ComeTogether: discovering communities of places in mobility data

Igo Ramalho Brilhante
Federal University of Ceará
Fortaleza, Brazil
Email: igobrilhante@lia.ufc.br

Roberto Trasarti
KDDLAB, ISTI-CNR, Pisa, Italy
Email: roberto.trasarti@isti.cnr.it

Michele Berlingerio
IBM Research & Development
Dublin, Ireland
Email: mberling@ie.ibm.com

Chiara Renso
KDDLAB, ISTI-CNR, Pisa, Italy
Email: chiara.renso@isti.cnr.it

Jose Antonio Fernandes de Macedo
Federal University of Ceará
Fortaleza, Brazil
Email: jose.macedo@lia.ufc.br

Marco Antonio Casanova
PUC-Rio
Rio de Janeiro, Brazil
Email: casanova@inf.puc-rio.br

***Abstract*—We analyse urban mobility and public places under a new perspective: how can we feature the places in a city based on how people move among them? To answer this question we need to combine places, like points of interest, with mobility information like the trajectories of individuals moving within a city. To accomplish this, we propose a methodology based on complex network analysis: we build a network of points of interests by connecting places by the individual trajectories passing through them. From such network we compute communities finding groups places highly connected by the mobility of the individuals. We present a case study on real trajectory dataset on the city of Milan, showing a complementary view on the urban mobility that is not covered by the state-of-the art techniques on mobility analysis.**

I. Introduction

People live in an environment where they move from one place to another. Therefore “places” are not only “static geographical objects” but they are also part of people lives. There is a two-way relationship between how the movements of people are affected by the location of places of interest, and how the places themselves are characterized and connected by the mobility of people. The way people move towards these places and the way they visit these places affects the overall movements, or mobility, of the environment. But the other way around is also true: city places like the Points of Interests (or POIs such as shops, restaurants, banks, hospitals and any locations that can be of interest for individuals) can be featured based on how people globally access them.

In this paper we offer a new perspective on observing how places are connected based on the mobility among them in a urban context. We believe that just counting, for example, the number of visits of a given place, although certainly giving a measure of the attractiveness of that place, is not enough to get a deep understanding on how that particular place is “lived by” people and how this place “relates” to other places. Which is the relationship between two or more places in terms of the mobility that connects them? Do people tend to visit places in the same communities? Are these communities related to mobility issues like traffic congestion or public transportation optimization? Although mobility analysis has become a hot research topic [5], [6], [12], those approaches fail to characterize, at global scale, the relationship among the POIs based on how people access them. The specific aspect of understanding how the objects interact at a global scale is usually associated to the paradigm of complex networks. Finding social interactions at a global scale is in the scope of this research area. A typical example is the study of the spreading of cell phone viruses thru GSM phone calls [10].

The contribution of this paper is twofold: on the one hand, we proposed a methodology for building a complex network combining Points of Interests and traces of people movements, from which we build communities of POIs. On the other hand, we also experimented this methodology in a real case study where trajectories are collected from private cars travelling in a city and Points of Interest are downloaded from the Web. We found different kinds of communities (e.g. *compact* where the movements are mainly inside the community, or *bridge* where the movements tend to connect two other communities). We discuss the possible exploitation of these results in the mobility and advertisement application fields.

This paper is organized as follows: Section III presents the problem definition and the methodology used in this paper. Section IV reports on the experimental results using a real dataset. Section V contains the conclusions and describes future works.

II. Problem Definition and Methodology

In this section we introduce the problem definition and the methodology to build the complex network of POIs. We assume basic knowledge of graphs and complex networks [**?**]. Let us introduce the concept of Point of Interest.

Definition 1. *A Point of Interest (POI) is a geographical object that is interesting for a specific application, usually associated to a human activity. Formally, we define a POI as a triple* $POI = (c, r, l)$ *where* c *is the representative spatial point,* r *is the spatial area representing the extent of the object and* l

is the label of the form cat:n where cat is the category of the POI and n is the POI name.

An example of POI is the Eiffel Tour: the representative spatial point c is the center of the tower, the extent is the area covered by the base of the tower and the label is the category (which can be, for example, "tourist attraction" or "monument" or "tower", depending on the application) and the name "Eiffel Tour"

The starting point of our process is the set of user position observations. Therefore, we define the mobility history of a single user as:

Definition 2 (User Mobility History). *Given a set of user's observations D_u, the user's history is defined as an ordered sequence of spatio-temporal points $H_u = \langle p_1 \dots p_n \rangle$ where $p_i \in D_u$, $p_i = (x_i, y_i, t_i)$, x_i, y_i are spatial coordinates, t_i is an absolute timepoint and $\forall (i, j)\ i \leq j \Rightarrow t_i \leq t_j$ holds.*

Problem Definition Given m traced moving users, a set of Points of Interest (POIs) V and the dataset collecting the users' histories: $D = \{D_1, \dots, D_m\}$, we want to group the POIs in V into groups (or communities) connected by the common mobility of the users.

To solve this problem, we must overcome the limitations of standard methods of grouping locations like spatial clustering, which is based only on the geographical aspect, to move towards a communities perspective where POIs are grouped by the mobility of the users.

The proposed methodology combines different aspects of mobility and graph analysis and it is composed of two main steps: the first step *builds a network* where each link represents the *relations* between two POIs in terms of mobility; the second step *extracts the communities* that identify groups of POIs which share a common mobility context. Furthermore, we define some measures to evaluate and compare the discovered communities. These steps are illustrated below.

A. Building the Network

In literature a network is defined as $G = (V, E)$ in which entities (the nodes in V) are linked by ties (the edges in E), representing any sort of connection, similarity or interaction. The network we propose is composed of a set of nodes which correspond to the set of POIs where the moving users stopped to perform some activity. In order to find these POIs we need to first distinguish the single *trajectory* as the part of the user history representing the movement associated to a specific activity, such as *going to work*, *shopping* etc. In order to distinguish between the different trajectories in a user history, we need to detect when a user stops for a long time so that this stop can be considered the end of that particular trajectory and the beginning of the next one. In the literature there are two main lines of approaches to detect trajectories from a user history: clustering-based [4] and heuristic-based [11]. However, for computational efficiency reasons here we propose a different method as a trade off between precision and efficiency. We search the points that change only in time. i.e. points that stays in the same spatial position for a certain amount of time quantified by the temporal threshold $MinStopTime$. Specularly, a spatial threshold $MaxStopArea$ is used to remove both the noise introduced by the imprecision of the device and the small movements that are of no interest for a particular analysis. These thresholds are used for detecting the *candidate stops* as defined below, where $area()$ is a function computing the size of the minimal convex region including a set of points and $\preceq$ is the operator of sequential inclusion without gaps.

Definition 3 (User's candidates stops). *Given the user history H_u, we define the sequence of candidate stops $S_u = \langle s_1 \dots s_m \rangle : s_k = (a, t, d), a = \langle p_i, \dots p_j \rangle \preceq H_u, area(a) \leq MaxStopArea, t = p_i.t, d = p_j.t - p_i.t \geq MinStopTime.$*

From this set of candidates we want to build the set of user trajectories by removing the cases of slow movements or long stops in a place. Examples of this long stops may be the home and the work places since usually users spend the night at home and the day at work. For this reason, we use a threshold called $MaxMoveTime$ to break the user history into distinct trajectories. This trajectory partitioning step is presented in [12]. Therefore, we define the user trajectories set as follows, where $contains$ is a spatial inclusion predicate between two spatial regions:

Definition 4 (User's trajectories set). *Given a set of Points of Interest V and given the sequence of candidates stops S_u for the user u , we define the user's trajectories set as $T_u = \{t_1, \dots, t_h\}$ where each trajectory is the maximal sequence $t = \langle v_1.l, \dots, v_k.l \rangle : \forall_{(i,j)}, 1 \leq i < j \leq k$, $\exists_{(w,q)}(s_w, s_q) \preceq S_u$, $contains(v_i.r, area(s_w.a)), contains(v_j.r, area(s_q.a))$ and $s_w.t - s_q.t \leq MaxMoveTime$.*

Having all the trajectories of all the users $T = \bigcup_{1 \dots m}^{u} T_u$, we compute the POIs network as:

Definition 5 (Points of Interest (POI) network). *Given a set of POIs V and a set of users trajectories T, we build the points of interest network $P_{oi}^N = (V, E, W)$ where $E = \{e_{i,j} : \exists t \in T, \langle v_i, v_j \rangle \preceq t \vee \langle v_j, v_i \rangle \preceq t\}$ and $W = \{w_{i,j} : w_{i,j} = |\{t_1, \dots, t_m\}|, \langle v_i, v_j \rangle \preceq t \vee \langle v_j, v_i \rangle \preceq t\}$.*

In other words, the POIs network is a undirected weighted graph which summarizes all the trajectories of the users and each edge is weighted by the number of trajectories which share the movement between the same pair of POIs.

An example of the stop computation process is presented in Fig.1. In Figure 1(a) the set of positional observation of a user are shown. The process starts building the user history as a continuous sequence of points ordered by time Fig.1(b). In Fig.1(c) the stops are identified considering $MaxStopArea$ = $50m^2$ and $MinStopTime$ = 30 minutes. Then the stops are spatially intersected with the set of POIs V as shown in Fig.1(d): the red edges between the two stops has a duration which is grater than the $MaxMoveTime$ (e.g. 4 hours) therefore it is removed cutting the user history into two trajectories. Finally, the two trajectories will contribute to the edges shown in Fig.1(e) where w and w' are the number of trajectories which share the same path respectively

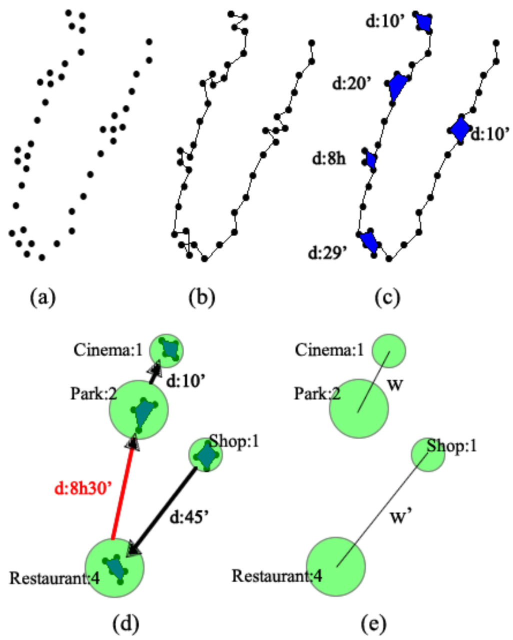

Fig. 1. The building process of places network from one user history: From positional observations in (a) to the user history in (b), the candidates stops in (c). The trajectories set in shown in (d) where a move of duration of 8h30' (thus exceding 4 hrs) splits the user history into two trajectories. The POIs network is depicted in (e)

$Shop : 1 \rightarrow Restaurant : 4$ and $Park : 2 \rightarrow Cinema : 1$.

B. Communities of Points of Interests

A branch of complex network research has been focusing on the discovery of structures called *communities*. Communities are groups of nodes highly interactive, densely connected, or, more in general, highly similar, for a given definition of similarity between any two nodes. Having the POIs network P^N_{oi} we can identify communities of POIs that are grouped based on the movements between them. This can be done using the state-of-the-art algorithm presented in [1] thus obtaining a set of communities $C = \{C_1, \ldots, C_n\}$ where each community is a subgraph of P^N_{oi}. Moreover, in order to evaluate the quality of discovered communities, we introduce three measures: the *Nodes sharing* measuring how similar communities are based on the nodes shared by them; the *Trajectories sharing* giving a measure of how the communities are similar from the point of view of the trajectories which pass through their edges; and the *Compactness* measuring how much the trajectories creating a community move inside the community itself. Formally:

$$Sharing_{Nodes}(C_i, C_j) = \frac{|V_i \cap V_j|}{|V_i|}. \quad (1)$$

$$Sharing_{Traj}(C_i, C_j) = \frac{|T(C_i) \cap T(C_j)|}{|T(C_i)|}. \quad (2)$$

$$Compactness(C_i) = \frac{|E_i|}{|distinct(T(C_i))|}. \quad (3)$$

where $T(C_k)$ is the set of trajectories traversing a community C_k and $distinct(T)$ set of edges traversed by the set of trajectories T.

III. Case Study

In this section we present the experiments carried out using a real trajectory dataset and a set of POI existing in the geographical area of the movements. Furthermore, we analyze and evaluate the generated network according to the analysis and measurements presented in Section III. In our experiments we use a set of positional observations collected by an Italian insurance company which offers a discount to the users who have an embedded GPS device in their car. The set of collected observations in one week in Milan (Italy) is composed by 1,806,293 points for 17,087 users in the Milan area. The POIs dataset of Milan has been downloaded from the web (OpenStreetMap [9]) obtaining a set of 2501 locations corresponding to commonly used POIs semantic categories such as banks, restaurants, cinemas, theaters, museums, etc.

According to the methodology defined in Section III, from the set of positional observations we computed the set of candidate stops for each users. The parameters used are 20 minutes as $MinStopTime$ and 150m^2 as $MaxStopArea$ (i.e. a car with speed less than 0.5 km/h) thus obtaining 216,523 candidate stops. Due to the fact that (1) it is possible to retrieve the representative point of the POIs, but not the precise extent and (2) the observations refer to the position of the car and not the user himself, we use an approximated area around the POIs of 150m. However a single stop matches since the POIs are very close (e.g. an open mall or the city center). To solve this problem we propose to perform a preprocessing step to group together the POIs that match a stop thus defining a *composite POI* represented by the union of all the extents of the close POIs. This has been done with spatial clustering (i.e. T-Optics [2]) and each cluster is handled as a single POI for the purpose of the network construction. The number of POIs after this clustering process is 347 for the single POI and 77 composite POI.

Having the candidates stops computed as described in Section III with the specified thresholds and the set of POIs, we can build the trajectory sets and then the POIs network. For this steps we used a $MaxMoveTime$ of 4 hours obtaining a network with 77 nodes and 677 links corresponding to the movements of the trajectories between the POIs. Figure 2 illustrates the generated POI network in Milan.

A. POI Network Characteristics

Figure 3 shows the distribution of the edge weights representing the number of trajectories. In Figure 3 we observe that the distribution follows a power law, i.e. there are few edges with a large number of trajectories while there is a large number of edges with a small number of trajectories. Intuitively we can conclude that few composite POIs are very popular, having many movements through them.

In order to understand the characteristics of the generated POI network, we computed the following network measures, widely used in standard complex network analysis: clustering coefficient, average shortest path and diameter. The clustering coefficient of P^N_{oi} is 0.329, the average shortest path length

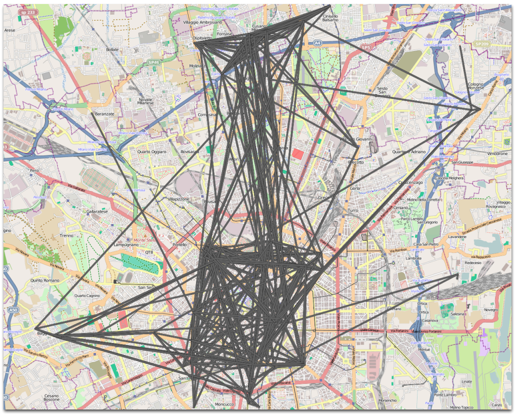

Fig. 2. The plot of the POIs network generated from our experiments with 77 nodes and 677 edges: nodes represent the composite POIs; and edges represent the movement of users' trajectories between the nodes

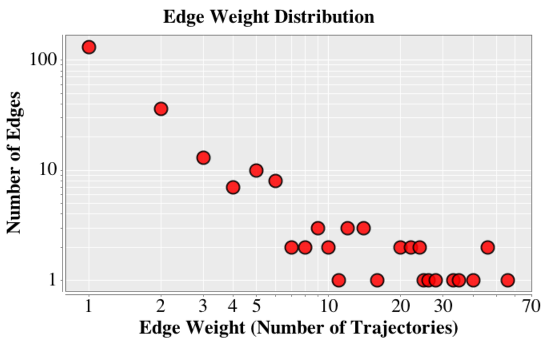

Fig. 3. Edge weight distribution of the network P_{oi}^{N} (Log-Log scale)

is 2.584, the diameter is 7. These results are similar to the ones found in many real world networks, such as biological networks, social networks, and citation networks [8]. They highlights a small world phenomenon where there is a high clustering coefficient and a small average shortest path.

Another interesting analysis is to understand how composite POIs are related, which may help in answering interesting questions such as: do the highly connected composite POIs preferentially connect other high-degree composite POIs? Or do they prefer to connect to low-degree ones? In other words, are we facing degree correlation? Degree correlation is a special case of assortative mixing of node degree [7], [8]. The correlation r is calculated by means of Pearson correlation between of the nodes at either end of the edges. We have found (results not reported here due to lack of space) that in the obtained network, nodes with high degree tend to connect with other nodes with high degree, thus the network looks assortative with a positive correlation of 0.326. This means that the obtained network has typical characteristics of a social network [7], and this is reasoable since this network represents social aspects of the users.

B. Communities Analysis

Discovering communities structures within a complex network is the key for finding tightly connected groups of nodes and thus understanding the functioning of the network [3].

After executing the community detection algorithm we obtained 109 communities from the generated POI network. Figure 4 illustrates P_{oi}^{N} in which the colors of the edges identify each discovered community. Furthermore, Figure 5 shows the community size distribution considering the number of edges. It is worth noting that a number of communities have only one edge while few communities have a large number of edges. Indeed, small communities can be formed by a single movement between two composite POIs while large communities require a large number of distinct composite POIs and movement among them. Moreover, a set of composite POIs that participate in large communities may form small communities among them.

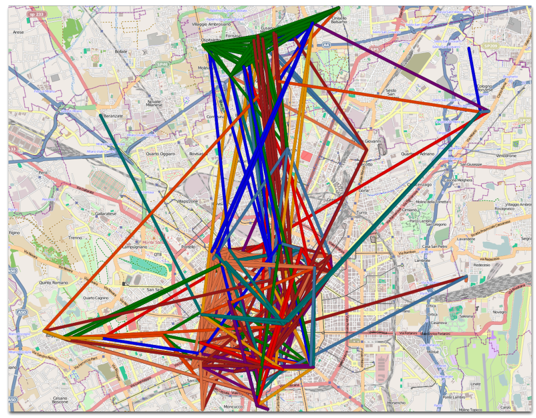

Fig. 4. The 109 communities discovered from the POIs network. The edge color identify the different communities.

Since communities are defined from trajectories, it is important to understand the relationship among them. Do trajectories that define a community tend to have their moves on the edges of that community or they also have some moves in another community? From this question we analyse the community by means of $Compactness$ measure defined in Section III. Figure 6 shows the $Compactness$ cumulative distribution. This plot illustrates these measures in the discovered communities. As we can notice, there is an evident gap at 0.5. It suggests two different behaviors of $Compactness$: one for the interval $(0, 0.5)$ corresponding to less compact communities and other for the interval $(0.5, 1.0)$ related to the communities more compact. Looking into these two groups, we analyze what may influence the $Compactness$ such as number of edges and number of types of places of interest involved into the community. Then, we analyze the correlation between the community properties and the $Compactness$.

For the interval $(0, 0.5)$ we can notice that small communities seem to be less compact, since the trajectories tend to go towards outside the community. In fact, the average

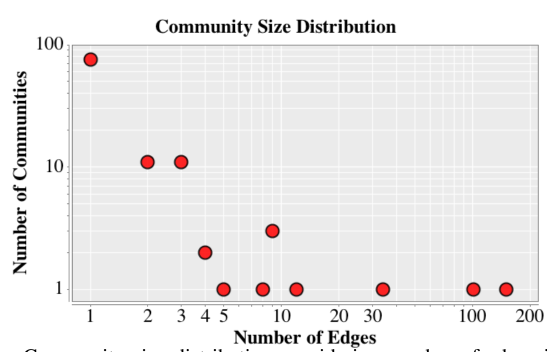

Fig. 5. Community size distribution considering number of edges in Log-Log scale. Many communities are formed by a few edges, whereas a few communities are composed by a higher number of edges.

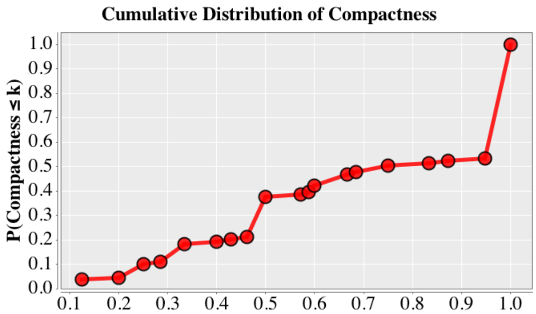

Fig. 6. Cumulative Distribution of $Compactness$. $P(Compactness \leq k)$ indicates the probability that $Compactness$ takes on a value less than or equal to k

community size in this interval is 2.4782 edges. Nonetheless, what may influence the $Compactness$? We have found that the highest correlation of $Compacted$ was with the number of POI types of 0.7874. This means that, for this interval, communities tend not to be so large and that communities with more POI types tend to be more compact. In other words, we could interpret this as the trajectories that tend to keep themselves inside the community since this community present different POI type.

The situation is changed for the interval $(0.5, 1.0)$ since in this case, the communities tend to be larger with an average of 29.4705 edges. There is still a correlation between compactness and the POIs categories since the values is 0.6371. However, the highest correlation found for this interval was between $Compactness$ and the number of edges, corresponding to 0.7257. Therefore, in this case the edges contributes for larger values in the compactness of the community. This means that the communities tend to be larger in number of POIs and, consequently, the trajectories tend to remain inside the community.

As consequence of the previous analysis we discovered two interesting types of communities: (i) the big communities with a large number of edges which cover multiple types of POIs, thus becoming compact, and (ii) the communities which are not large but since they cover several POIs categories they tend to form a compact structure. In the following we focus our attention on these two types of communities.

C. Large Communities

In this section, we focus on the top three larger communities with respect to the number of edges (72, 20 and 25). In Figure 7 we can see how they are distributed in space and how the communities 20 and 25 are interconnected since they share some nodes of the network while community 72 is very well separated. In fact, these three communities highlight how the center of the city is essentially divided into two major communities of POIs. This is confirmed by the table 7(a) and 7(b) where the $Sharing_{nodes}$ is 50% and 28% but the $Sharing_{Traj}$ is only 4% and 1% highlighting that only few users use both the communities. The community 72 is in the outskirts of the city and it describes a new gravitational point for the people activities. From tables 7(a) and 7(b) we can see that it is completely separated from the others considering both nodes and trajectories.

	20	25	72	63	76
20	–	**0.01**	0.00	0.00	0.00
25	**0.04**	–	0.00	**0.02**	0.00
72	0.00	0.00	–	**0.007**	**0.007**
63	0.00	**0.25**	**0.50**	–	0.00
76	0.00	0.00	**0.66**	0.00	–

(a)

	20	25	72	63	76
20	–	**0.28**	0.00	**0.08**	0.00
25	**0.50**	–	0.00	**0.07**	**0.14**
72	0.00	0.00	–	**0.15**	**0.35**
63	**0.40**	**0.20**	**0.60**	–	0.00
76	0.00	**0.22**	**0.77**	0.00	–

(b)

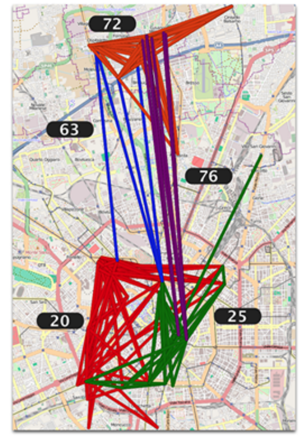

Fig. 7. The selected communities: the three largest communities by the number of edges are 72, 20, 25; and the two communities 76 and 63 act like a "bridge" between them characterizing the movement between two regions of the city. (a) $Sharing_{Nodes}$ between the communities; (b) $Sharing_{Traj}$ between the communities.

Looking at the other communities discovered in Figure 4 we note that there exists *bridge* communities, which connect this three large communities. Intuitively, a *bridge* community C_b is a community that shares nodes with two other communities C_i and C_j such that C_i and C_j do not share nodes with each other. Including them in the analysis (communities 63 and 76 in Figure 7) we can understand how they connect the center of the city with the peripheral area. We computed the sharing measure and we discovered that they share a large percentage of nodes and trajectories among them.

The results of such analysis could very valuable for a broad range of urban actors such as mobility managers or advertising companies, which can understand the dynamics and the interconnections of the city and be more accurate in their actions. For example, advertising companies could use this information to understand where to locate their posters to optimize the spreading of the information to all the three major communities exploiting, for example, of the *bridges* communities. Hence, considering the $Sharing_{Traj}$ and $Sharing_{Nodes}$ between the large communities and the bridges, the best places to put the posters are the shared nodes between community 20 and 25 and the shared nodes between these two communities and the *bridges*. Indeed, although communities 76 and 63 do not influence the central area, they share a large number of trajectories with community 72.

In the case of a mobility (or traffic) agency further information can be obtained by analyzing the temporal aspect of the communities. In fact, the communities change over time as shown in Figure 8. The communities are analyzed and compared to the entire network usage in terms of number of users who stop in a POI and the number of users who move between POIs is reported comparing them. The temporal analysis is performed using two different granularities, *days* and *hours*. We can observe how each community has its own distribution which follows the general behavior of the network. However, there are some specific periods where they clearly diverge. This result can be used by the mobility agency to better organize public security. For example, considering Figure 8(d), traffic agents could be allocated in community 72 at 16:00h (high movement) in order to guard movement among

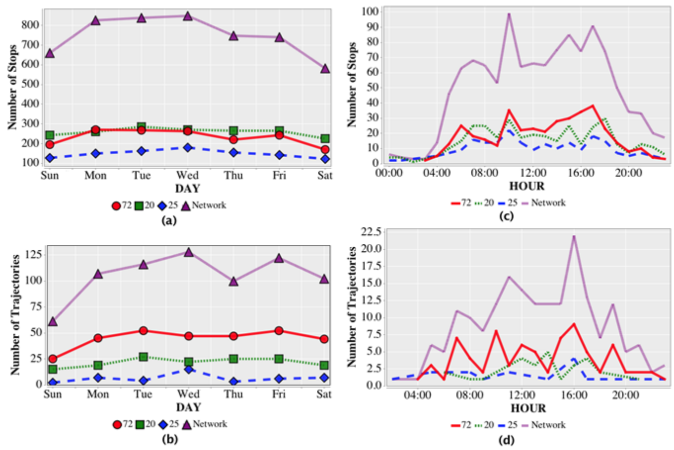

Fig. 8. (a) Stops in the week (b) Trajectories in all the week (c) Stops on Wednesday (d) Trajectories on Wednesday

community POIs. It's important to notice how a simple spatial clustering is not sufficient to obtain this result since these kinds of algorithms tend to partition the space (and therefore the groups of POIs) not considering the mobility information.

D. Compact Communities

In this section we focus on *compact* communities, which are characterized by containing trajectories that tend to remain inside the community. As we have shown and discussed section IV-B, there is an high correlation between the number of POI types and $compactness$. Here we focus on six compact communities selected from the intervals of compactness that have been discussed in Figure 6. We discover again the communities number 20 and 72 are among the most compact ones, but not the 25.

Focusing on the other communities in Figure 9 (i.e. 13, 6, 43 and 86), we can notice that some of them seem to be similar to the larger communities: community 86 is high related to the community 25, sharing a large percentage of nodes and trajectories (respectively 77% and 40%). Moreover, if we consider the topology of the community, this suggests the presence of a central core connecting almost all the POIs of the large community. A different relation exists between the communities 6 and 20 where the percentage of shared nodes is 90% but the shared trajectories is 0%: this means that it represents a different community which uses the same POIs. This observation highlights the complexity of the mobility in a city and the method discussed in this paper is a further step in trying to get an understanding of the phenomena.

IV. Conclusion

In this paper, we proposed an explorative study on the relation between people mobility and Points of Interest at the global scale, based on the complex network paradigm. We presented an algorithm to build a complex network that combines locations that people visit with the mobility of users represented as trajectories. From this network we computed the communities as the subgroups of Points of Interest related by the common users trajectories visiting them. We defined some interesting features of these communities such as the compactness or the presence of "bridge" communities. We observed

Community C_i	$Compactness(C_i)$
86	0.42
43	0.46
25	0.58
6	0.68
20	0.87
72	0.94

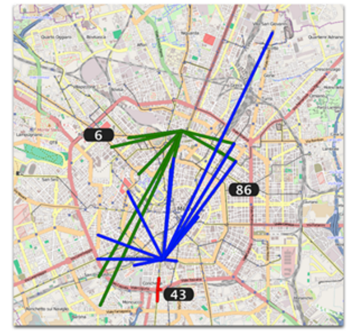

Fig. 9. Communities to illustrate the measure $Compactness$ considering different degrees of $compactness$: communities 104, 86 and 43 are less compact communities; 6 and 13 are more compact communities. The table contains the Compactness values of the new communities with the ones in Figures 7.

these measures discussing the possible interpretations in terms of applications such as traffic management or advertising.

Since this is a preliminary work, we envisage several future works. First of all, alternative ways of computing the stops and associating the POIs may be applied in order to better represent the actual activity of the user. Furthermore, we plan to extend this methodology to other real datasets to further validate the results. Naturally, we intend to investigate more in deep the possible applications that can be benefit from this analysis, for example going to the direction of POIs recommendation systems.

References

[1] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466, 2010.

[2] G. L. Andrienko, N. V. Andrienko, S. Rinzivillo, M. Nanni, and Dino Pedreschi:. A visual analytics toolkit for cluster-based classification of mobility data. In *SSTD 2009*, pages 432–435, 2009.

[3] S. Boccaletti, V. Latora, Y.Moreno, M. Chaves, and D.-U. Hwang. Complex networks: structure and dynamics. *Physics Reports*, (424):175–308, 2006.

[4] V. Bogorny, B. Kuijpers, J. Antônio Fernandes de Macêdo, B. Moelans, and A. A. Vaisman. A model for enriching trajectories with semantic geographical information. In *ACM-GIS*, 2007.

[5] F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, C. Renso, S. Rinzivillo, and R. Trasarti. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *Very Large Database*, 20(5), 2011.

[6] F. Giannotti and D. Pedreschi, editors. *Mobility, Data Mining and Privacy - Geographic Knowledge Discovery*. Springer, 2008.

[7] M E J Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20):5, 2002.

[8] M E J Newman. The Structure and Function of Complex Networks. *SIAM Review*, 45(2):167, 2003.

[9] OpenStreetMap. Openstreetmap, 2011.

[10] P. Wang, M.C. González, C.A. Hidalgo, and A.-L.Barabási. Understanding the spreading patterns of mobile phones viruses. *Science*, (324):1071–1076, 2009.

[11] J. Xiao. Clustering spatial data for join operations using match-based partition. In *CIMCA/IAWTIC*, 2005.

[12] Y. Zheng and X. Xie. Learning Location Correlation from GPS Trajectories. *2010 Eleventh International Conference on Mobile Data Management*, (49):27–32, 2010.