Ad Hoc Networks 10 (2012) 1520–1531

Contents lists available at SciVerse ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

Human-mobility enabled networks in urban environments: Is there any (mobile wireless) small world out there?

Angela Sara Cacciapuoti [a,1], Francesco Calabrese [b,2], Marcello Caleffi [a,*,1], Giusy Di Lorenzo [b,2], Luigi Paura [a,c,1]

[a] Department of Biomedical, Electronics and Telecommunications Engineering (DIBET), University of Naples Federico II, Naples, Italy
[b] IBM Research IBM Technology Campus, Damastown, Dublin 15, Mulhuddart, Ireland
[c] Laboratorio Nazionale di Comunicazioni Multimediali (CNIT), Naples, Italy

ARTICLE INFO

Article history:
Received 10 July 2010
Received in revised form 11 July 2011
Accepted 12 July 2011
Available online 30 July 2011

Keywords:
Measurement
Human
Mobility
Connectivity
Pocket switched networks
Delay tolerant networks
Small world

ABSTRACT

In the last 10 years, new paradigms for wireless networks based on human mobility have gained the attention of the research community. These paradigms, usually referred to as *Pocket Switched Networks* or *Delay Tolerant Networks*, jointly exploit human mobility and store-and-forward communications to improve the connectivity in sparse or isolated networks. Clearly, understanding the human mobility patterns is a key challenge for the design of routing protocols based on such paradigms. To this aim, we anonymously collected the positions of almost two thousand mobile phone users, spread over a metropolitan area greater than 200 km^2 for roughly one month. Then, with a multi-disciplinary approach, we estimated the mobility patterns from the collected data and, assuming Wi-Fi connectivity, we inferred the contact events among the devices to evaluate the connectivity properties of a human mobility-enabled wireless network. In a nutshell, the contribution of the paper is threefold: (i) it confirms some of the results obtained in smaller environments, such as the power-law distribution for contact and inter-contact times, allowing us to estimate the distribution parameters with high statistical significance; (ii) it addresses the feasibility of the transmission opportunities provided by human mobility to build a city-wide connected network for different forwarding strategies classes; (iii) it shows uncovered characteristics of the connectivity properties of human mobility, such as the presence of the *small world phenomenon* in wide-scale experiments.

1. Introduction

In the last 10 years, new paradigms for wireless networks based on human mobility have gained the attention of the research community.

These paradigms, usually referred to as *Pocket Switched Networks* [1] or *Delay Tolerant Networks* [2,3], assume that at each instant end-to-end paths may not exist in the network as a consequence of intermittent connectivity, low node density or the presence of isolated regions. Examples of these networks include those used in infrastructure-challenged environments found in developing countries, but also those used in developed countries when communication infrastructures are absent.

* Corresponding author. Address: Dipartimento di Ingegneria Biomedica, Elettronica e delle Telecomunicazioni (DIBET), via Claudio 21, I-80125 Napoli, Italy.

E-mail addresses: angelasara.cacciapuoti@unina.it (A.S. Cacciapuoti), fcalabre@ie.ibm.com (F. Calabrese), marcello.caleffi@unina.it (M. Caleffi), giusydil@ie.ibm.com (G. Di Lorenzo), paura@unina.it (L. Paura).

[1] This work is partially supported by the project "Global & Reliable End to End e-Commerce & On Line Service Platform (GRECO)" founded by the Italian National Program *Industria 2015* and by the projects "HarBour traffIc opTimizAtion sysTem (HABITAT)" and "DRIVEr monitoring: technologies, methodologies, and IN-vehicle INnovative systems for a safe and eco-compatible driving (DRIVE IN^2)" founded by the Italian National Program *Piano Operativo Nazionale Ricerca e Competitivitá 2007–2013*.

[2] This work was partially conducted while at the Department of Urban Studies and Planning, MIT.

1570-8705/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.adhoc.2011.07.006

In such paradigms, the end-to-end connectivity is thus provided by device mobility by adopting store-and-forward communications: intermediate devices store the data by waiting for transmission opportunities provided by mobility to deliver the data towards the final destination.

For many years the researchers have assumed that the mobility of the devices forming such networks was totally unpredictable. In reality, this assumption is unrealistic since devices are used by people, whose mobility patterns usually depend on the user habits [4]. Therefore, in order to better understand the opportunities offered by human mobility to enable end-to-end communications in sparse or isolated networks, we anonymously collected the positions of almost two thousand mobile phones for roughly one month. Such positions are spread over a region greater than 200 km^2 in east Massachusetts. Based on these traces, we first estimated the mobility patterns of the devices and then, assuming Wi-Fi connectivity, we inferred the contact events among the devices used in the subsequent connectivity analysis.

At first, we studied the distributions of two commonly adopted metrics for measuring human mobility, namely the contact and the inter-contact times. These metrics measure how long and how often two devices come in contact. We observed that the distributions of both the contact and the inter-contact times in a real-world wide-scale network exhibit a power-law distribution.

Moreover, we studied the impact of the human mobility in terms of hop distance and time distance, namely *separation degree* and *separation time*, for two different classes of forwarding strategy: the shortest-path and the shortest-delay. Based on our experimental observations, we deduced the presence of the *small world phenomenon* [5] in the network. Clearly, we do not claim that this observation can be generalized for every kind of human mobility-enabled wireless network. Rather, we claim that it confirms the possibility of setting up a connected network by exploiting human mobility outside the infrastructure boundaries.

Finally, we studied the distribution of the 'importance' of the devices in terms of network connectivity by accounting how frequently a device enables the connectivity between two devices.

The paper is organized as follows. Section 2 presents the related work, while in Section 3 we introduce the experimental datasets used in the paper. In Section 4 we analyze the results provided by the experiment and, finally, in Section 5 we conclude the paper with a brief discussion and suggested future work.

2. Related work

Several experiments have been conducted in the past 10 years aiming to collect human mobility data and to evaluate its impact on forwarding algorithms.

A first small experiment collecting human mobility traces in conference environment was conducted in 2005 [1]. Following this, several studies have been focusing on comparing different datasets and studying the impact of the derived human mobility patterns on the design of opportunistic forwarding algorithms [6–8]. Moreover, different temporal distance metrics able to quantify the speed of information diffusion processes have been proposed, see for instance [9].

However, none of the aforementioned studies has considered real wide-scale scenarios such as metropolitan areas. Only in [10], the authors have studied an urban setting (Cambridge, UK) but have relied on a rather small group of students.

In this paper, we instead evaluate the characteristics of human mobility: (i) at large spatial scales; (ii) not biased on involving volunteers, as traces were anonymously collected from mobile phone users of a major telecom operator. To the best of our knowledge, this is the first work in which a real-world wide-scale data set has been considered for understanding the impact of human mobility on network connectivity.

3. Measuring human mobility

As mentioned in Section 1, human mobility is an enabling factor for inducing connectivity in Delay Tolerant Networks. In order to explore the human mobility properties from a networking point of view, we conducted a data collection experiment involving one million mobile phone users of a US telecom operator in the Boston Metropolitan area, whose positions have been anonymously traced for roughly one month. The large user set allows us to limit biases in the obtained results, which are instead found in previous datasets where participation was restricted to students or conference attendees who volunteered to be tracked (see for instance [6]).

In this section, we first present the procedure adopted for measuring real world human positions. Then we describe how the mobility patterns have been inferred from the position traces. Finally, we illustrate how the contact events among the humans have been inferred based on the mobility patterns.

3.1. Tracing human positions

We start from an initial set M of 200 millions anonymous location measurements of one million mobile phones collected between October 1st and October 29th, 2009, and covering a region spread over 8 counties in east Massachusetts (Middlesex, Suffolk, Essex, Worcester, Norfolk, Bristol, Plymouth, Barnstable) with a population of 5.5 million people [11]. The location measurements have been generated each time a device connects to the cellular network, including:

- when a call is placed or received (both at the beginning and end of a call);
- when a short message is sent or received;
- when the device connects to the Internet (e.g. to browse the web, or through email programs that periodically check the mail server).

Each location measurement $m_i(t_i) \in M$ represents the position, i.e., the pair latitude and longitude, of the device

i estimated at time t_i through triangulation. The location estimation process has a greater uncertainty range than the one associated with GPS data, with an average absolute value of 320 m and median absolute value of 220 m [11]. Moreover, some peak errors appear when the user is connected to the network not using the closest cell phone tower. To overcome these issues, the data has been low-pass filtered, resampled every 10 min, following the approach proposed and evaluated in Rome [12,13].

3.2. Estimating mobility patterns

To infer human mobility patterns, we randomly extracted from the dataset M the location measurements of 1915 users living in Suffolk county, who make at least 100 cellular network connections per day and with individual inter-event time below 1 h in 75% of the cases. We selected the Suffolk county as it contains the densely populated Boston downtown area. As the selected area is relatively large (about 152 km^2) and contains both residential and work areas, we believe not to introduce significative biases. Studying the effect of the size of the monitored area on the number and nature contact events would be part of future work.

The selection of the users has a twofold purpose: (i) to limit the computational complexity of the data analysis, which becomes unfeasible for greater numbers of selected users; (ii) to improve the statistical meaning of the analysis results, by selecting people for which a sufficient number of location measurements is available. We note that such a user selection, based on the number of available location measurements, does not introduce bias in our results, since it does not account for the connectivity properties of the selected users.

The raw location measurements are then processed using the methodology described below to obtain traces with sampling rate of 10 min. In particular, since localization errors might generate fictitious trips, we adopted a pre-processing step as follows:

– We select the sequences of consecutive measurements $M_i^{q,z} = \{m_i(t_q), m_i(t_{q+1}), \ldots, m_i(t_z)\}$,

where the user i makes cellular network connections over a certain time interval, $t_z - t_q > 0$, into an area within the radius ΔS, i.e.

$$d(m_i(t_r), m_i(t_s)) < \Delta S \qquad \forall\, q \leqslant r, \quad s \leqslant z,$$

where $d(\cdot,\cdot)$ is the spatial (2-norm) distance and the spatial threshold ΔS has been defined as 1 km to take into account the localization errors.

– The locations $M_i^{q,z}$ are replaced by the corresponding centroid

$$C_i^{r,q} = \frac{1}{z-q}\sum_{t=t_q}^{t=t_z} m_i(t) \quad \forall\, q \leqslant r \leqslant z.$$

– The virtual locations become the origin or destination of a device movement.

Further details and statistics about the location measurements are given in [11].

3.3. Inferring contact events

To infer the contact events from the mobility patterns, we need to estimate the transmission opportunities among the mobile phones. Obviously, the transmission opportunities strictly depend on the adopted wireless technology. In this work, we assumed that each mobile phone is Wi-Fi enabled and so we consider ad hoc Wi-Fi transmission opportunities. This assumption is justified by the latest market research analyses, which estimate that 144 million Wi-Fi enabled mobile phones have been worldwide shipped in 2009 [14], and it predicts that the Wi-Fi enabled phones penetration would quadruple by 2015, reaching the 66% of all mobile phones shipments [15].

To take into account the Wi-Fi technology properties, we assumed a transmission range of 50 m for inferring the contact events. This value is reasonable since the effective indoor transmission ranges of the different Wi-Fi standards lie in the range [30–70] m [16]. According to this, two users experience a transmission opportunity in a certain time slot if their mutual distance is not greater than 50 m.

Formally, we model the network as a *temporal graph* [17]:

$$G(V, E, T_0, T_N, \tau) = \{G(V, E(T_0)), \ldots, G(V, E(T_0 + k\tau)), \ldots, G(V, E(T_N))\}, \tag{1}$$

where τ is the time slot and $G(V,E(T_0 + k\tau))$ is an un-direct graph in which the vertex $v_i \in V$ represents the ith mobile phone and the edge $e_{ij}(k) \in E(T_0 + k\tau)$ represents a communication link between nodes v_i and v_j observed in the time slot $[T_0 + k\tau, T_0 + (k+1)\tau)$.

The ith and jth users experience a transmission opportunity in time slot $[T_0 + k\tau, T_0 + (k+1)\tau)$, namely, $e_{i,j}(k) = 1$, if:

$$d(m_i(T_0 + k\tau), m_j(T_0 + k\tau)) < 50, \tag{2}$$

where $m_i(T_0 + k\tau)$ and $m_j(T_0 + k\tau)$ are the location measurements at time $T_0 + k\tau$, and $d(\cdot,\cdot)$ is the spatial (2-norm) distance.[3]

Two issues arise as a consequence of the assumed transmission range. The first is related to the localization error affecting the mobility traces, which can be greater than the transmission range. However, since the localization error of the measurements can be assumed independent and identically distributed with zero mean, and since we estimated the contact events by using a data set of eighty millions location measurements, no biases are introduced in the statistics of inferred contact events.

This consideration is confirmed by the results of a numerical simulation of contact events between pairs of nodes randomly placed in a 36 km^2 square area.[4] Based on the maximum transmission range of 50 m, we computed the total number of contact events between the pairs of nodes (noise-free contact events) and we compared it with the total

[3] Simulated contact event dataset with the same characteristics of the one used in this paper is available for download at http://wpage.unina.it/marcello.caleffi.

[4] The side of the square has been chosen to be greater than twice the maximum localization error seen in the data to limit the border effects in the simulation.

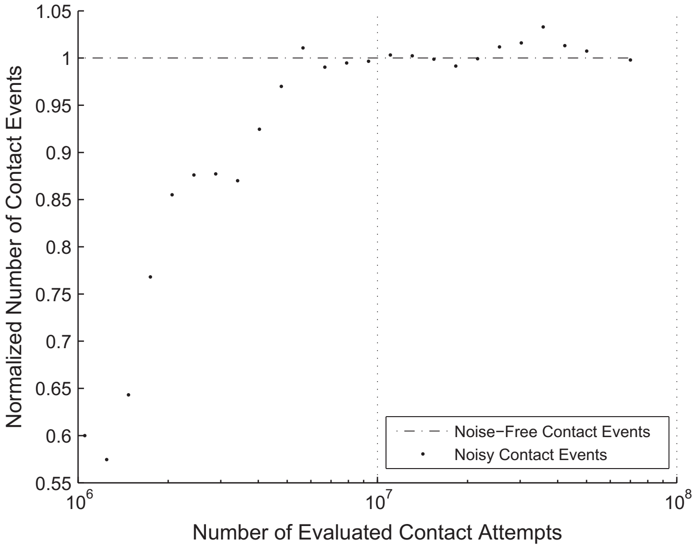

Fig. 1. Number of noisy contact events normalized to the number of noisy-free contact events, as a function of the number of evaluated contact attempts.

number of contact events if the positions of the nodes were effected by a localization error equivalent to the one found in the location measurements (noisy contact events).

Fig. 1 shows the total number of noisy contact events normalized to the total number of noise-free contact events as a function of the number of evaluated contact attempts. We observe that, as the number of evaluated contact attempts increases over 5 million, the two number of contact events are statistically equivalent. This condition is largely verified in our experiment, implying that the statistics on the inferred contact events can be assumed as not biased by the localization error.

The latter issue concerns the assumption of a deterministic transmission range, which does not allow us to account for the wireless propagations effects, such as link asymmetry and fading effects. Moreover, we have not taken into account further factors which affect contact events such as: (i) transmission data rate; (ii) sensitivity and transmission power of the wireless card along with the power management and the battery level of the device; (iii) user behavior, such as turning the mobile phone off or be not collaborative from a forwarding point of view. Despite this, the inferred contact events are a valuable source of real world human contact events spanning one month over a wide region and including almost two thousand devices. In addition, since we are interested in gaining a better understanding of the connectivity properties of a human mobility-based network, neglecting some limiting factors (such as user selfishness and device power management strategies) allows us to consider the best-case scenario for store-and-forward communications.

4. Experimental results

In this section we analyze and discuss the results obtained from the experiment, referred to as *GreaterBoston*, in terms of network connectivity as a consequence of the human mobility.

For the sake of comparison, we use publicly available traces measuring direct contact events between Bluetooth devices carried by a subset of the attendees at the InfoCom 2005 conference held in 2005 in Miami [1]. We note that the two datasets are different for at least two reasons, as shown by Table 1: (i) the methods used for collecting these traces are different; (ii) the social links among the InfoCom 2005 attendees are stronger than those that we expect among the people involved in GreaterBoston experiment, since all of the attendees move in the same building for most of the time (some of them are also members of the same research group). Nevertheless, as no citywide comparable dataset was available, we decided to still compare our results with available traces, and choose InfoCom 2005 as they have been widely used in the existing literature. By comparing the results, we can highlight some differences between the human mobility patterns in small-scale and wide-scale experiments.

In Sections 4.1 and 4.2, we analyze the duration and the frequency of the contact events. In Section 4.3, we investigate the consequences of the contact events in terms of network connectivity for two different store-and-forward strategies classes: the *shorthest-path* and the *shortest-delay*. This allows us to verify the presence of the *small world phenomenon* in human-enabled wireless networks when a shortest-path based strategy is adopted. Finally, in Section 4.4 we analyze the degree of interaction of the devices for

Table 1
Comparison of real world traces.

Experiment	InfoCom 2005	GreaterBoston
Participants	41	1915
Region	Conference	200 km^2
Devices	iMote	Cell phone
Wireless technology	Bluetooth	Wi-Fi
Contact type	Direct	Inferred
Duration	4 days	29 days
Granularity	120 s	<10 min

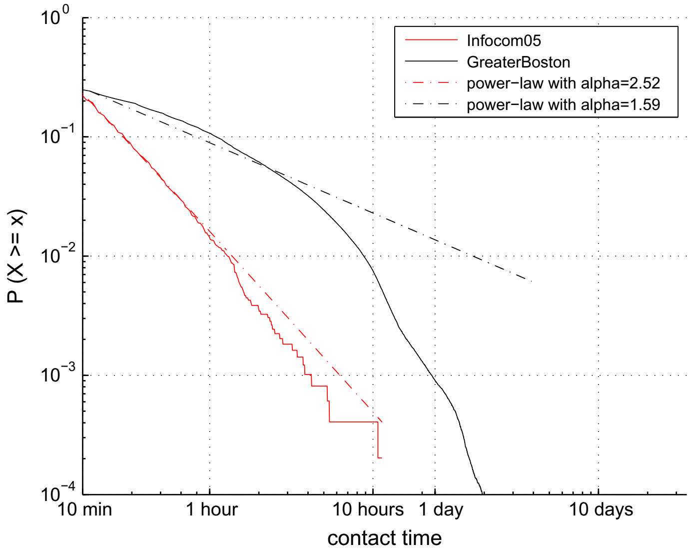

Fig. 2. Contact times complementary cumulative distribution function.

the considered forwarding schemes and discuss its impact on the design of networking protocols.

4.1. Contact and inter-contact times characterization

Given two devices v_i and v_j, we define the *contact time* $T_c(i,j)$ as the time interval in which the devices i and j experience a steady transmission opportunity, that is

$$T_c(i,j) = k\tau \iff \exists\, \tilde{n} : e_{i,j}(n) = \begin{cases} 1 & \text{if } n \in [\tilde{n}, \tilde{n}+k) \\ 0 & \text{if } n = \tilde{n}-1 \text{ or } n = \tilde{n}+k \end{cases}. \tag{3}$$

The complementary cumulative distribution functions of the contact times for both InfoCom 2005 and GreaterBoston experiments are given in Fig. 2. We observe that the distributions of both the experiments follow an approximate power-law distribution, whose exponent[5] α is indicated in the figure. Although the similar shape of the experimental curves suggests the presence of a similar distribution for the contact times in the two experiments, the different exponent values suggest that the probability of two nodes being connected for at least a certain amount of time in GreaterBoston experiment is significantly higher than the corresponding probability in the InfoCom 2005 experiment. In particular, in the time interval [1 h, 10 h] the two probabilities differ for roughly an order of magnitude. Moreover, we note that the contact time distribution of GreaterBoston experiment exhibits a smooth behavior in the time interval [10 min, 10 h], while it quickly falls out for values greater than one day. This behavior is consistent with the usual people working time.

Given a device v_i, we define the *any contact time* $T_{ac}(i)$ as the time interval in which the device experiences a steady transmission opportunity with at least another device:

$$T_{ac}(i) = k\tau \iff \exists\tilde{n} : \begin{cases} \exists v_j \in V : e_{i,j}(n) = 1 & \text{if } n \in [\tilde{n}, \tilde{n}+k) \\ \forall v_j \in V : e_{i,j}(n) = 0 & \text{if } n = \tilde{n}-1 \text{ or } n = \tilde{n}+k \end{cases}. \tag{4}$$

The complementary cumulative distribution functions of the any contact times are given in Fig. 3. We observe that, although the distributions of both the experiments follow an approximate power law distribution in the time interval [10 min, 1 h], the GreaterBoston approximating exponent is lower than the corresponding Infocom 2005 exponent. Despite this, there is a good agreement between the shapes of the two distributions in the time interval [30 min, 10 h], i.e. in the region of the distribution fall-off. By comparing these results with those in terms of contact-time distribution (Fig. 2), we can observe that, although GreaterBoston experiment is characterized by longer expected contact times between a pair of devices with respect to Infocom 2005, the any-contact time distributions exhibit similar behaviors.

We define the *inter-contact time* $T_i(i,j)$ as the time elapsed between two subsequent contact times:

$$T_i(i,j) = k\tau \iff \exists\tilde{n} : e_{i,j}(n) = \begin{cases} 0 & \text{if } n \in [\tilde{n}, \tilde{n}+k) \\ 1 & \text{if } n = \tilde{n}-1 \text{ or } n = \tilde{n}+k \end{cases}. \tag{5}$$

Fig. 4 shows the complementary cumulative distribution functions of the inter-contact times. We observe that the distributions for both the experiments follow an approximate power-law distribution over a long time interval ([10 min, 1 day]). Moreover, both the distributions exhibit a fall-off for long time intervals due to the duration of the experiments, which limit the inter-contact time values. However, the GreaterBoston curve exhibits a smoother fall-off with respect to InfoCom 2005 due to the larger

[5] We note that, with reference to InfoCom 2005 experiment, the difference in the exponent value with respect to that estimated in [1] is due to a different definition of the contact times. This note holds also in the following figures.

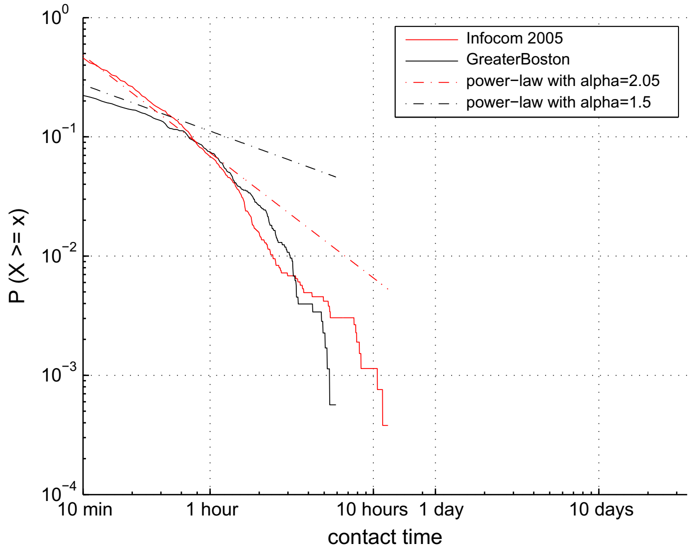

Fig. 3. Any contact times complementary cumulative distribution function.

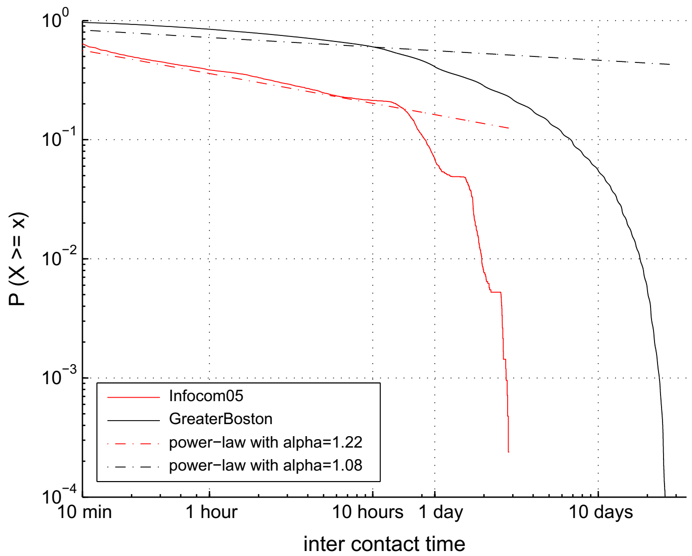

Fig. 4. Inter-contact times complementary cumulative distribution function.

available data set, which improves the statistical accuracy. Although the similarity in the shape of the experimental curves, the different exponent values suggest that the probability of two nodes being not connected for at least a certain amount of time in GreaterBoston experiment is significantly higher than the corresponding probability in the InfoCom 2005 experiment. In particular, we have that the probability of two devices being not connected for a time interval in the order of hours is very high, about 0.8. Nevertheless, the probability of the two devices being not connected for a time interval in the order of a week is significantly lower, about an order of magnitude. This behavior shows the presence of a periodicity in the mobility patterns of GreaterBoston experiment on a time scale of days or a week, which is intuitevely consistent with the human mobility periodicity.

We define the *any inter-contact time* $T_{ai}(i)$ as the time elapsed between two subsequent *any contact times* of the device:

$$T_{ai}(i) = k\tau \Longleftrightarrow \exists \tilde{n} : \begin{cases} \forall v_j \in V : e_{i,j}(n) = 0 & \text{if } n \in [\tilde{n}, \tilde{n}+k) \\ \exists v_j \in V : e_{i,j}(n) = 1 & \text{if } n = \tilde{n}-1 \text{ or } n = \tilde{n}+k \end{cases} \quad (6)$$

Fig. 5 shows the complementary cumulative distribution functions of the *any inter contact times*. We observe that

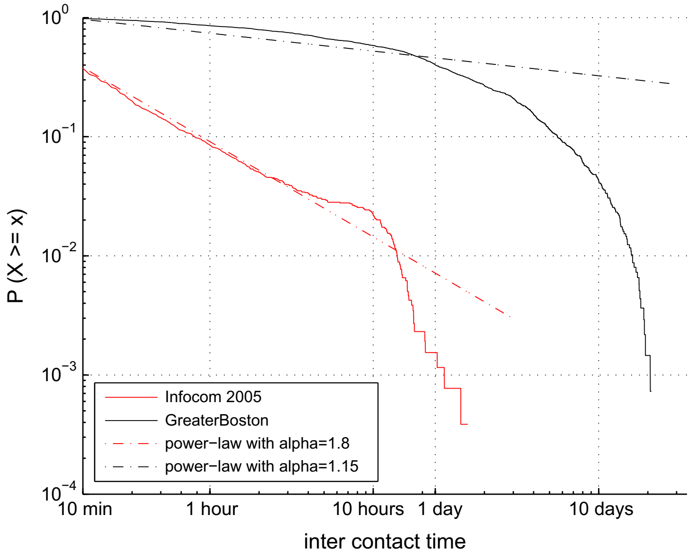

Fig. 5. Any inter-contact times complementary cumulative distribution function.

in GreaterBoston experiment the probability of a device being disconnected from any other device for a certain time interval is significantly higher than the corresponding probability of InfoCom 2005 experiment. This behavior is reasonable, since during InfoCom 2005 the people mobility was bounded by the conference location (hotel) and there was an high probability to meet at least another conference attendee in less than a day due to the social events. On the other hand, in GreaterBoston experiment people spread over a wide region without any global meeting location. Moreover, by comparing the distribution of the any inter contact times with the distribution of the inter contact times (Fig. 4), we have that the two distributions are quite similar for GreaterBoston, while they differ significantly in InfoCom 2005. As a consequence, we have that the probability of a device being connected with more than one device in the same time slot in GreaterBoston is lower than the correspondent probability for InfoCom 2005. This behavior is reasonable if we account for the differences in the mobility patterns of the two experiments.

Finally, if we analyze the results for the contact and inter-contact times with a networking prospective, we can highlight two key characteristics of GreaterBoston network that represent a sort of trade-off. In fact, although we can expect contact times long enough to allow the devices to exchange a significant amount of data (Fig. 2), we should also expect very long inter-contact times (Fig. 4) between the devices, comparable with the disconnected times (Fig. 5). These two aspects should be carefully considered in the design of the spreading algorithms for human mobility-enabled networks, although it is worthwhile to underline that the results reported here account for the connectivity of a wide-scale network formed by less then two thousand people. We can not predict which would be the results if we take into account the whole population, but it is reasonable to expect average inter-contact and any inter-contact times significantly shorter.

4.2. Contact events characterization

The number of distinct contact times experienced by devices v_i and v_j is defined as the *contact event number* $N_c(i,j)$[6]

$$N_c(i,j) = |\mathcal{N}|, \quad \text{where } \mathcal{N} = \{n : e_{i,j}(n-1) = 0 \quad \text{and } e_{i,j}(n) = 1\}. \tag{7}$$

Fig. 6 shows the complementary cumulative distribution functions of the number of contact events. We observe that the GreaterBoston distribution follows an approximate power law distribution for almost all the contact event values, while the Infocom 2005 distribution can be approximated by a power-law distribution only for small values of the contact events. Moreover, the approximating exponent for GreaterBoston is considerably lower than the corresponding exponent for Infocom 2005, with a difference between the two probabilities higher than an order of magnitude. Clearly, since the contact event distribution of InfoCom 2005 is strongly biased by the particular people mobility involved in that experiment, a comparison between the two distribution is far from be fair. Nevertheless, despite the low number of people involved and the wide covered area, the probability that two nodes meet each other is far from being negligible: more than 15% and less than 10% of node pairs meet each other at least once and twice, respectively.

The number of distinct any contact times experienced by node i is defined as the *any contact event number* $N_{ac}(i)$:

$$N_{ac}(i) = |\mathcal{N}|, \quad \text{where}, \\ \mathcal{N} = \{n : e_{i,j}(n-1) = 0 \; \forall v_j \in V \quad \text{and} \quad \exists j \in V : e_{i,j}(n) = 1\}. \tag{8}$$

[6] We denote with $|X|$ the cardinality of the set X.

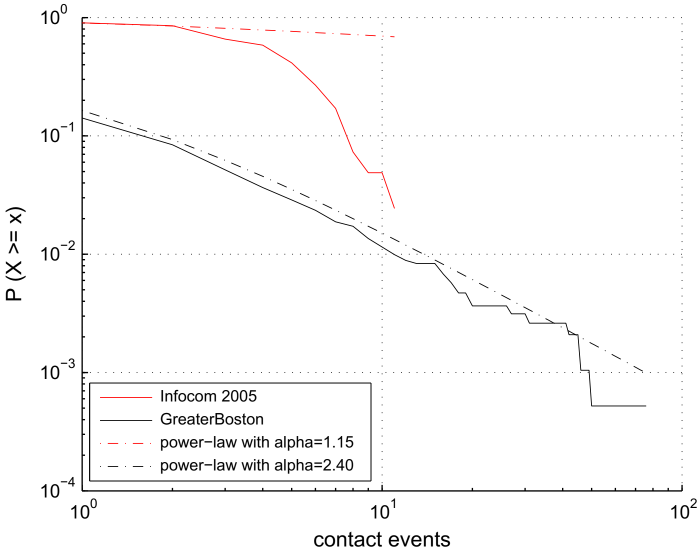

Fig. 6. Contact event number complementary cumulative distribution function.

Fig. 7 shows the complementary cumulative distribution functions of the any contact events. We observe that, although the distributions of both the experiments follow an approximate power law distribution in the time interval [1 event, 30 events], the approximating exponent for GreaterBoston is also in this experiment considerably lower than the corresponding exponent for Infocom 2005. If we focus our attention on the contact event distribution of GreaterBoston, we can observe that roughly 30% of the nodes meet at least another node once. Moreover, by comparing the distribution of the any contact events with the distribution of the contact events (Fig. 6), we have that the two distributions are quite similar for GreaterBoston, while they differ significantly in InfoCom 2005. This result confirms the considerations made in the previous section about the inter-contact and the any-inter contact times.

4.3. Separation degree and delay characterization

In this sub-section we present the complementary cumulative distribution functions of the *separation degrees* and of the *separation delays*. The former metric measures the shortest distances between a pair of nodes in terms of hop count, while the latter measures the shorter distance in terms of end-to-end delay. The two distributions are obtained by considering disjointed time intervals of 24 h and by averaging the results obtained in each time interval. For each metric we consider two different

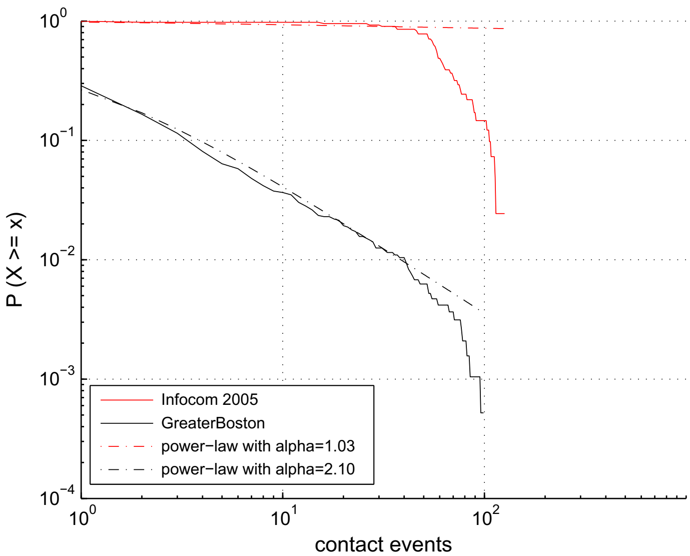

Fig. 7. Any contact event number complementary cumulative distribution function.

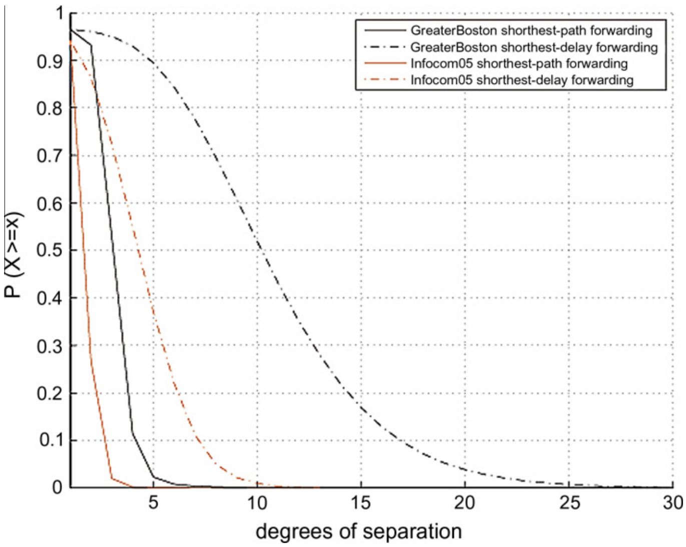

Fig. 8. Complementary cumulative distribution function of separation degrees.

store-and-forward strategies classes: the *shorthest-path* and the *shortest-delay*. The former minimizes the use of communication resources in packet forwarding, while the latter minimizes the delay.

We define the nodes v_s and v_d as *connected* in the time interval $[T_0 + n_{min}\tau, T_0 + n_{max}\tau]$ if it exists at least an ordered set of nodes $\mathcal{P}_{s,d} = (v_s = v_0, v_1, \ldots, v_l, v_{l+1} = v_d)$ and an ordered sets of temporal indexes $\mathcal{N}_{s,d} = (n_0, n_1, \ldots, n_l)$, with $n_{min} \leqslant n_0 \leqslant n_1 \leqslant \ldots n_l \leqslant n_{max}$, so that:

$$e_{v_k,v_{k+1}}(n_k) = 1 \quad \forall k = 0, \ldots, l. \tag{9}$$

According to the previous definition, by denoting with $\{P_{s,d}\}$ the set of all the ordered sets $\mathcal{P}_{s,d}$, we define the *separation degree* $\mathcal{D}_h(s,d)$ of the devices s and d as the length of the shortest path in terms of hop count:

$$\mathcal{D}_h(s,d) = \min_{\{\mathcal{P}_{f,\lceil}\}}\{|\mathcal{P}_{s,d}| - 1\}. \tag{10}$$

Clearly, if the nodes are not connected, their separation degree is assumed infinity.

Moreover, by denoting with $\{N_{s,d}\}$ the set of all the ordered sets $\mathcal{N}_{s,d}$, we define the *separation delay* $\mathcal{D}_d(s,d)$ as the shortest distance of the nodes in terms of time delay:

$$\mathcal{D}_d(s,d) = \min_{\{\mathcal{N}_{f,\lceil}\}}\{l\}\tau. \tag{11}$$

$\mathcal{D}_d(s,d)$ is assumed infinity if the nodes are not connected.

Fig. 8 shows the complementary cumulative distribution of the separation degrees for connected pair of nodes. We observe that, as expected, the separation degree for shortest-path forwarding is considerably lower than the one for shortest-delay forwarding for both the experiments. Surprisingly, although GreaterBoston experiment involves almost two thousand nodes spread over a huge area, the probability of reaching a node in more than six degrees of separation is negligible. For this, it is confirmed the presence of the *small world phenomenon* in a wide-scale experiment based on human mobility and the results corroborate the Milgram's famous experiment which spawned the popular saying *six degrees of separation* [5]. We do not claim that this observation holds for every kind of human mobility-enabled wireless network. Rather, we claim that it confirms the possibility of delivering packets in a wide ad hoc network by adopting multi-hop communications and by exploiting only the transmission opportunities provided by human mobility.

As regards to the complementary cumulative distribution of the separation delays for connected pairs of nodes shown by Fig. 9, we observe that the distributions of the two experiments are quite different and, in particular, we observe that the probability of packet being delivered with a delay between 5 and 14 h is negligible for both the InfoCom 2005 distributions. This behavior is reasonable, since these forwarding delays correspond to the day hours 10 p.m. – 7 a.m., in which we can suppose that the most of the conference attendees were sleeping.

4.4. Betweenness centrality characterization

In this sub-section we present the results regarding the degree betweenness centrality and the delay betweenness centralities metrics.

We define as *degree betweenness centrality* of a device v_i the ratio between the number of connected pair of devices for which the node belongs to the shortest-hop path and the number of all the connected pairs not involving v_i:

$$\mathcal{C}_h(i) = \frac{\sum_{v_j,v_k \in V}\sigma_h(i,j,k)}{(|V|-1)(|V|-2)}, \tag{12}$$

where we denote with $\sigma_h(i,j,k) = 1$ the event "device v_i belongs to the shortest-hop path between v_j and v_k" and with $\sigma_h(i,j,k) = 0$ the opposite event.

Similarly, we define as *delay betweenness centrality* of a node v_i the ratio between the number of connected pair of devices for which the node belongs to the shortest-delay path and the number of all the connected pairs not involving v_i:

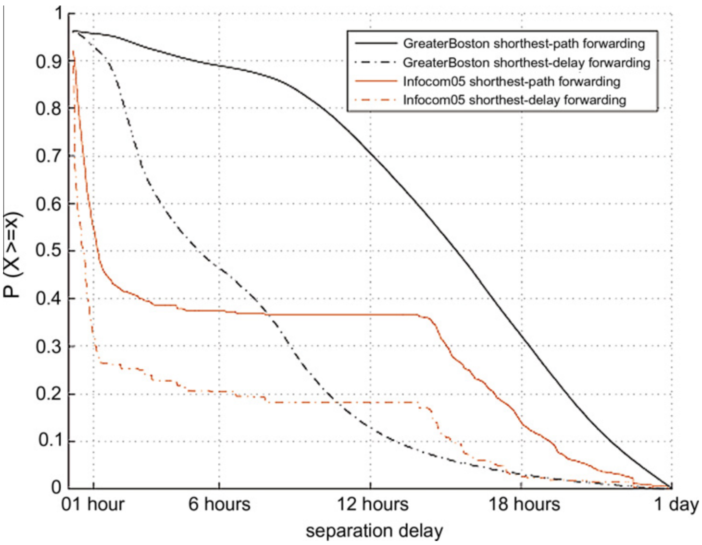

Fig. 9. Complementary cumulative distribution function of separation delays.

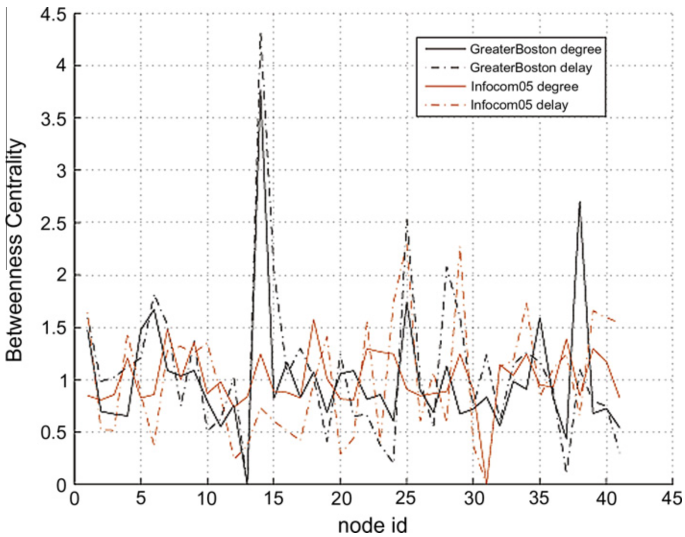

Fig. 10. Normalized betweenness centrality for a subset of devices.

$$\mathcal{C}_d(i) = \frac{\sum_{v_j, v_k \in V} \sigma_d(i,j,k)}{(|V|-1)(|V|-2)}, \tag{13}$$

where we denote with $\sigma_d(i,j,k) = 1$ the event "the node v_i belongs to the shortest-delay path between v_j and v_k" and with $\sigma_d(i,j,k) = 0$ the opposite event

We note that these definitions of betweenness centrality are different from the ones proposed in [7,18], where the betweenness centrality is defined as the frequency of node v_i acting as forwarder in all the shortest paths, not just in one. In this work, we adopt the definitions (12) and (13) since computing all the shortest paths for more than three millions pair of nodes for forty thousand time slots is not computational feasible.

Fig. 10 shows the betweenness centralities of a subset of devices (the first 41 according to the device id) normalized to the average value of the whole set of devices. The normalization allows us to highlight the spread of the values around the averages. From the figure, if is easy to observe the to contribution of devices in enabling end-to-end communication is not uniform. In other words, some devices are preferred for acting as forwarder due to their mobility patterns. Moreover, we observe that GreaterBoston experiment exhibits a significantly higher spread of the values for both the metrics with respect to InfoCom 2005 experiment. This behavior is reasonable and consistent with the differences between the mobility patterns of the two experiments: long distances and different geographic regions for GreaterBoston, and small distances and a conference location for InfoCom 2005. Finally, we note that the same device can perform differently for the different metrics: with reference to GreaterBoston

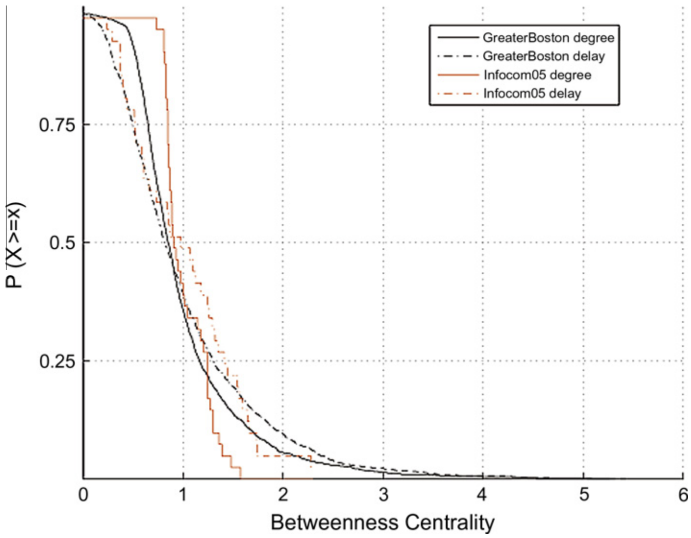

Fig. 11. Complementary cumulative distribution function of the normalized betweenness centrality.

experiment, we have for example that device 38 exhibits a high degree betweenness centrality, and at the same time it exhibits an average delay betweenness centrality.

The qualitative measurement proposed in Fig. 10 is confirmed by the results shown in Fig. 11, which shows the complementary cumulative distribution functions of the normalized betweenness centralities. At first, we observe that the GreaterBoston distributions exhibits a smoother shape with respect to the InfoCom 2005 ones. This result confirms the higher spread of the values for both the metrics of GreaterBoston experiment with respect to InfoCom 2005 one obtained with the qualitative measurement (Fig. 10). Moreover, we observe that for both the experiments the distribution of the delay betweenness centrality is smother than the distribution of the degree betweenness centrality. This confirms the result obtained with the qualitative measurement shown in Fig. 10.

If we analyze the results for the betweenness centralities with a networking prospective, we can highlight two key characteristics, common to both the experiments although more evident in GreaterBoston one. The first characteristic is that there is a significant difference in term of potential contribution to the forwarding process among the devices. As a consequence, a forwarding algorithm for human-enabled networks should be able to capture this difference for increasing the probability of information delivery. The second characteristic is that different route metrics imply different potential contribution for the same node to the forwarding process. We believe that this dependence between the route metric and the device centrality is a key factor for enabling efficient human mobility-enabled networks which requires a coupled design of both the route metric and the forwarding algorithm.

5. Conclusions

In the last 10 years, new paradigms for wireless networks which aim to exploit human mobility to improve the connectivity in sparse or isolated networks have gained attention of the research community. In this paper, we describe the results of an experiment in which the positions of almost two thousand humans, spread over a metropolitan region greater than 200 km^2, have been anonymously traced for roughly one month. The results regarding the contact and inter-contact times are in agreement with those obtained in previous small-scale experiments in working and conference environments, confirming that a power law approximation holds. We then analyzed the characteristics of the contact events distributions, and found that in this case the power law approximation holds practically always. In addition, we explored the degree and the delay of separation for two different forwarding strategies classes. For the class which aims to minimize the separation degree, we found that the probability that a node be more than six degrees of separation away from another node is negligible. As far as we know, this is the first experiment involving wide-scale mobile ad hoc wireless networks in which the *small world phenomenon* has been discovered. Finally, we analyzed the node popularity distributions in terms of betweenness centrality for both the forwarding strategies, and we found a dependance between the popularity of a device and the adopted routing metric. As future work, we are planning to analyze the mobility patterns for larger populations, and the related contact events in presence of emergency events. This will help us design and evaluate routing protocols for human mobility-enabled networks based on the insights acquired with wide-scale experiments.

References

[1] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, C. Diot, Pocket switched networks and human mobility in conference environments, in: WDTN '05: Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-Tolerant Networking, 2005, pp. 244–251.

[2] E.M. Daly, M. Haahr, The challenges of disconnected delay-tolerant manets, Ad Hoc Networks 8 (2) (2010) 241–250.
[3] A.S. Cacciapuoti, M. Caleffi, L. Paura, Mobile p2p: peer-to-peer systems over delay tolerant networks, in: A. Vasilakos, Y. Zhang, T. Spyropoulos (Eds.), Delay Tolerant Networks: Protocols and Applications, CRC Press, 2011, pp. 1–35.
[4] M. Gonzalez, C. Hidalgo, A.L. Barabasi, Understanding individual human mobility patterns, Nature 453 (7196) (2008) 779–782, doi:10.1038/nature06958.
[5] S. Milgram, The small world problem, Psychology Today 2 (1967) 60–67.
[6] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, J. Scott, Impact of human mobility on opportunistic forwarding algorithms, IEEE Transactions on Mobile Computing 6 (6) (2007) 606–620.
[7] P. Hui, J. Crowcroft, E. Yoneki, Bubble rap: social-based forwarding in delay tolerant networks, in: MobiHoc '08: Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing, 2008, pp. 241–250.
[8] E.M. Daly, M. Haahr, Social network analysis for routing in disconnected delay-tolerant manets, in: MobiHoc '07: Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing, 2007, pp. 32–40.
[9] J. Tang, M. Musolesi, C. Mascolo, V. Latora, Temporal distance metrics for social network analysis, in: WOSN '09: Proceedings of the 2nd ACM Workshop on Online Social Networks, 2009, pp. 31–36.
[10] J. Leguay, A. Lindgren, J. Scott, T. Friedman, J. Crowcroft, Opportunistic content distribution in an urban setting, in: CHANTS '06: Proceedings of the 2006 SIGCOMM Workshop on Challenged Networks, 2006, pp. 205–212.
[11] F. Calabrese, G. DiLorenzo, C. Ratti, Human mobility prediction based on individual and collective geographical preferences, in: IEEE International Conference on Intelligent Transportation Systems, Madeira Island, Portugal, 2010.
[12] F. Calabrese, C. Ratti, Real time rome, Networks and Communications Studies 20 (3–4) (2006) 247–258.
[13] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, C. Ratti, Real-Time urban monitoring using cell phones: a case study in rome, IEEE Transactions on Intelligent Transportation Systems 12 (1) (2011) 141–151.
[14] Wi-fi capable handsets: residential and enterprise markets for wi-fi and dual-mode handsets. Market research analysis; ABI Research, 2009.
[15] Wi-fi enabled mobile phone handsets in the us, 2010–2015, Market research analysis; Coda Research Consultancy, 2010.
[16] G. Anastasi, E. Borgia, M. Conti, E. Gregori, Ieee 802.11b ad hoc networks: performance measurements, Cluster Computing 8 (2–3) (2005) 135–145.
[17] V. Kostakos, Temporal graphs, Physica A: Statistical Mechanics and Its Applications 388 (6) (2009).
[18] J. Tang, M. Musolesi, C. Mascolo, V. Latora, V. Nicosia, Analysing information flows and key mediators through temporal centrality metrics, in: Proceedings of the 3rd ACM Workshop on Social Network Systems (SNS '10), ACM, Paris, France, 2010.

Angela Sara Cacciapuoti received the Dr. Eng. degree summa cum laude in Telecommunications Engineering in 2005, and the Ph.D degree with score "excellent" in Electronic and Telecommunications Engineering in 2009, both from University of Naples Federico II. Since 2008, she is with the Dept. of Biomedical, Electronic and Telecommunications Engineering (DIBET), University of Naples Federico II, as postdoctoral research fellow. From 2010 to 2011, she has been with Broadband Wireless Networking Lab, Georgia Institute of Technology, as visiting researcher. In 2011, she has also been with NaNoNetworking Center in Catalunya (N3Cat), Universitat Politecnica de Catalunya, as visiting research. Her research activities lie in the area of statistical signal processing, digital communications, and communication systems. In particular, her current interests are focused on non-stationary signal processing, cognitive networks and wireless communications cross-layering.

Francesco Calabrese is an advisory research staff member at the IBM Dublin Research Laboratory and a research affiliate at the SENSEable City Laboratory of the Massachusetts Institute of Technology. His research interests include ubiquitous computing, intelligent transportation systems, urban network analysis, and distributed control systems design. Calabrese has a Ph.D in computer and systems engineering from the University of Naples Federico II, Italy. He's a member of IEEE and the IEEE Control Systems Society. Contact him at fcalabre@ie.ibm.com.

Marcello Caleffi received the Dr. Eng. degree summa cum laude in Computer Science Engineering from University of Lecce in 2005, and the Ph.D. degree in Electronic and Telecommunications Engineering from University of Naples Federico II in 2009. Since 2008, he is with the Dept. of Biomedical, Electronic and Telecommunications Engineering, University of Naples Federico II, as postdoctoral research fellow. From 2010 to 2011, he was with Broadband Wireless Networking Lab, Georgia Institute of Technology, as visiting researcher. In 2011, he was with NaNoNetworking Center in Catalunya (N3Cat), Universitat Politecnica de Catalunya, as visiting researcher. His research activities lie in the area of wireless networking. In particular, his current interests are focused on cognitive ad hoc networks and human-enabled wireless networks.

Giusy Di Lorenzo is a research scientist at the IBM Dublin Research Laboratory and a research affiliate at the SENSEable City Laboratory of the Massachusetts Institute of Technology. Her research explores the analysis of urban dynamics and human mobility using data gathered from sensor networks. In particular, she's interested in developing context-aware applications to improve urban living experiences of citizens. Di Lorenzo has a Ph.D in computer and systems engineering from the University of Naples Federico II. Contact her at giusydil@ie.ibm.com.

Luigi Paura received the Dr. Eng. degree summa cum laude in Electronic Engineering in 1974 from University of Napoli Federico II. From 1979 to 1984, he was with the Dept. of Biomedical, Electronic and Telecommunication Engineering, University of Napoli Federico II, first as an Assistant Professor and then as an Associate Professor. Since 1994, he has been a Full Professor of Telecommunications: first, with the Dept. of Mathematics, University of Lecce, Italy; then, with the Dept. of Information Engineering, Second University of Naples; and, finally, from 1998 he has been with the Dept. of Biomedical, Electronic and Telecommunication Engineering, University of Naples Federic o II. He also held teaching positions at University of Salerno, at University of Sannio, and at University Parthenope of Naples. In 1985–86 and 1991 he was a Visiting Researcher at Signal and Image Processing Laboratory, University of California, Davis. At the present time, his research activities are mainly concerned with statistical signal processing, digital communication systems, and cognitive networks. He is co-author of more than one hundred and ten technical papers on international journals and conference proceedings.