quasi-clique. When the task ends we get for each member the set of intervals during which the user was in a specific location.

A.2 Co-location detection

Finally, the co-location detection task exploits the temporal tick representation of the mobility traces and performs a simple counting. In detail, in line 10, the concatenation of all the lists of temporal ticks is performed. This results in F_l. Then the procedure `Occur-`

`renceCount` takes F_l and counts the occurrences of each single temporal tick; as output, it produces the list of occurrences identified by the tuple $\langle tick, \{co\text{-}located\ members\}\rangle$. In the next step, the procedure `FilterMembershipCardinality` filters out the occurrences in which the number of co-located members is below the threshold η. In the last step, the procedure `OccurrenceToInterval` transforms the list of occurrences in a list of intervals by aggregating adjacent temporal ticks. When this second task terminates, we obtain the set of temporal intervals.

A.3 Co-location filtering algorithm complexity

Now we briefly discuss the time complexity of the co-location algorithm using the following notation: n as the number of records of the mobility trace, m as the number of the quasi-clique's members, and l as the number of potential locations.

The temporal filling task is performed m times, one for each member, and its time complexity is dominated by the `TimestampSort` procedure which is $O(n \log n)$, obtained by using a classical sorting algorithm. The procedures `TimestampToInterval` and `FillInterval` are linear in the number of records because they perform a transformation of all elements by taking $O(1)$ time for each element. The procedure `MergeOverlapping` is also linear w.r.t. n because it exploits the ordering and the equal length of the time intervals. Thus, the checking of the overlapping condition is limited to two consecutive intervals only. The resulting time spent by the algorithm to perform the filling task over all members is $O(m \cdot n \log n)$.

The co-location detection task is performed once per each location and is linear in the number of records, $O(n)$, because the procedures `OccurrenceCount` and `FilterMembershipCardinality` iterate over all records by performing constant time operations; and the procedure `OccurrenceToInterval` can be optimized in order to perform a linear scanning over all occurrences by checking the adjacent condition between two consecutive temporal ticks only.

The two previously discussed tasks are performed l times, one per each location. Thus, the overall time complexity of the co-location algorithm is $O(l \cdot (m \cdot n \log n + n))$. It is worth of noting that in a real application scenario the number of users belonging to a quasi-clique is very small and, due to the high regularity of the users' mobility, the set of locations visited by a single user is small [18]. Consequently, we have $m \ll n$ and $l \ll n$ and we can rewrite the time complexity as $O(n \log n + n)$.

The proposed algorithm is highly parallelizable. At the highest level, each quasi-clique can be analyzed separately. As further optimization each location can be processed in parallel. Moreover, the temporal filling task can also be parallelized.

Acknowledgements
Not applicable.

Funding
This research has been funded by the Project Phydia (2016–2018) through the transition grant of the University of Milan.

Abbreviations
CDR, Call Detail Record; SMS, Short Text Message; CDF, Cumulative Distribution Function; AM, Arithmetic mean; GM, Geometric Mean.

Availability of data and materials
The data that support the findings of this study are available from the cellular network operator but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the cellular network operator.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
MZ, CQ, SG and GPR conceived and designed the experiments. MZ and CQ performed the experiments. CQ, MZ, SG and GPR analyzed the data. CQ, MZ, SG and GPR wrote the paper. All authors read and approved the final manuscript.

Endnotes
[a] For the purpose of ensuring customer anonymity, each subscriber is identified by a surrogate key.
[b] Website http://unwiredlabs.com/.
[c] The choice of the six closest cells is due to the conventional representation of cells as hexagons.
[d] Whether or not an anonymized number belong to an operator's customer has been provided by the mobile operator.
[e] Inequality of arithmetic and geometric means states that the arithmetic mean of a set of non-negative numbers is greater than or equal to its geometric mean.
[f] We use the term visit to indicate that at least one co-location event occurred in a given location.
[g] SnapChat is a self-declared camera-company.

Publisher's Note

Received: 24 September 2018 Accepted: 22 February 2019 Published online: 04 March 2019

References
1. Conti M, Das SK, Bisdikian C, Kumar M, Ni LM, Passarella A, Roussos G, Tröster G, Tsudik G, Zambonelli F (2012) Looking ahead in pervasive computing: challenges and opportunities in the era of cyber-physical convergence. Pervasive Mob Comput 8(1):2–21
2. Utz S, Muscanell N, Khalid C (2015) Snapchat elicits more jealousy than Facebook: a comparison of Snapchat and Facebook use. Cyberpsychol Behav Soc Netw 18(3):141–146
3. Kostakos V, O'Neill E, Penn A, Roussos G, Papadongonas D (2010) Brief encounters: sensing, modeling and visualizing urban mobility and copresence networks. ACM Trans Comput-Hum Interact 17(1):2
4. Wang Z, Zhang D, Zhou X, Yang D, Yu Z, Yu Z (2014) Discovering and profiling overlapping communities in location-based social networks. IEEE Trans Syst Man Cybern Syst 44(4):499–509
5. Grauwin S, Szell M, Sobolevsky S, Hövel P, Simini F, Vanhoof M, Smoreda Z, Barabási A-L, Ratti C (2017) Identifying and modeling the structural discontinuities of human interactions. Sci Rep 7:46677
6. Allen SM, Chorley MJ, Colombo GB, Jaho E, Karaliopoulos M, Stavrakakis I, Whitaker RM (2014) Exploiting user interest similarity and social links for micro-blog forwarding in mobile opportunistic networks. Pervasive Mob Comput 11:106–131
7. Min JK, Cho SB (2011) Mobile human network management and recommendation by probabilistic social mining. IEEE Trans Syst Man Cybern, Part B, Cybern 41(3):761–771
8. Wang R, Gou Q, Choi TM, Liang L (2018) Advertising strategies for mobile platforms with 'apps'. IEEE Trans Syst Man Cybern Syst 48:767–778
9. Yang D, Zhang D, Zheng VW, Yu Z (2015) Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. IEEE Trans Syst Man Cybern Syst 45(1):129–142
10. Atzmueller M (2014) Social behavior in mobile social networks: characterizing links, roles, and communities. In: Mobile social networking: an innovative approach. Springer, Berlin, pp 65–78
11. Quadri C, Zignani M, Capra L, Gaito S, Rossi GP (2014) Multidimensional human dynamics in mobile phone communications. PLoS ONE 9(7):103183
12. Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási A-L (2007) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci 104(18):7332–7336
13. Naboulsi D, Fiore M, Ribot S, Stanica R (2015) Large-scale mobile traffic analysis: a survey. IEEE Commun Surv Tutor 18(1):124–161
14. Calabrese F, Smoreda Z, Blondel VD, Ratti C (2011) Interplay between telecommunications and face-to-face interactions: a study using mobile phone data. PLoS ONE 6(7):20814
15. Wang D, Pedreschi D, Song C, Giannotti F, Barabasi A-L (2011) Human mobility, social ties, and link prediction. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining. KDD'11. ACM, New York, pp 1100–1108
16. Phithakkitnukoon S, Smoreda Z, Olivier P (2012) Socio-geography of human mobility: a study using longitudinal mobile phone data. PLoS ONE 7(6):39253
17. Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. Nature 453(7196):779–782
18. Papandrea M, Jahromi KK, Zignani M, Gaito S, Giordano S, Rossi GP (2016) On the properties of human mobility. Comput Commun 87:19–36
19. Nika A, Ismail A, Zhao BY, Gaito S, Rossi GP, Zheng H (2016) Understanding and predicting data hotspots in cellular networks. Mob Netw Appl 21:402–413
20. Wasserman S, Faust K (1994) Social network analysis: methods and applications, vol 8. Cambridge University Press, Cambridge
21. Abello J, Resende MGC, Sudarsky S (2002) Massive quasi-clique detection. In: Rajsbaum S (ed) LATIN 2002: theoretical informatics: 5th Latin American symposium, 2002 proceedings. Springer, Berlin, pp 598–612

22. Sekara V, Stopczynski A, Lehmann S (2016) Fundamental structures of dynamic social networks. Proc Natl Acad Sci 113(36):9977–9982
23. Nanavati AA, Singh R, Chakraborty D, Dasgupta K, Mukherjea S, Das G, Gurumurthy S, Joshi A (2008) Analyzing the structure and evolution of massive telecom graphs. IEEE Trans Knowl Data Eng 20(5):703–718
24. Blondel VD, Decuyper A, Krings G (2015) A survey of results on mobile phone datasets analysis. EPJ Data Sci 4(1):1
25. Lambiotte R, Blondel VD, De Kerchove C, Huens E, Prieur C, Smoreda Z, Van Dooren P (2008) Geographical dispersal of mobile communication networks. Phys A, Stat Mech Appl 387(21):5317–5325
26. Karsai M, Kaski K, Barabási A-L, Kertész J (2012) Universal features of correlated bursty behaviour. Sci Rep 2:397
27. Li M-X, Palchykov V, Jiang Z-Q, Kaski K, Kertész J, Miccichè S, Tumminello M, Zhou W-X, Mantegna RN (2014) Statistically validated mobile communication networks: the evolution of motifs in European and Chinese data. New J Phys 16(8):083038
28. Zignani M, Quadri C, Bernadinello S, Gaito S, Rossi GP (2014) 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, pp 408–415
29. Xu K, Zhang X (2012) Mining community in mobile social network. Proc Eng (2012 International workshop on information and electronics engineering) 29:3080–3084
30. Li M-X, Xie W-J, Jiang Z-Q, Zhou W-X (2015) Communication cliques in mobile phone calling networks. J Stat Mech Theory Exp 2015(11):11007
31. Zignani M, Quadri C, Gaito S, Rossi GP (2015) Calling, texting, and moving: multidimensional interactions of mobile phone users. Comput Soc Netw 2(1):13
32. Leung IXY, Hui P, Liò P, Crowcroft J (2009) Towards real-time community detection in large networks. Phys Rev E 79:066107
33. Tibély G, Kovanen L, Karsai M, Kaski K, Kertész J, Saramäki J (2011) Communities and beyond: mesoscopic analysis of a large social network with complementary methods. Phys Rev E 83(5):056125
34. Uno T (2010) An efficient algorithm for solving pseudo clique enumeration problem. Algorithmica 56(1):3–16
35. Seufert M, Hoßfeld T, Schwind A, Burger V, Tran-Gia P (2016) Group-based communication in WhatsApp. In: 2016 IFIP networking conference (IFIP networking) and workshops, pp 536–541
36. Calabrese F, Smoreda Z, Blondel VD, Ratti C (2011) Interplay between telecommunications and face-to-face interactions: a study using mobile phone data. PLoS ONE 6(7):20814
37. Backstrom L, Huttenlocher D, Kleinberg J, Lan X (2006) Group formation in large social networks: membership, growth, and evolution. Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining. KDD'06. ACM, New York
38. Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B (2007) Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. IMC'07. ACM, New York
39. Mcauley J, Leskovec J (2014) Discovering social circles in ego networks. ACM Trans Knowl Discov Data 8(1):4
40. Yang J, Leskovec J (2012) Defining and evaluating network communities based on ground-truth. In: 2012 IEEE 12th international conference on data mining (ICDM). IEEE, pp 745–754
41. Song C, Qu Z, Blumm N, Barabási A (2010) Limits of predictability in human mobility. Science 327(5968):1018–1021
42. Zignani M, Gaito S, Rossi GP (2016) Predicting the link strength of newborn links. In: Proceedings of the 25th international conference companion on World Wide Web, International World Wide Web Conferences Steering Committee, pp 147–148
43. Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in Facebook. In: Proceedings of the 2nd ACM workshop on online social networks. WOSN'09. ACM, New York
44. Dunbar R, Arnaboldi V, Conti M, Passarella A (2015) The structure of online social networks mirrors those in the offline world. Soc Netw 43:39–47
45. Miritello G, Moro E, Lara R, Martínez-López R, Belchamber J, Roberts SGB, Dunbar RIM (2013) Time as a limited resource: communication strategy in mobile phone networks. Soc Netw 35(1):89–95
46. Gaito S, Manta G, Quadri C, Rossi GP, Zignani M (2014) Groo-me: handling the dynamics of our sociality on mobile phone. In: Wireless and mobile networking conference (WMNC), 2014 7th IFIP. IEEE, pp 1–4
47. Onnela J-P, Saramäki J, Hyvönen J, Szabó G, de Menezes MA, Kaski K, Barabási A-L, Kertész J (2007) Analysis of a large-scale weighted network of one-to-one human communication. New J Phys 9(6):179
48. Pappalardo L, Simini F, Rinzivillo S, Pedreschi D, Giannotti F, Barabási A-L (2015) Returners and explorers dichotomy in human mobility. Nat Commun 6:8166
49. De Nadai M, Staiano J, Larcher R, Sebe N, Quercia D, Lepri B (2016) The death and life of great Italian cities: a mobile phone data perspective. In: Proceedings of the 25th international conference on World Wide Web. WWW'16. International World Wide Web Conferences Steering Committee, Switzerland, pp 413–423.
50. Dunbar RI, Spoors M (1995) Social networks, support cliques, and kinship. Hum Nat 6(3):273–290