Mobile Opportunistic Traffic Offloading

(MOTO)

D3.2: Spatiotemporal characterization of contact patterns in dynamic networks

Document information	
Edition	13
Date	29/04/2014
Status	Edition
Editor	UPMC
Contributors	UPMC, TCS, CNR, INNO

Contents

1 Executive summary

The efficiency of opportunistic offloading inherently relies on how the proposed strategies get advantage from the dynamics of the underlying mobile network. The patterns that govern encounters between nodes have direct influence on the opportunities to implement device-to-device content exchanges, relieving thus the cellular infrastructure. In this deliverable, we report the main contributions of the MOTO project to better understand such patterns. We structure our work around complementary points of view, ranging from the generalization of mobility modeling, to the impact of duty cycling, to a more complete investigation of node vicinity. For the sake of completeness, we also consider security trends that may arise when connectivity is intermittent.

2 Introduction

To achieve efficient opportunistic offloading, it is fundamental to understand how nodes move around and, most importantly, how they meet creating communication opportunities. Research in this area has been very active in the latest years, leading to both fundamental discoveries and practical observations. Although the community has achieved a reasonable level of maturity in this area, several fundamental issues remain open. In fact, most of the related work considers the network as a whole. In a context where every opportunity for communication counts, it becomes important to capture contact opportunities in a fine-grained fashion and characterize mobility at the microscopic level. For example, nodes that meet frequently but apart from the rest of the network have limited contribution to the capacity of the system. Or nodes may sleep and miss contact opportunities. Similarly, a node that shows strong contact activity but with the same neighbours might be poor diffusers. Also, in crowded spaces users might switch on and off their devices, thus generating a non-mobility-induced opportunistic network.

In the MOTO project, we advocate that it is no further sufficient to investigate the average behaviour of the network. We propose to tackle the problem from a spatiotemporal viewpoint, by considering contacts not only as individual phenomena but also as an atomic event. This is a very challenging problem, as it introduces several variables into the equation. In particular, we need to define the role that a particular node can take in the network by considering its spatial mobility and its interactions with the other nodes. Last but not least, such a role must take into account the other interactions happening throughout the network and possibly future events predicted by models based on the history of the system.

These considerations lead to specific questions that are central to the activities of the MOTO project:

- Would it be possible to create mobility models that easily account for specific requirements and features, thus better reproducing contact and intercontact patterns?
- What is the impact of duty cycling on contact and intercontact patterns?
- Would it be worth relaxing the definition of "proximity" in opportunistic networks? When two nodes are not in direct contact, are they nearby or really far away?

Providing answers to these questions can significantly help design protocols and algorithms to achieve higher offloading gains. In fact, the utility of the MOTO framework is only worth it if encounters between nodes are sufficient to allow frequent device-to-device exchanges (reducing the load on the cellular infrastructure).

In this deliverable, we report scientific contributions to fill the gaps described above. In a nutshell, we describe the following achievements:

- **A unifying framework for contact patterns (Section 3).** Existing mobility models are limited in terms of flexibility and controllability. The main reason is that they are in too general (thus unrealistic) or too specific (thus dedicated), which prevents us to easily adapt their parameters to any kind of observed mobility. We propose SPoT, a framework that combines social, spatial, and temporal features to overcome such limitations.

- **Duty-cycling and encounters (Section 4).** Device-to-device communications consume significant energy, so nodes are typically operated in duty cycling mode. During these periods, nodes may miss contact opportunities, with potentially severe consequences to the offloading scheme. We investigate how power saving techniques, which may effectively reduce the number of usable contacts, affect intercontact times.
- **Node proximity beyond one hop (Section 5).** Existing analyses of meeting patterns rely on the binary assumption that nodes are either in contact or in intercontact. We revisit this assumption and provide arguments toward the adoption of an extended view of the neighborhood. We provide evidences that considering nodes that are not in contact, but still reachable, can be of great help in the design of efficient opportunistic networks.
- **Predicting node proximity (Section 6).** Recent studies have revealed that, under the right prediction method and predictive features, contacts between mobile users are to a certain extent predictable. We extend such an analysis to the case of extended neighborhood (previous bullet) and show that there is a higher potential on relying on extended neighborhood prediction compared to the traditional one-hop case.
- **Security issues (Section 7).** As security is an important keyword in the MOTO project, we decided to investigate the relationships between contact patterns and security trends. In particular, we are interested in understanding the impact of connectivity disruptions on confidentiality, integrity, and availability.

For the sake of completeness, we provide a number of research papers related to the different sections in the Appendix of this deliverable.

3 Contact patterns under a unifying framework

The current approach to human mobility modeling is based on trying to incorporate in the model the newest features of mobility properties as they come up from trace analysis, typically focusing on just a few of them. So, for example, the class of social-based mobility models aims to exploit the relation between sociality and movements, and to formalize social interactions as the main driver of human movements [1] [2]. The disadvantage of this approach is that the proposed models are intrinsically bound to the current state of the art on trace analysis, and need to be redesigned from scratch any time a new discovery is made. In addition, with current mobility models it is typically difficult, if not impossible, to fine tune the mobility properties (e.g., obtaining intercontact times featuring a probability distribution with controllable parameters). Overall, flexibility (i.e., allowing for different distributions of mobility properties – e.g., intercontact times – without redesigning the mobility model) and controllability (i.e., obtaining a predictable output starting from a given input) are currently missing from available models of human mobility.

The goal of this research activity is to address the above limitations. To this aim, we have proposed a mobility framework (SPoT – Social, sPatial, and Temporal mobility framework) that incorporates the three dimensions of human mobility and that is flexible and controllable. In the following, we will summarise the main ideas of SPoT, we will discuss how to configure it relying on the analysis of real traces of human mobility, and we will show that SPoT is both flexible and controllable [3].[1]

3.1 The SPoT Mobility Framework

In the following we summarise how the main components of our SPoT mobility framework work. SPoT is designed around the three main dimensions of human mobility, i.e., the social, spatial and temporal (see Figure 1). The social dimension is explicitly captured in the framework by taking a graph of human

[1] Reference available in the appendix.

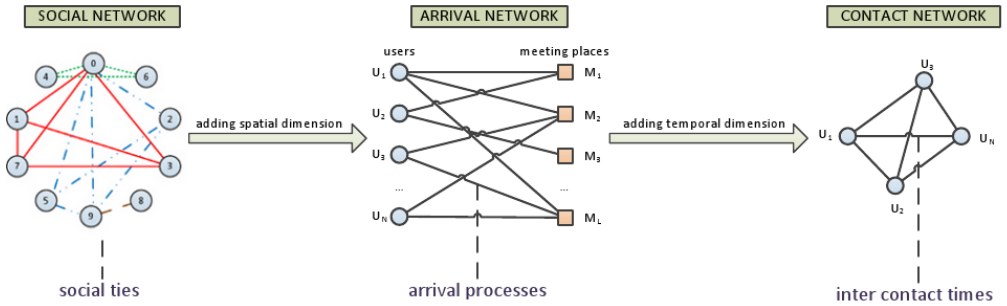


Figure 1: Framework overview

social relationships as an input parameter. This graph can be any well known graph, such as random graphs [4] or scale-free graphs [4], or it can be extracted from real traces. Then, SPoT adds the spatial dimension to the social ties by generating an *arrival network*, which is a bipartite graph that connects users and meeting places. Specifically, user i and a meeting place l are connected in the arrival network if user i visits meeting place l. Finally, in order to add the temporal dimension to the model, each link in the arrival network is associated to a stochastic point process that characterises how arrivals of the user to the meeting place are distributed over time. By sampling from the random variables representing the time between consecutive arrivals, we obtain the time sequences of the visits of a user to a given location. Then, the contact network, i.e., the network describing the contacts between nodes, can be obtained by assuming that two nodes are in contact with each other if they happen to be at the same time in the same meeting place.

3.1.1 The social and spatial dimensions of human mobility

Social interactions between users have emerged as one of the key factors defining human mobile behavior, because individuals belong to *social communities* and their social ties strongly affect their movement decisions [5] [6]. We consider proximity-based communities, i.e., communities whose members share a common meeting place (e.g., offices, bars, apartments). Since all members of the community visit a shared meeting place, it implies that users are socially connected with all other members of the community, and, therefore, form fully connected components (i.e., cliques) in the social graph. Such cliques in realistic social networks exhibit an *overlapping* and *hierarchical structure* [7] [8]. Each user belongs to several overlapping cliques, representing different social circles (e.g., friends, relatives, colleagues). On the other hand, each clique is itself composed of a number of nested cliques, which share additional meeting places that are not common to all the users of a parent clique. For example, a company shares a set of offices visited by all its employees, while each subdivision has its own working places.

As anticipated, we represent the relation between the spatial and the social dimension of human mobility by means of a bipartite graph of users and meeting places, which we call arrival network. In the algorithm (summarized in Table 1) for generating the arrival network starting from the input social graph we mainly need two components: a clique finding algorithm (which also detects overlapping cliques) and a way for reproducing hierarchical cliques.

The first component corresponds to steps 1 and 2 in Table 1. In each round, the social graph is divided into a set (called *cover*) of overlapping cliques, such that each link of the social graph is assigned to exactly one clique. To this purpose, we use the BronKerbosch algorithm [9]. The cover of each round tries to capture the biggest possible cliques. For each of the newly identified cliques, we create a new meeting place and assign all members of the clique to that meeting place. In other words, we create a new meeting place vertex in the arrival network and we add links between this vertex and all members of the community. As an example, we describe in Figure 2 how cliques identified in the social graph are reflected into corresponding meeting places. The second component (step 3 in Table 1) of the algorithm for generating the arrival network allows us to generate nested cliques. More specifically, our algorithm

Table 1: Algorithm for building the arrival network - Input: social graph G and removal probability α

1. Divide input social graph G into a set of overlapping cliques, such that the sizes of the cliques are maximum and each link is assigned to exactly one clique. To this aim, the BronKerbosch algorithm [9] can be used.
2. To each clique assign a separate meeting place, i.e., create a new meeting place and a set of links between this place and each member of the clique in the arrival network.
3. Remove randomly each link in the social graph with probability α, inducing emergence of new nested cliques.
4. Proceed to the next round starting from the first step, until there are no links left in the input graph.

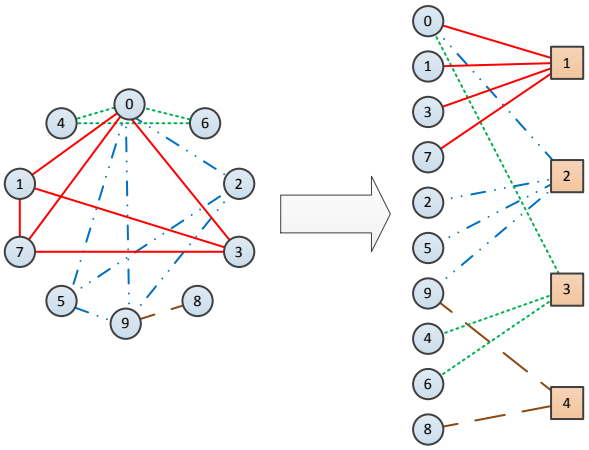


Figure 2: A round of assigning social cliques to meeting place; cliques are marked with different line styles

tries to identify cliques of lower size nested into those identified in the previous round. To do so, cliques are split according to a very simple random process, according to which every link in the social graph is randomly removed with a constant, configurable, probability α (*removal probability*). This leads to the emergence of smaller cliques, which are indeed nested into the original ones. This simple strategy has also the advantage of allowing for a fine control of the number of meeting places shared by users. In fact, each link participates into a geometrically distributed (with parameter α) number of rounds of meeting place assignments. As each link is assigned to at most one clique per round, also the number of cliques that includes that link will be geometrically distributed. This implies that the pair of users i, j with which this link is associated will share a number L_{ij} of cliques (and thus of meeting places) that is itself geometrically distributed with parameter α. The algorithm for generating the arrival network stops (step 4 in Table 1) when there are no more links in the social graph to be removed.

3.1.2 From meeting places to geographical locations

The analysis of the algorithm in Table 1 reveals that the number of meeting places generated grows with the number of cliques. Thus, the more cliques in the input social graph, the more meeting places are required. The proliferation of meeting places is not of big concern as meeting places might correspond to very small geographic areas (e.g., offices). However, in order to improve the realism of the generated

scenario, we combine these meeting places into a fixed number L of wider physical locations (e.g., this is equivalent to combining offices into a business center). To assign meeting places to geographical locations we explore the observation that, intuitively, the places that share many common frequent visitors should be located geographically close to each other, like in the case of different buildings of a university campus or different offices of a company. In [3] we have validated and confirmed this observation using three datasets extracted from location-based online social networks. In order to quantify the closeness between two meeting places, we define the strength F_{ij} of ties between a pair of meeting places i and j as the summary co-appearing frequency across all the users the two places share. More formally, we can write F_{ij} as follows:

$$F_{i,j} = \sum_{u \in \mathcal{U}_{i,j}} f_u^i \times f_u^j \tag{1}$$

where f_u^i is the frequency of user u's visits to location i and $\mathcal{U}_{i,j}$ is the set of all users shared between place i and j. The higher the arrival frequency of user u to both places i and j, the higher the strength between the two places.

Our goal now is to distribute meeting places on the 2D plane such that pairs of places with stronger ties in terms of shared visitors would be located closer. To this aim, we use a variation of the energy model for graph drawing described in [10, 11]. In this model, the places are represented as particles, where particles connected with a link attract each other proportionally to the power of the strength of the link and inversely proportional to the power of the distance between the particles. Similarly, particles that are not connected with a link repulse each other. The final spatial positioning of the meeting places is achieved through simulation, where initial positions of the places are selected randomly in a rectangle of size $w \times h$. As a result of applying attraction and repulsing forces to the nodes, the system eventually reaches an equilibrium state in which tightly connected meeting places are situated close to each other, thus achieving our desired goal.

3.1.3 The temporal dimension of user visits to meeting places

The arrival network that we have built in the previous section tells us which are the meeting places visited by each user. Here we add the temporal properties of such visits. To this aim, we assign to each link in the arrival network a discrete stochastic point process A_i^l that describes the arrivals of user i to a meeting place l over time. In this work, we have considered only discrete point processes, leaving the continuous case for future work. In a discrete point process, the time is slotted. During a time slot, each node visits a set of locations, where this set is determined by the evolution of the arrival processes.

We assume that processes A_i^l are independent, i.e., that nodes arrive to locations independently of each other. This implies that also the resulting contact processes are independent. In real traces, contacts can be synchronized [12], but coordination[2] between nodes may drastically complicate the mathematical analysis of mobility frameworks. For this reason, keeping in mind our target of controllability, we decided to limit the focus to independent arrival processes. The comparison with traces (where these assumptions in general do not hold) presented in Section 3.3 shows that the accuracy of the model is good, nevertheless.

Once we have characterized the time at which users visit their assigned meeting places, we can build the contact graph of the network (Figure 1). In fact, a contact between two users happens if the two users appear in the same meeting place at the same time slot. A contact duration is measured as the number of consecutive time slots in which two users have at least one commonly visited location. The contact graph can be fully mathematically characterized (we provide an example of this characterization in Section 3.4 for the case of arrival processes being heterogenous Bernoulli processes) or it can be obtained from simulations.

[2] A weaker coordination involving only pairs of nodes has been sometimes assumed in the literature for modelling purposes. With pairwise coordination, pairs of nodes can still meet on purpose, but independently of the other pairs. However, since the tractability of our analytical framework would not benefit from this assumption, we decided to use the strongest independence condition.

3.2 Analysis of real user movements

The SPoT framework takes as input the social graph of the network users and the arrival processes describing how users visit locations. The properties of the user social graph have been extensively studied in the literature [13, 14], thus making their configuration easy. Instead, the statistical characterisation of user arrivals has been little explored, especially for what concerns the individual user-pair behavior. In order to address this open point for realistically configuring the framework, we have considered three datasets of real user movements, extracted from the location-based online social networks Gowalla [15], Foursquare [16], and Altergeo [17]. In location-based online social networks, users check-in at places (e.g., restaurants, offices) and share their location with their friends. Thus, the concept of check-ins is very similar to the arrivals considered in the SPoT mobility framework. In fact, both notions represent records of the time at which users visit particular venues. For this reason, we chose to take check-ins as proxies for user arrivals at places and to use them to measure the temporal characteristics of arrival sequences.

From a preliminary analysis we observed that across a significant population of user-place pairs the distribution of inter-arrival times has the shape of a straight line in lin-log scale, which roughly corresponds to a geometric distribution in the discrete case. Similarly, a preliminary observation of the pairwise intercontact time yielded again a geometric distribution. We aim to validate this hypothesis by fitting individual inter-arrival time and intercontact time distribution to a geometric distribution and evaluating the goodness of fit across all user-place pairs and user-user pairs, respectively, in the dataset. The fitting is performed using Maximum Likelihood Estimation [18], and we test whether it is plausible that our data come in fact from such fitted distribution using one of the most popular goodness of fit tests, the Pearson's chi-squared test [18].

The percentage Q of user-place pairs for which the assumption on the geometric distribution of inter-arrival times is not rejected at different significance levels is shown in Table 2. We observe that for the majority of pairs across all datasets, i.e., $Q_{GO}^{geom} = 60\%$, $Q_{FS}^{geom} = 66\%$, $Q_{AG}^{geom} = 54\%$, the inter-arrival time distribution follows a geometric distribution. This result is important as a geometric distribution of inter-arrival times can be modeled with a simple Bernoulli arrival process, which, as we discuss in Section 3.4, is very convenient for mathematical analysis. The physical implication behind Bernoulli arrivals is that users tend to visit places with a fixed rate. This matches the common finding [19] that users tend be quite regular in their movements. Relying on these three datasets, we have also analysed the pairwise intercontact times. The pairwise results will be later compared against the mathematical results in Section 3.4, showing that data and model predictions are totally in agreement. The main obstacle in computing intercontact times in our datasets is that there are no check-out records, i.e., no records of the time when users leave places. For this reason, we have to make some assumptions about the duration of the sojourn time at a location. In [12], the intercontact times for the Gowalla trace (the exact same trace that we consider in this work) were measured assuming that a contact between two users happen if they have checked-in less than 1 hour apart at the same place. The rationale for this choice lies behind the nature of location-based online social networks like Gowalla, Foursquare, and Altergeo. In fact, these applications capture mostly users going out for eating or entertainment, for which the 1-hour choice appears reasonable. Thus, also in our work we kept the 1 hour assumption. The results of this analysis are shown in Table 3. Summarizing, the chi-squared test does not reject the geometric hypothesis for $Q_{GO}^{geom} = 80\%$, $Q_{FS}^{geom} = 87\%$, $Q_{AG}^{geom} = 78\%$ of pairs in our datasets.

Table 2: Percentage of pairs for which the geometric distribution hypothesis for arrivals is not rejected, at different significance levels

α	Gowalla (%)	Foursquare (%)	Altergeo (%)
0.001	0.71	0.77	0.57
0.01	0.60	0.66	0.54
0.05	0.50	0.51	0.49

Table 3: Percentage of pairs for which the geometric distribution hypothesis for intercontact times is not rejected, at different significance levels

α	Gowalla (%)	Foursquare (%)	Altergeo (%)
0.001	0.89	0.94	0.84
0.01	0.80	0.87	0.78
0.05	0.51	0.74	0.56

3.3 Testing the framework flexibility

As discussed before, the capability of allowing for different distributions of mobility metrics (which we have called flexibility) is one of the properties of SPoT. In order to showcase this property, in the following we show that SPoT, once configured for the settings observed in a real mobility trace, generates the same aggregate characteristics (aggregate intercontact times, specifically) as those seen in traces. We chose the aggregate over the pairwise statistics in this case because, from the mathematical characterisation of the framework (Section 3.4), we know that SPoT, once configured with Bernoulli arrivals, will generate geometric intercontact times. Since we also know from trace analysis (Section 3.2) that intercontact times in the dataset can be approximated with a geometric distribution for a large fraction of pairs, a match between the intercontact times generated by the framework and those seen in traces would be quite expected. Less obvious, instead, is the capability of reproducing also a realistic aggregate behaviour starting from pairwise controlled parameters.

In order to use the framework, we need to configure the following quantities: the social graph G, the removal probability α, and the arrival processes A_i^l for each user i visiting a location l. We extract this information from the data traces themselves (details on how this is done can be found in [3]). In Figure 3 we show the aggregate intercontact time generated by SPoT against those observed in the traces. As we can see from the plot, the aggregate behavior observed in traces (red squares) is in good agreement with the corresponding results from the simulation (blue crosses). This confirms the flexibility of the framework to capture a desired realistic behavior seen in real traces.

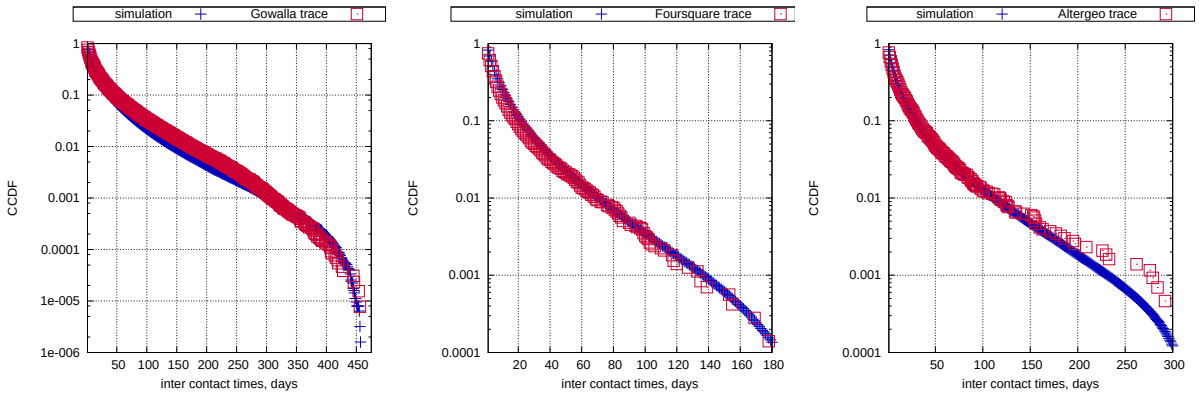


Figure 3: Aggregate intercontact times obtained from traces (red squares) and from simulations (blue crosses)

3.4 Testing the framework controllability

Besides flexibility, the other goal of SPoT is controllability. Thus, in [3] we have derived mathematically how the SPoT framework is able to produce different, controllable outputs depending on its initial con-

figuration. To this aim, we exploit the data analysis results and we focus on Bernoulli arrivals, which we have shown to represent the behaviour of the majority of user-place pairs. Using the Bernoulli assumption, we fully characterise the pairwise dynamics of the framework and we also analytically derive the conditions under which heavy-tailed and exponentially-tailed aggregate intercontact times, two cases often observed in real traces, emerge. Below we will summarise the results obtained, whose detailed description can be found in [3].

We assume that each user U_i visits place M_l according to a Bernoulli process A_i^l with rate $\rho_{A_i^l}$. Under this assumption, it can be derived the intercontact times distribution between a pair of users U_i and U_j, meeting at a number L_{ij} of meeting places, is geometric with the following rate:

$$\rho = 1 - \prod_{l=1}^{L_{ij}} (1 - \rho_{A_i^l} \times \rho_{A_j^l}). \tag{2}$$

Please note that the above result is in agreement with what we have seen in traces (Section 3.2).

As for the characterisation of the aggregate, we have derived three conditions under which exponential, exponentially-tailed, and heavy-tailed aggregate intercontact times emerge in SPoT. More specifically, when individual arrival processes are independent Bernoulli point processes with homogeneous rates $\rho_{A_i^l} = \beta$ and the number of shared meeting places L_{ij} between pairs of users is constant, i.e., $L_{ij} = L$, then the aggregate intercontact times feature a discrete exponential (i.e., geometric) distribution with CCDF $F(\tau) = e^{-\gamma\tau}$ ($\gamma = -L\ln(1-\beta^2)$). Instead, when individual arrival processes are independent Bernoulli point processes with homogeneous rates $\rho_{A_i^l} = \beta$ but the number of shared meeting places L_{ij} between pairs of users is a geometric random variable with parameter α, then the CCDF of the aggregated intercontact times has an exponential tail, i.e., $F(\tau) \sim e^{-\delta\tau}, \tau \to \infty$, where $\delta = -\ln(1-\beta^2)$. Finally, if individual arrival processes are independent Bernoulli point processes, the rates $\rho_{A_i^l}$ of the processes are drawn such that $\rho_{A_i^l} = e^{-\frac{1}{2}Y^2}$ (where Y is a standard normal random variable) and the number of shared meeting places L_{ij} between pairs of users is a geometric random variable with parameter α, the CCDF of the aggregated intercontact times is given by Equation 3.

$$F(\tau) = \frac{a + a^2}{(\tau + a)(\tau + a + 1)} \quad \left(x \to \infty \Rightarrow F(\tau) \sim 1/\tau^2\right). \tag{3}$$

3.4.1 Validation

In this section, we validate the results obtained above comparing analytical predictions against simulation results. In order to instantiate the proposed framework, we need to define its input parameters: the social graph G, the removal probability α, and the arrival processes A_i^l for each user i visiting a location l. We use the state-of-the-art Barabási-Albert model [20] to generate input social graphs with realistic characteristics (e.g., scale-free degree distribution, short average path length). Thus we consider the two graphs G_{n_1,m_1} and G_{n_2,m_2} of $n_1 = 500$ and $n_2 = 1000$ users and growth parameters $m_1 = 50$ and $m_2 = 30$. We evaluate both graphs G_{n_1,m_1} and G_{n_2,m_2} when the removal probability used by the algorithm for generating the arrival network is $\alpha_1 = 0.5$ and $\alpha_2 = 0.2$. As a result, we obtain four arrival networks with different structural parameters which we explore in simulations. For each of these arrival networks, we study the resulting intercontact times obtained changing the characteristics of the arrival processes A_i^l of users to meeting places. More specifically, we focus on two cases discussed in the previous sections, namely, when the arrival processes are homogeneous and when the arrival rates features specific distribution that leads to the heavy-tailed aggregate intercontact times. Simulations are run for 10000 time units of simulated time, and results are shown with a confidence level of 99.9%.

We start with the heavy-tailed case. To this aim, we assign rates $\rho_{A_i^l}$ of the Bernoulli arrival processes such that $\rho_{A_i^l} = e^{-\frac{1}{2}Y^2}$, where Y is a standard normal random variable. These settings correspond to the case which we have mathematically characterised in the previous section. Figure 4 depicts the result of simulations for each of the arrival networks. For instance, Figure 4.a depicts simulation results for the network with parameters $n = 500$, $m = 50$ and $a = 0.5$. As we can see from the figure, the resulting

aggregate intercontact time CCDF for this network decays as a power law of exponent $\gamma = -2$, i.e., $F(\tau) \sim \tau^{-2}$. In the other arrival networks we observe similar results, which are in agreement with the theoretical predictions.

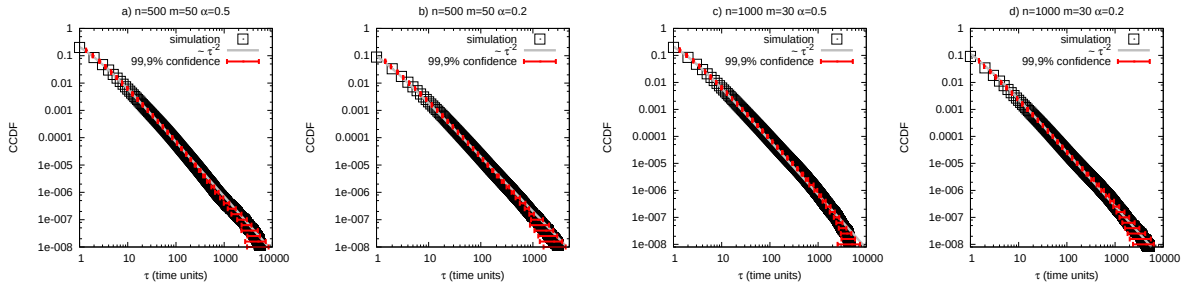


Figure 4: The aggregate intercontact times distribution for different arrival networks

In the second set of simulations we consider arrival networks where arrival processes are Bernoulli processes with identical rates. More specifically, we model two networks with same parameters $\{n = 500, m = 50, a = 0.5\}$, in which all the rates of arrival processes are identical and equal to $\rho^{(1)}_{A^l_i} = 1/2$ for the first network, and $\rho^{(2)}_{A^l_i} = 1/3$ for the second. From Figure 5 we can see that the resulting distribution of the aggregate intercontact times decays as an exponential function with exponent $\delta_{(1)} = 0.29$ in the first case and $\delta_{(2)} = 0.12$ in the second. This result is in agreement with the theoretical prediction ($\delta = -\ln[1 - \rho^2]$, where ρ is the rate of the arrival process).

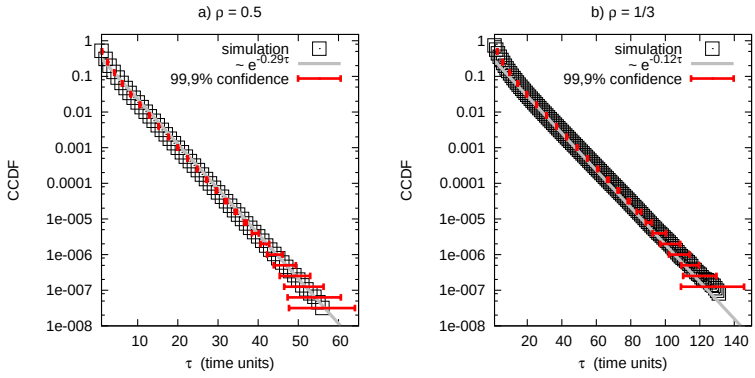


Figure 5: The aggregate intercontact times distribution for arrival network with identical arrival rates

3.5 Final remarks

In this section we have discussed SPoT, a mobility framework that incorporates the spatial, social, and temporal dimensions of human mobility. The social and spatial dimensions are added imposing that people belonging to the same social community are assigned to the same location, which is where the people of that community meet. Then, the way users visit their assigned locations over time (corresponding to the temporal aspects of mobility) is described by means of a stochastic process. In order to provide a realistic instantiation of the framework in terms of the arrival process of users to meeting places, we have analyzed three datasets containing traces of human check-ins at real locations, extracted from the online location-based social networks Gowalla, Foursquare, and Altergeo. The analysis of these datasets has revealed that human arrivals to places can be reasonably approximated, for the majority of user-place

pairs, by Bernoulli processes. In the third part of the section we have focused on the flexibility and controllability of the framework. First we have shown that the SPoT framework can be easily instantiated to accurately reproduce the mobility behavior seen in the Gowalla, Foursquare, and Altergeo traces. Second, as far as the controllability is concerned, we have analytically derived the conditions under which aggregate heavy-tailed and exponentially-tailed intercontact times emerge, and we have shown that these analytical predictions are totally in agreement with simulation results.

4 Impact of duty cycling on contact patterns

A possible roadblock in the scenario of offloading through ad-hoc WiFi or Bluetooth technologies is the fact that direct, opportunistic, communications between nodes consume significant energy. To address this, nodes are typically operated in duty cycling mode, by letting their WiFi (or Bluetooth) interfaces ON only for a fraction of time. The joint effect of duty cycling and mobility is that devices are able to directly communicate with each other only when they come in one-hop radio range *and* both interfaces are ON. The net effect of implementing a duty cycling scheme is thus the fact that some contacts between nodes are missed because the nodes are in power saving mode. Hence, detected intercontact times, defined as the time between two consecutive contact events during which a communication can take place for a pair of nodes, are longer than intercontact times determined only by mobility, when a duty-cycling policy is in place. This heavily affects the delay experienced by messages, since the main contribution to message delay is in fact due to the intercontact times. For this reason, we believe it is of paramount importance to understand how power saving techniques, which may effectively reduce the number of usable contacts, affect the intercontact time [21].[3]

4.1 Problem statement

We assume that user mobile devices alternate between ON and OFF states, whose duration is fixed. We denote as duty cycle Δ the ratio between the duration of the ON and OFF states, and as T their sum. We assume that when a node is in the ON state it is able to detect contacts with other nodes. Please refer to [21] for a discussion on how to apply this model to popular technologies such as Bluetooth and WiFi Direct. For the sake of simplicity, coarse synchronisation (e.g., controlled by the cellular infrastructure in the case of mobile data offloading) can be used to guarantee that ON intervals overlap between any pair of nodes, such that they can communicate during a contact if this overlaps with their ON phases. In the following we model the duty cycle function as in Equation 4 and we assume that the first ON interval of d starts at s_0 and ends at s_1 (with $\tau = s_1 - s_0$).

$$d(t) = \begin{cases} 1 & \text{if } t \bmod T \in [s_0, s_1] \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

In the following, we focus on the intercontact process between a generic pair of nodes A, B and, to make the analysis more tractable, we assume that a contact event is detected only if it starts during an ON period. This is reasonable, since the duration of a contact is typically one-two orders of magnitude smaller than the intercontact time [22, 23], hence the probability that the contact lasts until the next ON interval is negligible. We assume that the time between two consecutive contacts between the same pair of nodes can be modelled as a continuous random variable S, and that intercontact times between a given pair of nodes are independent and identically distributed (while they can follow different distributions for different pairs). Hence, the contact process can be modelled as a renewal process, where $S_i \sim S$ denotes the time between the i-th and the $(i+1)$ contact event. Similarly, we denote with $\tilde{S}$ the random variable representing the detected intercontact times, and with $\tilde{S}_i \sim \tilde{S}$ the time between the i-th and the $(i+1)$ detected contact event (Figure 6). In the following, without loss of generality, we assume that there is a contact event at t_0 during the first ON period after $t = 0$. Consider the case in which $i-1$ contacts are missed after the one happening at t_0 and the i-th is detected. If we neglect contact duration, it is clear

[3] Reference available in the appendix.

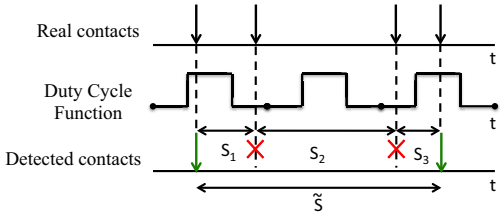


Figure 6: Contact process with duty cycling

that the time between the two detected events is given by the sum of the interarrival times of the events up to the i-th. Thus, for $\tilde{S}$, the following definition holds.

Definition 1. *The detected intercontact time $\tilde{S}$ can be obtained as $\tilde{S} = \sum_{i=1}^{N} S_i$, where N is the random variable describing the number of contacts needed to have one contact detected.*

Therefore, $\tilde{S}$ is a random sum of i.i.d. variables. Note that Definition 1 is general, i.e., holds for any type of continuous intercontact time distribution and for any type of duty cycling policy.

4.2 The case of exponential intercontact times

According to Definition 1, in order to derive the detected intercontact time, we first need to model N and then to compute the sum $\sum_{i=1}^{N} S_i$. In the next sections we will briefly discuss these two points (all proofs and additional results can be found in [21]).

4.2.1 Computing N

In [21] we have derived the probability distribution of N, defined as the number of contacts needed in order to detect the first one. Specifically, we have provided a general formulation (i.e., holding for any distribution of intercontact times S_i) for the PDF of N under the assumption that the probability that two undetected contacts fall in the same OFF interval is very low. Unfortunately, this expression is not in a closed form and, in general, finding a closed form for the distribution of N might be prohibitive (please note that numerical solutions can still be obtained). However, when intercontact times are exponential, a closed form solution is available, as shown below.

Theorem 1 (N with exponential intercontact times)**.** *When real intercontact times S_i are exponential with rate[4] λ, the probability density of N can be approximated by:*

$$\begin{cases} P\{N=1\} = 1 + \frac{e^{-\lambda\tau}-1}{\lambda\tau} + \frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{\lambda\tau(e^{\lambda T}-1)} \\ P\{N=k\} = e^{\lambda\tau}\frac{(1-e^{-\lambda\tau})^2}{\lambda\tau(1-e^{-\lambda T})}\left[\frac{\lambda(T-\tau)}{e^{\lambda T}-1}\right]^{k-1}, \quad k \geq 2 \end{cases} \tag{5}$$

The above result holds when the probability of two consecutive contacts happening during the same OFF period is small. Due to this approximation, the PDF of N does not adds up to 1 but remains below. The closer the total probability to 1, the smaller the error that we make. Exploiting these considerations and the result in Theorem 1, it is possibile to compute the error function $\mathcal{E}(\tau, T, \lambda)$ as follows.

Corollary 1. *The error $\mathcal{E}$ (defined in $[0,1]$) introduced by the approximation in Theorem 1 can be expressed as:*

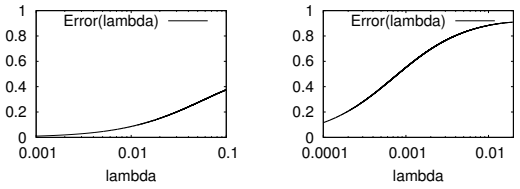
$$\mathcal{E}(\tau,T,\lambda) = \frac{1-e^{-\lambda\tau}}{\lambda\tau} - \frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{\lambda\tau(e^{\lambda T}-1)} - \frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{1-e^{-\lambda T}}\frac{T-\tau}{\tau(e^{\lambda T}-1-\lambda T+\lambda\tau)}. \tag{6}$$

[4]For ease of notation, we omit subscript A, B for λ, since it is unambiguous that we are referring to the tagged node pair A, B. Please note, however, that the network model we are referring to is still heterogenous.

Let us now derive a condition under which the above error goes to zero. Intuitively, from a physical standpoint, our approximation holds when the average intercontact time is significantly larger than the duration of the OFF interval. Under this condition, in fact, the OFF interval is, on average, too short to accommodate two intercontact times. Thus, the condition under which we expect our approximation to be very accurate is $E[S] \gg T - \tau$, or, equivalently, $\frac{1}{\lambda} \gg T - \tau$. The worst case for this condition happens when τ is very small. In fact, when $\tau \to 0$ the OFF interval can be very large (i.e., occupying almost all T) and, consequently, the probability that two undetected contacts fall into it maximum. Thus, if we ensure that $\frac{1}{\lambda} \gg T$ our condition will hold. We have mathematically proved this result in [21], from which the lemma below follows.

Lemma 1. *When $\lambda T \ll 1$, the error $\mathcal{E}$ introduced by the approximation of Theorem 1 approaches zero.*

For the sake of example, let us now explore the parameter space τ, T, λ in order to characterise how $\mathcal{E}$ goes to zero when the condition introduced above is satisfied. At first we set $\tau = 5$ and $T = 15$ (as in the RollerNet experiment – see Section 4.2.3) and we plot $\mathcal{E}$ varying λ (Figure 7(a)). As $\frac{1}{\lambda}$ represents the mean intercontact time $E[S]$, the smaller λ the bigger the mean intercontact time. And in fact, the error goes to zero as $\lambda \ll \frac{1}{T-\tau} = \frac{1}{10}$ and our condition is confirmed. We now keep the same τ value and increase T, setting it to $120s$, which is the value used for the Infocom experiment (Section 4.2.3). We expect from condition (i) that the error increases for small λ values with respect to the previous case, and this is confirmed in Figure 7(b).



(a) $\tau = 5$ and $T = 15$ (b) $\tau = 5$ and $T = 120$

Figure 7: $\mathcal{E}$ varying λ (logarithmic scale for the x-axis)

4.2.2 Computing the detected intercontact times

Now that we have found the distribution of N, we discuss how to compute the first and second moment of the detected intercontact time $\tilde{S}$ for a generic node pair A, B. The relation between S and $\tilde{S}$ is stated by Definition 1, i.e., $\tilde{S} = \sum_{i=1}^{N} S_i$. Thus, $\tilde{S}$ is a random sum of random variables, and we can exploit its well known properties to compute its first and second moment.

Proposition 1. *The first and second moment of $\tilde{S}$ are given by $E[\tilde{S}] = E[N]E[S]$ and $E[\tilde{S}^2] = E[N^2]E[X]^2 + E[N]E[X^2] - E[N]E[X]^2$.*

While the above formula holds in general, in the case of exponential intercontact times it is possible to derive an even stronger result, described in Theorem 2 below. This result is the key derivation of this work, and it tells us that, under condition $\lambda T \ll 1$, exponential intercontact times are modified by duty cycling only in terms of the parameter of their distribution but they still remain exponential.

Theorem 2. *When $\lambda T \ll 1$, the detected intercontact times $\tilde{S}$ follow approximately an exponential distribution with rate $\lambda\Delta$.*

4.2.3 Validation

In this section, we consider the average rate measured in real datasets of human mobility and we verify (i) whether our assumption $\lambda T \ll 1$ is reasonable and (ii) whether our model correctly predicts

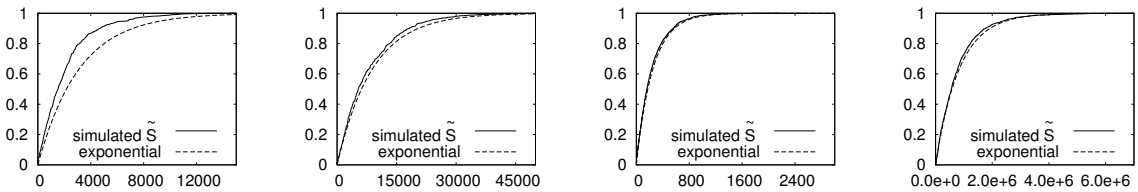
the distribution of the detected intercontact times assuming that the original intercontact times were exponential.

For opportunistic networks, there are several publicly available datasets obtained from experiments monitoring contacts between device pairs, and in almost all these experiments some simple duty cycling policies similar to the one described in Section 4.1 were implemented. Here we consider four popular datasets often used in the related literature: Infocom05 [22], Infocom06 [22], RollerNet [24], and Reality Mining [25]. In Table 4 we report the average pairwise rate extracted from these traces (corresponding to $\tilde{\lambda}$, i.e., to rates after duty cycling) and the associated λ values obtained applying Theorem 2, under the assumption that the underlying ICT are exponential[5] In addition, in Table 4 we also highlight the duration of the ON period and the period T of the duty cycling process. Using these parameters, we are able to compute λT and check whether our approximation holds (due to space reasons, here we perform an average analysis, while a pairwise analysis is provided in [21]). As it can be seen in Table 4, λT is smaller than 1 in all cases. Clearly, the farther from 1 the better, since we require $\lambda T \ll 1$. Thus, we expect the approximation that we make to be quite good for all datasets except for the Infocom05.

Table 4: Quality of approximation in popular datasets

Dataset	T	τ	$\tilde{\lambda}$	λ	λT
Infocom05	120	5	$3.2 \cdot 10^{-4}$	$7.7 \cdot 10^{-3}$	0.92
Infocom06	120	5	$1.13 \cdot 10^{-4}$	$2.7 \cdot 10^{-3}$	0.33
RollerNet	15	5	$4.07 \cdot 10^{-3}$	$1.2 \cdot 10^{-2}$	0.18
Reality	300	5	$1.2 \cdot 10^{-6}$	$7.2 \cdot 10^{-5}$	0.02

In order to complement the theoretical analysis presented above, here we verify that our prediction for the distribution of $\tilde{S}$ actually matches simulation results exploiting the parameters of real experiments. Specifically, we take a tagged node pair and we assume that the meeting rate of this pair corresponds to the average meeting rate (average across all pairs of nodes in the traces) measured from the traces in Table 4. With this approach we are able to represent the behavior of the average node pair. Then, we draw 10000 samples (100000 for the Reality Mining case, due to the long duty cycle period which led to fewer detections) from an exponential distribution, configured with the parameters λ in Table 4. The sequence of these samples corresponds to the contact process between the tagged node pair. To this contact process we apply a duty cycling function with $\Delta = \frac{\tau}{T}$, where τ and T are taken again from Table 4. Then we measure $\tilde{S}$ after each detected contact and we plot its CDF for all the four datasets in Figure 8. As expected, for the Infocom05 scenario, there are discrepancies between the actual and predicted values. For the other scenarios, in which the product λT is closer to zero, model prediction are very close to simulation results.



(a) Infocom05 (b) Infocom06 (c) RollerNet (d) Reality Mining

Figure 8: CDF of $\tilde{S}$ in the different scenarios

[5]While this assumption is very strong in general, Tournoux et al. [24] have shown that it is acceptable for a significant percentage of pairs in some of these datasets.

4.3 The effect of duty cycling on the delay

The great advantage of knowing how intercontact times are modified by duty cycling is that their distribution that we have derived can be fed to one of the modelling frameworks available in the literature in order to compute some important metrics for opportunistic networking. Theorem 2 tells us that the detected intercontact times are still exponential, so we can choose any of the many modelling frameworks exploiting exponential intercontact times. In this work we have focused on deriving the first two moments of the delay and to this aim we have relied on the model that we have presented in [26]. The goal of this evaluation is to study how the first two moments of the delay are affected by energy saving techniques. Please note that while the analysis in the previous section focused on a tagged pair of nodes, in this section we study the whole network. So, assuming exponential intercontact times, we denote their rates as λ_{ij} for node pair i, j.

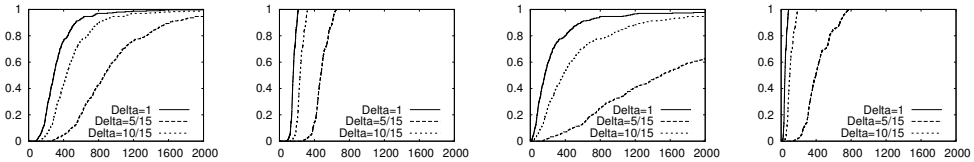
The forwarding model that we exploit represents the forwarding process in terms of a Continuous Time Markov Chain (CTMC). The chain has as many states as the nodes of the network and transitions between states depend both on the meeting process between nodes (i.e., their intercontact times) and on the forwarding protocol in use. Denoting the delay of messages from a generic node i to a tagged node d as D_i, and using standard Markov chain theory, it is possibile to derive the first two moments of D_i as in [26]. In [26], we have defined a set of abstract policies able to capture significant aspects of popular state-of-the-art forwarding strategies. In the following we will focus on two of these policies, the Direct Transmission and the Direct Acquaintance (the analysis for other policies is provided in [21]). Under the Direct Transmission (DT) forwarding scheme, the source of the message is only allowed to hand it over to the destination itself, if ever encountered. The Direct Acquaintance (DA) is a social-aware policy, in which each intermediate forwarder hands over the message to nodes that have a higher probability of bringing the message closer to the destination, according to some predefined forwarding metrics. With Direct Acquaintance the forwarding metric is the contact rate with the destination ($\frac{1}{E[S_{id}]}$): a better forwarder is one with a higher contact rate with respect to the node currently holding the message.

In the following we assume that nodes intercontact times are exponential. The fitting analysis presented in [23] has shown that contact rates in the traces already considered in Section 4.2.3 follow a Gamma distribution. Below, we focus on the distribution parameters for the RollerNet scenario reported in [27], i.e., shape $\xi = 4.43$, rate $r = 1088$. We consider a network made up of 25 nodes and we solve the forwarding model described above in the case of duty cycle equal to $\frac{5}{15}$, $\frac{10}{15}$ and 1 (no duty cycling). Figures 9(a)-9(d) show the CDF of the moments of the delay in this case. As expected, both the first and second moment become larger as we reduce the ON interval in the duty cycle. In fact, as discussed before, the neat effect of duty cycling is to effectively reduce the number of usable contacts to only those happening during an ON period. The shorter the ON period, the fewer the usable contacts every T, the longer the delay. Let us now see what happens to the coefficient of variation of the delay. From Figures 9(e)-9(f) it can be seen that the coefficient of variation can be either bigger or smaller than one. This means that the delay can be approximated with a hyper-exponential or a hypo-exponential distribution [28]. In the next section, we will use this representation of the delay in terms of the hyper-exponential or hypo-exponential distribution in order to compute the volume of traffic carried by the network. Another interesting observation from Figures 9(e)-9(f) is that the coefficient of variation does not depend on the duty cycle Δ (in fact, all curves overlap). This means that, in the case of exponential intercontact times, the duty cycle does not affect the variability of the delay experienced by messages.

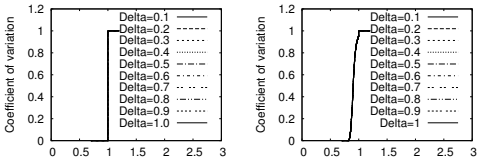
4.4 Energy, traffic, and network lifetime

In this section we investigate the benefits of implementing a duty cycling policy on the nodes of the network. These benefits are in terms of energy saved by nodes and, consequently, increased network lifetime. In fact, assuming that nodes have an energy budget L (expressed as the amount of time they can be on when no duty cycling is implemented), by definition nodes' lifetime is extended to $\frac{L}{\Delta}$ when energy saving strategies are in place[6]. At the same time, however, as we have seen in the previous section,

[6] We consider only the part of the energy budget related to networking activities.



(a) First moment - DT (b) First moment - DA (c) Second moment - DT (d) Second moment - DA



(e) Coefficient of variation - DT (f) Coefficient of variation - DA

Figure 9: CDF of the delay moments and coefficient of variation

the gain in terms of energy is counterbalanced by a loss from the delay standpoint. In fact, the expected delay increases as Δ decreases, so the network lifetime is longer but nodes also need more time to deliver messages. In [21] we have studied the following three aspects. First, what is the relationship between the energy consumed with and without duty cycling. Second, what is the volume of traffic carried by the network with and without duty cycling. Third, whether there exists an optimal duty cycle value for which the loss in terms of delay is minimum and the gain in terms of traffic carried by the network is maximum.

We first assume that messages are all generated at time $t = 0$ and that L is very large. The goal here is to understand how much energy is saved by duty cycling, without considering the limited network lifetime, i.e., just taking into account the delivery of standalone messages without temporal limitations. Throughout the section we use a simple energy model in which nodes consume a certain power w (measured in watts) during ON intervals and zero otherwise. Below we focus on the expected delay across the whole network, for having a compact representation. Please note however that the behaviour is the same for the single pairs of nodes. We measure the energy consumed as the product between power w and the length of the time interval for which the network is ON. Without duty cycling, the network is ON for the whole time it takes to deliver a message (hence, for $E[D]$), while in case of duty cycling, the network is ON only for a fraction Δ of the time ($E[D_\Delta]$, where we denote with D_Δ the delay under duty cycling Δ) it takes to complete the delivery. In order to measure the relationship between the two quantities, we study the following:

$$R_\Delta = \frac{wE[D]}{w\Delta E[D_\Delta]} = \frac{E[D]}{\Delta E[D_\Delta]}. \tag{7}$$

We find that, exploiting the same parameters of the RollerNet scenario used above, the ratio stays around 1 independently of the specific duty cycle value Δ (the plot can be found in [21]). This result is very interesting, because it shows that under exponential intercontact times the energy consumed for a standalone message is the same, regardless of the value of Δ. With $\Delta = 1$ the energy budget needed for sending the message is simply used all at once, while with $\Delta < 1$ this budget is spread across different intervals of duration T, during which the nodes of network are partially turned off. This is consistent with the assumption of unlimited network lifetime and standalone messages (all generated at $t = 0$) that we have made so far.

For a more realistic evaluation, we hereafter assume that messages are generated according to a Poisson process with rate μ. By definition, the number of messages created in disjoint timeslots are independent. Thus, the number of messages arriving during a time interval of length dt is given by μdt. When $\Delta = 1$ these messages keep arriving until L, after which the network has exhausted all its energy

budget and turns off indefinitely. Instead, when $\Delta < 1$ the network takes longer to consume all its energy budget, thus stays on until $\frac{L}{\Delta}$. In the following we study the volume of messages delivered by the network with and without duty cycling, measured as the number of messages delivered in $\frac{L}{\Delta}$. Then, the following result holds, whose proof can be found in [21].

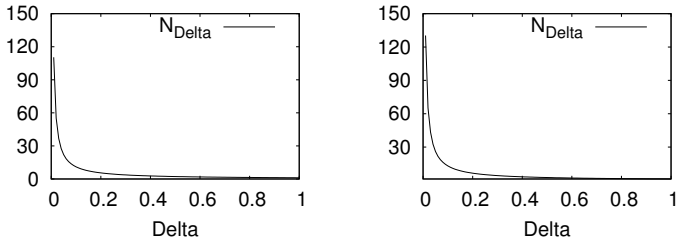
Theorem 3. *When the coefficient of variation of the delay c belongs to the interval $[\frac{\sqrt{2}}{2}, 1]$, the volume N_Δ of messages delivered by the system under duty cycling Δ is given by:*

$$N_\Delta = \frac{\mu L}{\Delta} - \mu E[D_\Delta] \cdot \left\{ 1 - \frac{1}{4\alpha} \left[(1+\alpha)^2 e^{-\mu_1 \frac{L}{\Delta}} - (1-\alpha)^2 e^{-\mu_2 \frac{L}{\Delta}} \right] \right\}, \tag{8}$$

where $\alpha = \sqrt{1 + 2(c^2 - 1)}$, $\mu_1 = \frac{2}{E[D]}(1+\alpha)^{-1}$, $\mu_2 = \frac{2}{E[D]}(1-\alpha)^{-1}$.
When the coefficient of variation of the delay c is smaller than one, the volume N_Δ of messages delivered by the system under duty cycling Δ is given by:

$$N_\Delta = \frac{\mu L}{\Delta} - \mu E[D_\Delta] \left[1 - \frac{1}{2} e^{\frac{-L}{E[D_\Delta]\Delta}} \left(e^{\left(1+\sqrt{\frac{c^2-1}{c^2+1}}\right)} + e^{\left(1-\sqrt{\frac{c^2-1}{c^2+1}}\right)} \right) \right]. \tag{9}$$

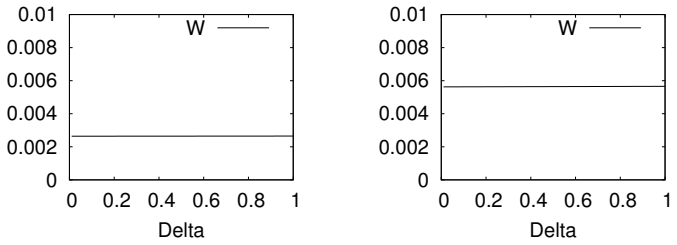
Basically, N_Δ is given by the number of messages generated during the network lifetime (the first term in the above equations minus the number of messages that are not delivered before the energy budget is depleted. The latter quantity is a function of the expected value of the delay and of its variability. We now exploit Theorem 3 in order to study N_Δ. Specifically, in Figure 10 below we show how N_Δ varies with different duty cycles, where we assume that each node generates 1 message every 10 minutes (so $\mu = \frac{1}{600}$). The plot is drawn for a tagged node pair for the sake of readability, but the same results hold for the other pairs. We see that the volume of traffic carried by the network (i.e., the number of messages delivered on average during network lifetime) increases as the duty cycle Δ decreases. So, as expected, increasing the network lifetime more messages get a chance of being delivered but the price to pay, as seen in Section 4.3, is a larger expected delay.



(a) DT (b) DA

Figure 10: N_Δ varying Δ with different forwarding algorithms

In the final part of our evaluation, we study whether it exists an optimal working point that minimises the expected delay and maximises the volume of messages exchanged. We borrow the definition of *power of the network* (which we denote with $\mathcal{W}$) from traditional queueing theory [29]. Quantity $\mathcal{W}$ measures the trade-off between the traffic N_Δ carried by the network (function of the message injection rate μ) and the expected delay $E[D_\Delta]$. The power is then defined as $\mathcal{W} = \frac{N_\Delta}{D_\Delta}$. In traditional queueing systems, the above trade-off was regulated by contention. In fact, under limited resources, we could not increase indefinitely the quantity of messages successfully delivered without affecting the resulting expected delay (because, e.g., under heavy traffic, packets start to be discarded from buffers). In our case, we do not have contention, since we assume that there are no limitations on buffers and bandwidth. Our knob is instead the duty cycle. When Δ approaches 1, delays are as short as possible given the underlying mobility, but a lot of energy is consumed and the network lifetime is shorter. If we want to increase network lifetime, we have to sacrifice the expected delay.



(a) DT (b) DA

Figure 11: $\mathcal{W}$ varying Δ in the different forwarding algorithms ($\mu = \frac{1}{600}$)

Figure 11 shows how $\mathcal{W}$ varies depending on Δ. It can be clearly seen that $\mathcal{W}$ remains practically constant, which implies that whatever one gains in network lifetime is immediately lost in expected delay. Thus, under exponential intercontact times, there is no optimal choice of Δ and all working points are equivalent. When configuring a duty cycling policy, the operator can thus only maximise one metric at a time. Note, anyway, that if the total number of messages to be delivered is of primary concern, at the cost of additional delay, then reducing the duty cycling is clearly effective.

4.5 Final remarks

In this section we have investigated the effects of duty cycling on intercontact times, delay, and energy consumption in opportunistic networks. To this aim, we have provided a general formula for the derivation of the intercontact times under duty cycling, and we have specialised this formula obtaining a closed-form expression for the case of exponential intercontact times. Surprisingly enough, under condition $\lambda T \ll 1$ satisfied by most popular contact datasets, the intercontact times after duty cycling can be approximated as exponentially distributed with a rate scaled by a factor $\frac{1}{\Delta}$. Exploiting this result, we have then studied the first two moments of the delay under duty cycling, showing that these moments both increase as Δ becomes smaller. Finally, we have focused on how the network lifetime is impacted by duty cycling, highlighting the fact that a larger volume of traffic is handled by the network when a duty cycling policy is in place, because the network lifetime is increased even if at the expense of the delay experienced by messages. In addition, we have found that it is not possible to derive an optimal duty cycle value that maximises the network lifetime while at the same time minimally impacting the expected delay.

As future work, two main different directions can be pursued. First, we plan to extend the analysis to intercontact times featuring a distribution different from the exponential, e.g., the Pareto intercontact times case, which also represents a popular hypothesis for the intercontact times considered in the literature [30]. Second, we also plan to evaluate different duty cycling policies, e.g., some in which the duration of ON and OFF intervals is not fixed but varies according to a specified distribution.

5 Contacts and intercontacts beyond one hop

A common point among these works is that they rely on the binary assumption that nodes are either in contact or in intercontact. *We revisit this assumption and provide arguments toward the adoption of an extended view of the neighborhood.* Let us motivate our work through the example shown in Fig. 12. This figure represents a snapshot where nodes in "group 1" are in contact with A (i.e., they are within A's communication range). In the usual binary vision, all remaining nodes are, by definition, in intercontact. Still, we notice that there is a fundamental difference among nodes in "group 2". None of the nodes in "group 2.a" are in contact with A; nevertheless, they do have a contemporaneous path to A. On the other hand, nodes in "group 2.b" do not have any path to A. In opportunistic networking, where we need to gather as much knowledge as possible to achieve efficient communication standards, deeming both cases

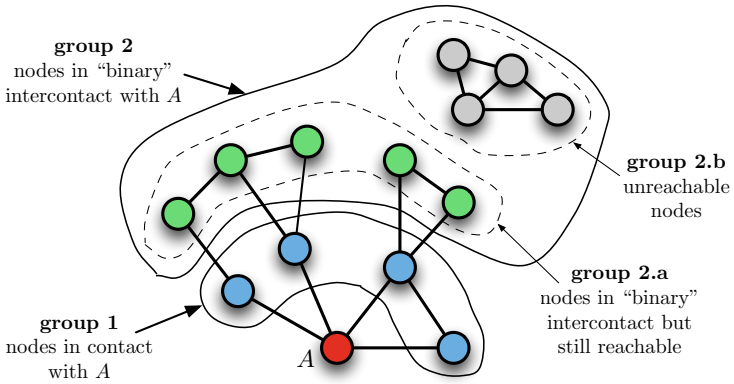


Figure 12: Motivating example. From node A's point of view, we see that nodes in "group 1" are in contact. Using the usual binary vision, we conclude that all other nodes are in intercontact (i.e., in "group 2"). However, nodes in "group 2.a" are essentially different from nodes in "group 2.b". A has end-to-end paths toward the first and no paths at all to the latter.

of intercontact under the same definition results in waste of information. Suppose A needs to send a message to one of the nodes in group 2.a. With the binary vision, A does not know that the destination is nearby, and may miss an opportunity to communicate if, for example, the destination moves after some time to group 2.b.

We provide evidences that considering nodes that are not in contact, but still reachable, can be of great help in the design of efficient opportunistic networks. For instance, we show that in some experiments more than half of the pairs of nodes do come within each other's vicinity but do not get in direct contact. Understanding what occurs within this zone as well as between this zone and the rest of the network is a key to understand the network. To this end, we formally define in this work the notion of κ-vicinity, which is the extended neighborhood of a node containing all the nodes within κ-hop reach [31].[7]

As a summary, our contributions are:

- **Vicinity definition in opportunistic networks.** We formalize the vicinity concept in DTN and define temporal measures to characterize the interactions within the vicinity and with the full network.
- **The revelation of the binary assertion.** We illustrate our plea by showing the importance of missed transmission possibilities when considering only contacts in opportunistic techniques.
- **κ-vicinity analysis and a rule of thumb.** We analyze the composition of the vicinity through various parameters and provide a way to infer vicinity size using only contacts.

We produce results for a wide variety of datasets, from real-world to synthetic ones. Each of them reflects a situation where opportunistic networking could happen. For instance, we observe how the κ-contact and κ-intercontact distributions also follow power laws up to a characteristic time then exponentially decay. We also show how local density may influence these distributions. To reduce vicinity monitoring costs, we also hint how monitoring a $\{3,4\}$-vicinity is enough to catch most meaningful events in a node's vicinity.

5.1 Defining a new vicinity for opportunistic networks

To formalize the vicinity notion in DTN, we choose to use a node-centered point of view, i.e., they do not need to rely on any global knowledge to gather vicinity information. The κ-vicinity notion also adds

[7] Reference available in the appendix.

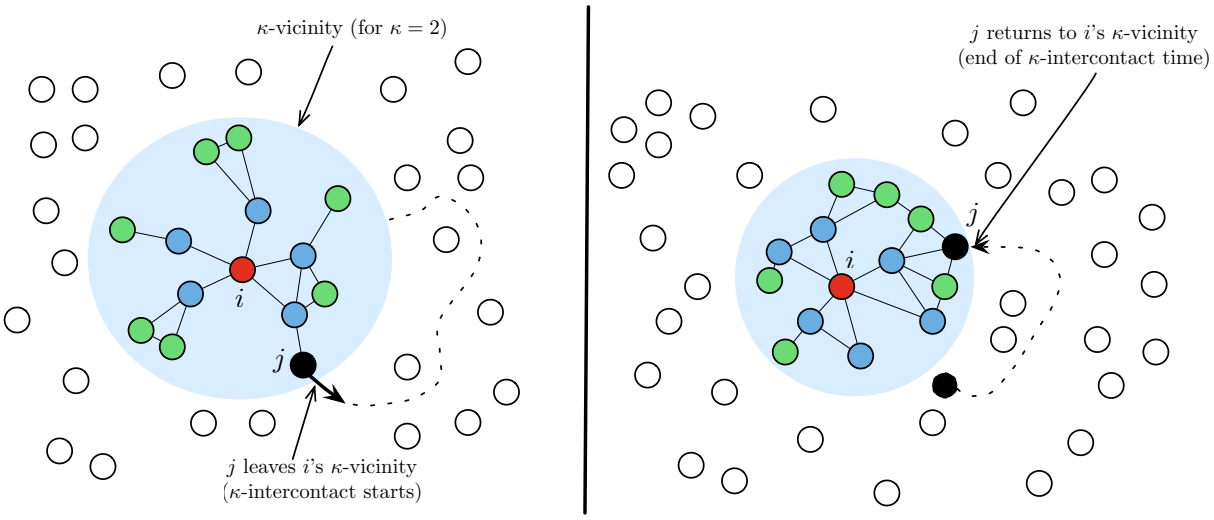


Figure 13: Node i's κ-vicinity and the κ-intercontact phenomenon. For the sake of clarity, we only display i's connectivity links within the κ-vicinity.

a hop-based discrimination. This differentiation helps us limit our vision according to our needs as well as identify neighbor properties. We discriminate a node i's vicinity according to the number of hops between i and its surrounding neighbors.

Definition 2. ***κ-vicinity.*** *The κ-vicinity $\mathcal{V}_\kappa^i$ of node i is the set of all nodes whose shortest distance to i is of at most κ hops.*

This definition implies that $\mathcal{V}_{\kappa-1}^i \subset \mathcal{V}_\kappa^i$. In Fig. 13, we illustrate the 2-vicinity of node i. The κ-vicinity brings the immediate surrounding knowledge. This is an interesting point of view for opportunistic networks because it extends a node's knowledge to immediately useable communication opportunities. The κ-vicinity empowers a node's reach in the network [32].

Vicinity knowledge may come from different techniques. For instance, we can use link state protocols to gather information about a node's connected component. There are many ways to do so, but they all are costlier than getting information from contacts only. Of course, there is a tradeoff between knowledge and overhead to collect vicinity information. The good news is that, as we will show later, even only a few hops around the node (typically 2–4 hops) significantly helps getting representative behavior of the node's surroundings.

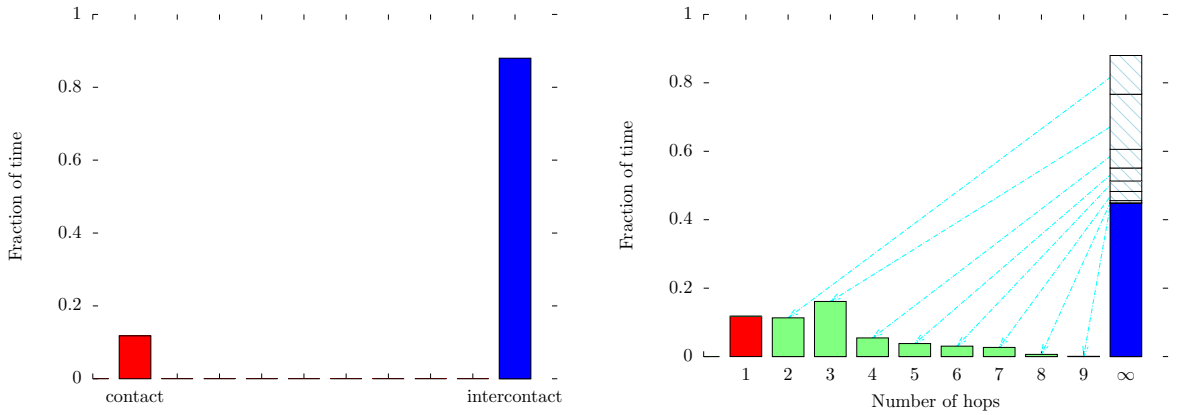
The κ-vicinity defines a node's neighborhood, its new *zone* in the network. To characterize this zone relationships to node i, we must define some temporal measures relating to times neighbors spend in the zone and times outside the zone, namely κ-contact and κ-intercontact. We maintain a pairwise definition for these measures.

Definition 3. ***κ-contact.*** *Two nodes are in κ-contact when they dwell within each other's κ-vicinity, with $\kappa \in \mathbb{N}^*$. More formally, two nodes i and j are in κ-contact when $\{i \in \mathcal{V}_\kappa^j\}$ and $\{j \in \mathcal{V}_\kappa^i\}$.*

In other words, there is a contemporaneous path of length at most κ links i and j. We also need to grasp the intercontact observations for our vicinity viewpoint. The literature definition of mere intercontact is when two nodes are not in contact. Therefore, we consider κ-intercontact when two nodes are not in κ-contact. These are complementary notions. Another way to see it is as follows: if node i maintains knowledge about its κ-vicinity, it is in κ-intercontact with any node beyond its κ-vicinity. In Fig. 13, node j leaves i's κ-vicinity and then gets back some time later, characterizing a κ-intercontact interval.

Definition 4. ***κ-intercontact.*** *Two nodes are in κ-intercontact while they do not belong to each other's κ-vicinity. Formally speaking, two nodes i and j are in κ-intercontact when $\{i \notin \mathcal{V}_\kappa^j\}$ and $\{j \notin \mathcal{V}_\kappa^i\}$. There is no path of length κ or less linking i and j.*

Note that 1-contact matches the contact notion and 1-intercontact corresponds to usual binary intercontact. Our zone point of view integrates previous binary network vision and also extends it via nearby nodes. The binary assertion misleads our vision of a network. For huge networks, considering the binary abstraction that when nodes are not in contact, they are in intercontact and devoid of path between them may be true in most cases but it is wrong in few cases. These few cases are the ones opportunistic networking should leverage. In the next section, to better understand the binary assertion, we will focus on pairwise relationships.



(a) Binary assertion

(b) Real vision

Figure 14: Example of time-distance distribution from the *RT* dataset. In Fig. 14(a), nodes spend 10% the time in contact (1-hop). With the binary vision, we then consider that nodes spend around 90% of their time in intercontact. Fig. 14(b) shows that in reality, they dwell at a distance 2 for around 10%, at a distance 3 for 16%. Real intercontact deprived of multi-hop path represents only 50% of the time (∞).

5.2 The limits of the binary assertion

Before investigating where the binary assertion fails to reflect the proximity structure in opportunistic networks, let us describe the datasets that we will use to illustrate our findings.

5.2.1 Datasets

For our study, we observe the binary assertion and vicinity properties in real-world experiments as well as synthetic datasets (denoted by (S)) described hereafter. We use realistic measurements to observe the extent of vicinities in real-life situations. We also confronted the vicinity notion to synthetic datasets to observe its presence in dedicated mobility patterns.

Infocom05 measurement was held during a 5 day conference in 2005 [33]. 41 attendees carried iMotes collecting information about other iMotes nearby within a 10 m wireless range. We study a 12-hour interval bearing the highest networking activity. Each iMote probes its environment every 120 seconds. *Infocom05* represents a professional meeting framework.

Sigcomm09 counted 76 attendees with dedicated smartphones probing their surroundings during 5 days [34]. Smartphones sensed their surroundings using Bluetooth every 120 seconds. *Sigcomm09* is another example of a professional meeting scene.

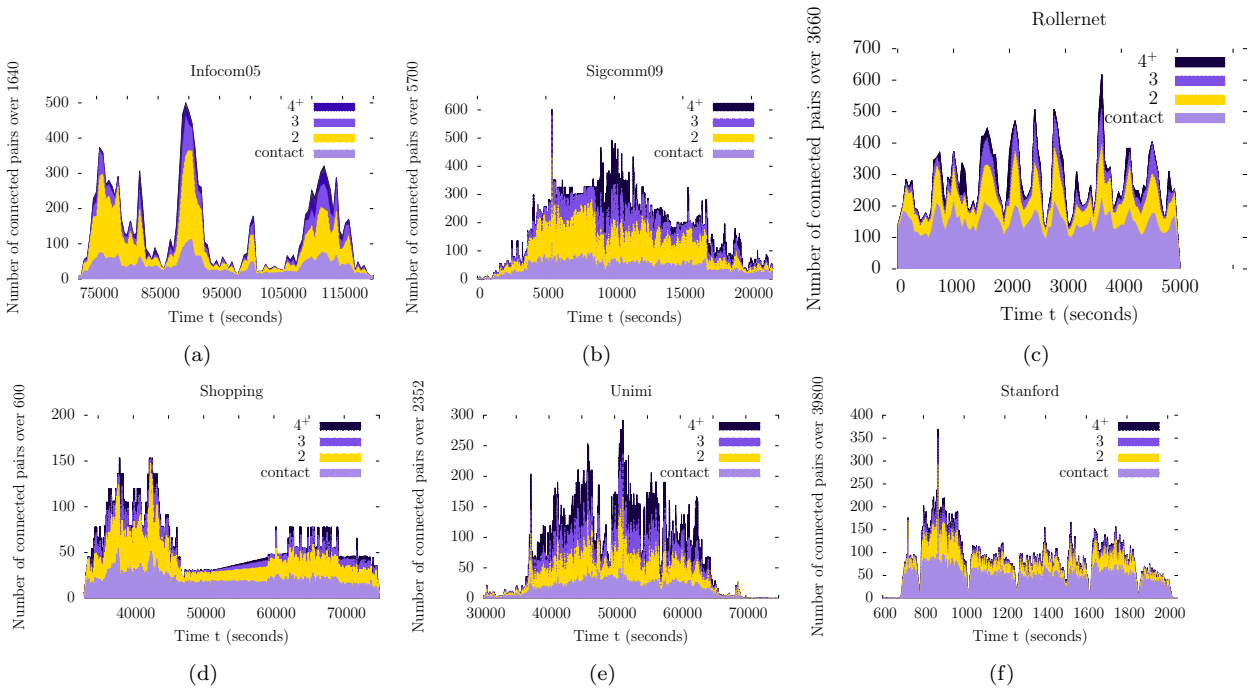


Figure 15: Datasets sociostructures presenting the amount of pairs connected by contacts, two-hop distance, three-hop and so on in a layered mode according to time. We notice the omnipresence of pairs connected by two or more hops. They often overcome the possibilities offered by contact only (bottom layer). As a result, contact opportunities only represent a minor part of all end-to-end opportunities between two nodes. The binary assertion overlooks these possibilities by blending all nodes in intercontact under a unique concept.

Rollernet had 62 participants measuring their mutual connectivity with iMotes during a 3 hour rollerblading tour in Paris [35]. These iMotes sent beacons every 30 seconds. This experiment shows a specific sport gathering scenario.

Shopping used 25 dedicated devices in a shopping mall over 6 days [36]. Galati and Greenhalgh gave 25 devices to shop owners and planted 8 others at various locations in the mall. Devices performed neighborhood discovery every 120 seconds. *Shopping* reflects the working day routine of shop owners as well as some of their customers.

Unimi is a dataset captured by students, faculty members, and staff from the University of Milano in 2008 [37]. The experiment involved 48 persons with special devices probing their neighborhood every second. *Unimi* provides a scholar and working environment scenario.

Stanford used TelosB motes – detecting contacts up to a 3 m range [38]. We focus on a set of 200 participants in an U.S. high school. TelosB motes send beacons every 20 seconds. *Stanford* expresses a settings with a majority of teenagers who have a tendency to dwell in groups of interests.

RT (S) for *RT* is a mobility model correcting flaws from the Random Waypoint model [39]. We sampled the behavior of 20 nodes following this model on a surface of 50 $\times$ 60 m^2 with speed between 0 and 7 m/s and a 10 m range. We choose to simulate 20 nodes over this surface to recreate office conditions.

Community (S) is a social-based mobility model [40]. It tends to collocate socially related nodes in specific positions at the same time like groups of friends would do. We simulated 50 nodes with a 10 m wireless range on a 1,500 $\times$ 2,500 m^2 plane during 9 hours. In *Community*, we tried to reproduce a city-wide setting.

Table 5: Datasets characteristics, where # indicates the number of nodes in the datasets, *Length* indicates its duration, *Probing* presents the probing frequency, *Range* provides the probing devices wireless range, and *Type* presents the scenario type.

Dataset	#	*Length*	*Probing*	*Range*	*Type*
Infocom05	41	12h	120s	10 m	Conference
Sigcomm09	76	4 days	120s	10 m	Conference
Rollernet	61	1h30	30s	10 m	Sport
Shopping	25	6 days	120s	10 m	Mall
Unimi	50	19 days	1s	10 m	Scholar
Stanford	200	1h	20s	3m	Scholar
RT (S)	20	9h	1s	10 m	Work
Community (S)	50	9h	1s	10 m	City

We recapitulate the main datasets characteristics in Table 5. In the following, we use some of the presented datasets to introduce and illustrate our vicinity concepts.

5.2.2 Binary assertion illustration

Fig. 14 illustrates an example of the binary assertion using the datasets described previously. For a given pair of nodes of the *RT* dataset, we compare the amount of time they spend in contact and in intercontact. They spend around 10% of their time in contact and around 90% in binary intercontact. Now let us consider the extended vision; for the same pair of nodes, they still spend 10% of their time in direct contact. However, they also spend around 10% of their time at a 2-hop distance, 18% at 3 hops, 5% at a 4-hop distance, and so on. The actual time they spend without any communication possibility from one to the other is only around 50% of the experiment duration.

More than just limiting our vision, the binary definition prevents us from leveraging our environments and performing simple yet efficient nearby end-to-end transmissions. A natural question that arises is whether our observation is peculiar or not.

5.2.3 Missed transmission possibilities

To quantify how many end-to-end transmission opportunities the binary assertion misses, we present what we call *aggregated network sociostructures* in Fig. 15 [41]. For each real life-based dataset, we plotted (in layered mode) the number of connected pairs for each shortest distance. This means that layer 2 shows the amount of pairs connected via a two-hop path, layer 3 represents connection via three-hop paths, and so on. The bottom layer symbolizes the amount of pair of nodes in contact. The binary assertion does not recognize such relations.

In Fig. 15(a), for *Infocom05*, we observe several density peaks of connected pairs. Being a conference-based measurement, these peaks indicate morning arrivals, lunch, afternoon break, and end of sessions. During high density peaks, an unexpected observation is the large proportion of pairs of nodes connected by two-hop paths. Places with high density ignite transmission possibilities beyond mere contact.

For *Sigcomm09*'s sociostructure, we focus in the first density peak of this conference dataset (see Fig. 15(b)). The number of pairs in contact remains non null during the observation. At some point, like 10,000 seconds, pairs linked by two or more hops represent more than four times the number of pairs currently in contact. As a result, in such a scenario, 2^+*-hop transmissions should be more helpful than mere contact transmissions* or pure DTN techniques.

In Fig. 15(c), we witness *Rollernet*'s accordion phenomenon, i.e., the stretching and shrinking of the crowd due to urban obstacles preventing the crowd from moving forward [35]. *Rollernet* has a dynamic setting with a compulsory path. Nodes do not have as much movement liberty as they have in other datasets. Contacts are prominent in *Rollernet*, but we still observe many pairs of nodes connected by two or more hops.

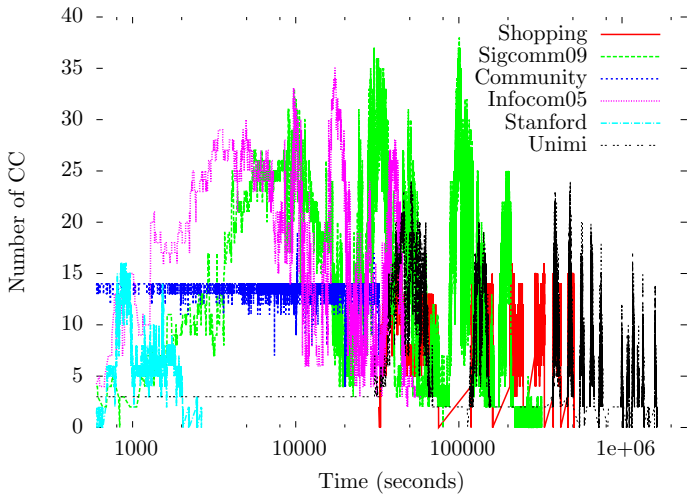


Figure 16: The presence of connected component (CC) of size higher than 2 according to time. As long as nodes find themselves in their "working" environment, they display numerous simultaneous CC. The conference datasets *Infocom05* and *Sigcomm09* due to their specific nature bear the highest number of concurrent CC (up to 38). The same pattern applies to other datasets but with a lower number of CC values. Still, they are most of time higher than 10. *Community* displays a peculiar number of CC evolution. Note the logscale on the x-axis.

For *Shopping* and *Unimi* (Fig. 15(d) and Fig. 15(e)), we notice the omnipresence of pairs connected at a two-hop distance and how they may overcome contact opportunities. In the *Unimi* dataset, we notice even more pairs connected by three or more hops.

The *Stanford* sociostructure from Fig. 15(f) occurs during a school day and shows the different groups found in high school. The majority of students stay close enough to be connected via contacts; still, we can observe the omnipresence of a significant share of two-hop links between nodes.

The aforementioned plots illustrate the limits of the binary assertion. If we maintain contact-only knowledge in DTN, we miss the omnipresent power of nodes at two or more hops. For all datasets, contemporaneous multi-hop paths are far from being negligible. These paths are potential transmission opportunities as they only involve few relays that could reduce significantly end-to-end delays. Considering only contacts provides a limited vision of what happens in the network. Observing a node's vicinity at a two-hop distance may, in some situations, more than double the transmission opportunities (as seen for example in the *Infocom05* dataset at 90,000 seconds).

We advocate that it is important to observe nodes beyond simple contacts. In the next section, we analyze κ-intercontact distributions to provide vicinity-related temporal characterization of the network.

5.3 κ-vicinity analysis

To uncover vicinity properties in disruption-tolerant networks, we begin by observing all network's connected components.

5.3.1 The seat of κ-vicinities: connected components

Network connected components are the loci of κ-vicinity. Studying a network's connected components is a good indicator of the κ-vicinity. Each node is its own connected component. We use the definition of connected components by Cormen et al. [42]. Since κ-vicinity need at least two nodes to make sense (i.e., a contact to happen), we will investigate connected components of size larger than two.

Table 6: Average size, diameter, and gravity $\mathcal{G}$ of dataset's largest connected component.

Dataset	Size (s)	Diameter (d)	Gravity ($\mathcal{G} = s/d$)
Infocom05	20.2	5.0	4.0
Sigcomm09	16.3	4.5	3.6
Rollernet	12.9	6.0	2.1
Shopping	11.0	4.0	2.7
Unimi	9.1	4.1	2.2
Stanford	6.7	3.2	2.1
RT	12.4	5.9	2.1
Community	13.8	3.6	3.8

Presence. In Fig. 16, we represent the number of connected components (CC) of size larger than two for each of the datasets. We first observe a wide range of the number of connected components. For each scenario, there is a clear evolution through time. The size of the CC may be close to zero at night when people are at home or outside the measurement environment. Conversely, when people find themselves at work, school or a conference, they bring density to the picture, contributing to the formation of numerous CC. We often observe more than 10 simultaneous different CC [43].

For both conference datasets (*Infocom05* and *Sigcomm09*), we get the highest number of simultaneous CC. The conference setting where many people stay in different rooms to listen to talks may explain this high value. People stay together at different places, therefore, we have many small groups instead of a bigger one. For the sake of clarity, we did not display the *Rollernet* dataset here (for information, *Rollernet* displays between 5 and 35 concurrent CC during its course). *Shopping* and *Unimi* have a conference-like behavior but at a lower level; they show periods of large numbers of CC but still less than the ones observed in conference datasets. The *Stanford* dataset has a smaller number of components, but this stems from the nature of the logging devices used in the experiment. *Community* has a quasi constant number of components, which may be due to the synthetic nature of the dataset. The *Community* mobility model generates group-based meeting patterns.

There are many simultaneous CC in the observed datasets. Their number show how widespread are the potential κ-vicinities.

Size and diameter. In opportunistic networks, there are no connectivity graph diameters as such. The diameter notion only applies on a fully connected graph. However, the notion of diameter is still meaningful for each network connected component. In Table 6 , we present the average size of the largest CC, its average diameter for each dataset, and their corresponding "gravity". The gravity $\mathcal{G}$ is defined as the ratio between the largest connected component average size (CC^{+}_{size}) and the largest connected component average diameter ($CC^{+}_{diameter}$):

$$\mathcal{G} = \frac{CC^{+}_{size}}{CC^{+}_{diameter}}. \tag{10}$$

The gravity is somehow an indicator of how concentrated the largest CC is. Datasets with gravity values ≥ 3, like *Infocom05*, *Sigcomm09*, and *Community* display largest CCs that are more condensed than in all the other datasets. The distribution of nodes within these CC is tighter than in other situations. Therefore, their proximity in terms of connectivity is stronger. They form components with stronger connectivity and allow resulting κ-vicinities to be more redundant link wise. Remaining datasets still bear a gravity ≥ 2, allowing numerous yet more loosely connected κ-vicinities.

5.3.2 κ-vicinities $\mathcal{V}^{i}_{\kappa}$ local density

With the κ-vicinity, we can measure the potential of such nearby companions in terms of opportunistic communications. Yet, obtaining a information about a node's vicinity comes with a cost. To reduce the tradeoff between additional vicinity information and gathering costs, we can wonder how far a node must probe its vicinity to obtain the maximum information with the lowest probing cost. We may ponder this

Table 7: Average number of neighbors $\mathcal{D}_\kappa$ in a node's κ-vicinity.

	κ							
Dataset	1	2	**3**	**4**	5	6	7	8+
Infocom05	2.4	6.7	9.5	10.7	11.1	11.3	11.4	11.4
Sigcomm09	1.0	2.6	3.7	4.2	4.4	4.4	4.5	4.5
Rollernet	1.2	2.4	3.4	4.0	4.3	4.5	4.6	4.7
Shopping	1.6	3.3	4.3	4.6	4.7	4.7	4.7	4.7
Unimi	0.6	1.3	1.8	2.1	2.3	2.3	2.3	2.3
Stanford	0.5	0.7	0.8	0.8	0.8	0.8	0.8	0.8
RT	2.2	4.3	6.3	7.6	8.4	8.8	9.0	9.0
Community	2.0	4.1	4.7	4.8	4.8	4.8	4.8	4.8

cost by limiting the scope of κ-vicinity to a given κ. To maximize the κ-vicinity utilization, we need to capture the most closeby nodes and events as possible. The more neighbors and events we observe, the better we can use them to perform opportunistic operations.

To get more neighbors, we must extend the κ value. The first aspect to analyze is each κ-vicinity local density. Let $\mathcal{D}^i_\kappa$ be the density of nodes around i, given by

$$\mathcal{D}^i_\kappa = \frac{card(\mathcal{V}^i_\kappa)}{\tau_t}, \tag{11}$$

where $card(\mathcal{V}^i_\kappa)$ is the number of nodes in i's κ-vicinity and τ_t is the sum of all moments where $card(\mathcal{V}^i_\kappa)$ was not null. κ-vicinity internal composition influences a node's κ-vicinity behavior. For a given probability p of having nodes at $\kappa + 1$ distance when there is a node at κ hops from i, the more κ-contacts a node has, the more chances it has of getting $\{\kappa + 1\}$-contacts. In Table 7, we present the average $\mathcal{D}_\kappa$ according to κ. For all datasets except *RT* and *Rollernet*, above a certain threshold ($\kappa_t = \{3, 4\}$), their $\mathcal{D}_\kappa$ does not increase anymore and is limited by the network diameter. More dynamic or inconsistent patterns – *RT* and *Rollernet* – display logarithmic increase in $\mathcal{D}^i_\kappa$. For all cases, we verify card$(\mathcal{V}^i_\kappa)$ growth with κ indicating the presence of nearby nodes useable as relays for κ-contacts.

For all datasets, observing only contacts shows limited $\mathcal{D}_\kappa$. For instance, the average $\mathcal{D}_1$ for *Infocom05* and *Sigcomm09* is respectively 2.4 and 1.0 neighbors. Observing the κ-vicinity up to a few hops ($\kappa = \{3, 4\}$) increases $\mathcal{D}_\kappa$ by at least a factor of two. *Infocom05* and *Rollernet*'s average $\mathcal{D}_4$ is 10.7 and 4.2. By observing their four-vicinity, we increased their number by a factor of more than four. For $\kappa > 4$, the increase rate is less striking or even null. Nevertheless, longer κ-contacts in terms of path length may not be interesting because of potential path inconsistency due to all relays movements. Monitoring κ-vicinity up to $\kappa = \{3, 4\}$ brings most of the local density a node can use.

5.3.3 Not-in-contact neighbors

An interesting situation occurs when pairs of nodes do not come into contact but belong to each other's κ-vicinity. Usual protocols miss this knowledge by overlooking the potential of nearby nodes. To analyze the impact of such situations, we studied the closest distance between nodes for all pairs of nodes.

For *Unimi*, *Infocom05*, and *Sigcomm09* we find that respectively 92%, 91%, and 80% of the pairs of nodes come to contact at least once. This can be explained by the nature of the datasets where people are coworkers. However, we find that even in these cases, some nodes never meet (respectively for 6%, 7%, and 12% of them). Datasets of different nature, like *Rollernet*, *Community*, and *Shopping* show that contact only represent 31%, 42%, and 61% of the lowest distances. There, respectively more than 51%, 46%, and 35% of the nodes stay, at the closest, between two and four hops. In the *Stanford* dataset, around 70 % of pair of nodes come into contact, but the remaining 30% of the nodes come at most at a two-hop distance (mostly between two and three hops). In *RT*, all pair of nodes come into contact.

By observing the $\{3, 4\}$-vicinity, we manage to monitor additional situations of non-contact between nodes. As a result, a small value of κ is enough to catch most κ-contacts occurring in a node's vicinity.

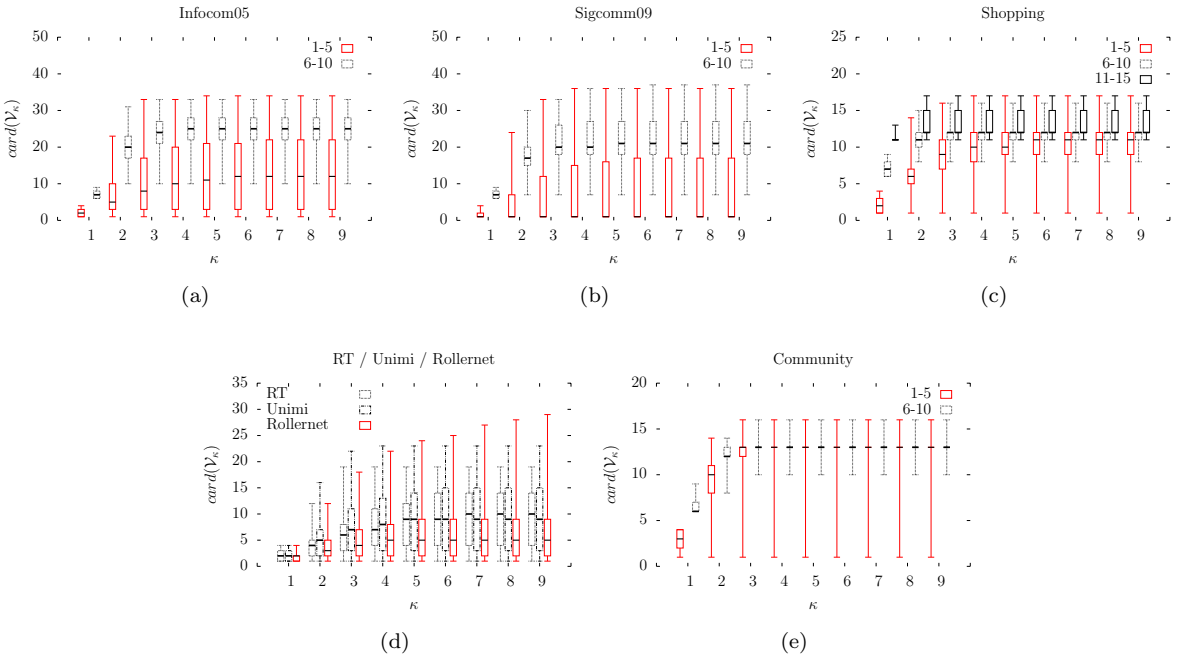


Figure 17: Contact-based $card(\mathcal{V}_\kappa)$ according to κ for each bin [n:m] (bin size = 5) . The x-axis represents the considered κ value for the κ-vicinity. The y-axis indicates the κ-vicinity size. For each bin, the candlebar displays from bottom to top, the lowest vicinity size, the first quartile, the median size, the third quartile, and the highest size value.

5.3.4 A rule of thumb for $card(\mathcal{V}_\kappa^i)$

The strength of κ-vicinity lays in its size and extent. For a given node i, the most straightforward information relies on its current number of contacts ($\mathcal{C}^i$). To facilitate the deployment of our proposal without the costly neighborhood probing, we propose an heuristic based on $\mathcal{C}^i$ to derive node i's current κ-vicinity size – $card(\mathcal{V}_\kappa^i)$. We want to investigate the relationship between the number of nodes in contact and the current $card(\mathcal{V}_\kappa^i)$. To ease data understanding and their representation, we group the number of nodes in contacts by bins of 5 consecutive values. Fig. 17 presents this bin-based vision of κ-vicinities. The x-axis indicates the considered κ. The y-axis represents $card(\mathcal{V}_\kappa^i)$. We observe for each 5-bins the distributions of κ-vicinity sizes. For its corresponding bin, each candlestick displays from bottom to top: the lowest vicinity size, the first quartile, the median size, the third quartile, and the highest value.

In the *Infocom05* figure, we observe that when nodes have between [1:5] nodes in contacts, they have a median three-vicinity size of 8 nodes. The minimum size observed was one and the largest three-vicinity size was 33. The first and third quartiles are respectively 3 and 17. This means that for a node having one to five nodes in contact, 50% of the corresponding three-vicinities have sizes between 3 and 17. Considering the second bin value of [6:10] contacts, the interquartile difference becomes more interesting. If a node has between 6 and 10 nodes in contacts, its median three-vicinity size will be 24, the first and third quartiles are 21 and 27. As a result, when a node has 6 to 10 nodes in contact, it can quite safely bet to have at least a 27 nodes in its three-vicinity and all the more neighbors to use. For κ-vicinities with $\kappa \geq 3$, for the second bin [6:10], the vicinity size the interquartile value is restricted the interval 25±3. The same observation for bin [6:10] holds for *Sigcomm09*.

The other datasets bear a "low" gravity. Their loosely connected vicinities may explain the low values we observed here. *Shopping* is the only dataset to have nodes with more than 10 simultaneous contacts and therefore, three different contact bins. We notice a logarithmic growth of its median $card(\mathcal{V}_\kappa^i)$. As *RT*, *Unimi*, and *Rollernet* had only [1:5] concurrent contacts, we displayed them in a single figure. Their

average size grow logarithmically with κ. Low gravity values indicate that the most secure connection occur in contacts, therefore most of the neighbors are located near the considered node.

The *Community* dataset displays a different behavior. Its synthetic nature provides κ-vicinities for $\kappa > 2$ of constant size whatever the number of current nodes in contact.

As a rule of thumb, for datasets with large gravity values like *Infocom05* and *Sigcomm09*, their $card(\mathcal{V}_\kappa^i)$ for $\kappa \geq 3$ is on average at least the double of their ceiling bin value (especially for bin [6:10]). For low-gravity datasets, most events occur in contacts, so their $card(\mathcal{V}_\kappa^i)$ for $\kappa \geq 3$ is at least $\mathcal{C}^i$ or the ceiling bin value.

5.4 Final remarks

In this section, we raised the weakness of the binary assertion in disruption-tolerant networking. We showed how this paradox misleads a node's vision of its surroundings and quantified how much end-to-end communication opportunities a node can miss. To address this issue, we defined the notion of vicinity in opportunistic network as well as temporal measure to characterize this novel vicinity notion. We made some interesting observations concerning κ-intercontact and κ-contacts. We confirmed previous results from Karagiannis et al. Aggregated κ-intercontact distributions follow power laws up to a knee point and decay exponentially afterwards. This allows current DTN protocols to leverage their κ-vicinity without too much change in their functioning. Concerning κ-contact patterns, they bear two main patterns which are density-related. Dense settings bring logical results of κ-contact duration extension with higher κ while light settings present result in paradoxical patterns. For a node, gathering vicinity knowledge comes with a cost. The resulting give-and-take between gathering costs and additional vicinity information may become an issue when deploying our proposal.

An important takeaway of our work is that monitoring a node's κ-vicinity with small values of κ is enough to benefit from most κ-vicinity advantages. Finally, we focused in the κ-vicinity notion by analyzing their seat in the considered datasets and presenting the gravity measure. To ease the early stage deployment of our proposal, we also presented a rule of thumb to infer the number of neighbors in a node's κ-vicinity based only on its contacts. As future work, we plan to analyze the improvements of vicinity integration in several opportunistic approaches.

6 On the predictability of contact patterns in the vicinity of a node

Recent studies have addressed the problem of contact prediction – predict if two nodes are going to be in direct transmission range – and have revealed that, under the right prediction method and predictive features, contacts between mobile users are to a certain extent predictable [44]. This result is valuable as it allows one to predict the evolution of the network of human interactions that can be used to design more effective DTN communication protocols.

But contacts between nodes are not the only type of relationship between mobile users. Often, users may find themselves not in direct transmission range but still in the nearby vicinity. Thus, to have a broader view on the available communication opportunities, the extended notion of contact, namely κ-contact, has recently been proposed [32]. Previous analyses showed that considering only contacts between nodes ends up in a biased suboptimal network understanding while studying κ-contacts provides a more complete understanding on the available end-to-end communication opportunities.

In this work, we study the predictability of extended contact opportunities in DTNs. Using data from three human-based contact traces, we show that κ-contact opportunities are more predictable than direct contact relationships. To measure the possible impact of this finding in a real-world application, we propose an experimental setting that supports the idea that κ-contact prediction has an interesting potential usage [45].[8]

Our contributions can be summarized as follows:

[8] Reference available in the appendix.

Table 8: Distance average duration (in seconds).

	κ						
Dataset	1	2	3	4	5	6	7
Infocom05	399	296	224	175	131	154	212
Sigcomm09	149	83	41	25	18	13	11
Rollernet	48	65	76	89	105	114	129

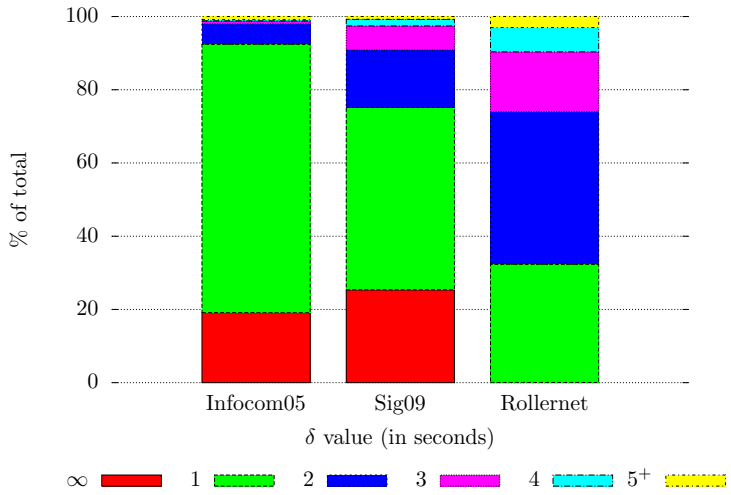


Figure 18: Pairwise minimum distance for *Infocom05*, *Sigcomm09*, and *Rollernet*.

- We provide insights on the κ-contacts relationships between mobile nodes and show that considering only direct contacts covers a limited part of the end-to-end transmission possibilities. We reveal the instability of periods of time nodes stay at the same distance, and that κ-contact intervals display better predictability characteristics: intervals frequency and length.
- Using a supervised prediction framework, we study the predictive nature of κ-contacts and compare it with the traditional case of predicting contacts between nodes. Our results indicate that, in highly dynamic mobile settings (e.g. rollerblading scenario), predicting that nodes will remain at a distance of at most two hops from one another, can attain twice the performance of a direct contact prediction.
- Through simulations, we evaluate the impact of κ-contact prediction in a service that would benefit from predicting contacts between mobile users. The experimental results show that there is a higher potential on relying on κ-contact prediction compared to the traditional contact case.

6.1 Pairwise interactions under the κ-contact case

Given the new definitions of contact and intercontact we analyze different characteristics of the pairwise interactions. For more detailed information concerning κ-contact and κ-intercontact, please refer to [31].

6.1.1 Pairwise minimum distance

We begin by studying the pairwise minimum distance, i.e., how close nodes come to each other during the duration of a trace. For instance, if two nodes meet at least once, we mark this distance as 1. If they come as close as 3 hops, we consider the minimum distance to be 3. For nodes that never come in κ-contact, we consider this distance as ∞.

Table 9: κ-contact average duration (in seconds).

	κ						
Dataset	1	2	3	4	5	6	7
Infocom05	399	322	274	247	230	224	224
Sigcomm09	149	101	72	60	54	51	50
Rollernet	48	61	68	75	81	86	90

We represent the results in Fig. 18. In terms of pairs of nodes that come in direct contact, we observe that in conference settings, characterized by a high number of nodes in restricted physical spaces, the number of connected pairs is reasonably high: 49% for *Sigcomm09* and 73% for *Infocom05*. *Rollernet* on the other hand shows a lower network connectivity, with only 33% of nodes coming in a direct contact. But the analysis of contact alone yields an incomplete picture as there is a considerable amount of nodes who come close to each other but never in direct contact. For example, the percentage of pairs that come at a distance of 2 is 5% for *Infocom05*, 16% for *Sigcomm09*, and 41% for *Rollernet*. For *Rollernet* the percent of nodes that come at a 2-hops distance is even higher than the nodes that come in direct contact and one can observe that a non negligible amount of nodes advance up to a distance 3 (6%) and 4 (16%).

6.1.2 Average distance duration

In Table 12, we present the average duration of an interval during which nodes remain at a distance of κ-hops from one another. For *Infocom05* and *Sigcomm09*, we observe that close connections are more stable, with smaller average durations as the distance between nodes increases. This shows how for conference settings, network stability comes from the core of the κ-vicinity. However, we observe the opposite phenomenon for *Rollernet* dataset. With larger κ we have an increase of the average duration that nodes spend at a certain distance from one another. Thus, due to nodes' movement in a highly dynamic scenario, meeting between users lasts for very short periods of time but nodes spend a significant amount of time in the nearby vicinity.

6.1.3 Average κ-contact duration

We also study the average κ-contact durations (see Table 9), i.e., we observe the average duration of any κ-contact interval. Following our logic, since we cover a wider spatial range with our κ-vicinity, nodes coming closer are likely to be in κ-contact earlier and leave the κ-contact later, therefore we should obtain longer κ-contact intervals. With *Rollernet*, we observe that the greater the value for κ, the longer the durations. Surprisingly for *Infocom05* and *Sigcomm09*, this is not the case, we actually notice the opposite phenomenon. With larger κ, we seem to have smaller κ-contact intervals. So does that mean that increasing our network vision with the κ-vicinity reduces the duration of end-to-end transmission possibilities?

Table 10 shows how wrong this conclusion may be. In this table, we show the actual number of κ-contact intervals for each κ and each dataset. For all of them, the greater the value of κ, the greater the number of κ-contact intervals. So, with higher κ values, we multiply the possibility of observing a κ-contact interval. They may be on average of shorter length (for *Infocom05* and *Sigcomm09*) yet we multiply the possibility of having pairwise end-to-end paths. In addition, the cumulated κ-contact duration grows with larger κ.

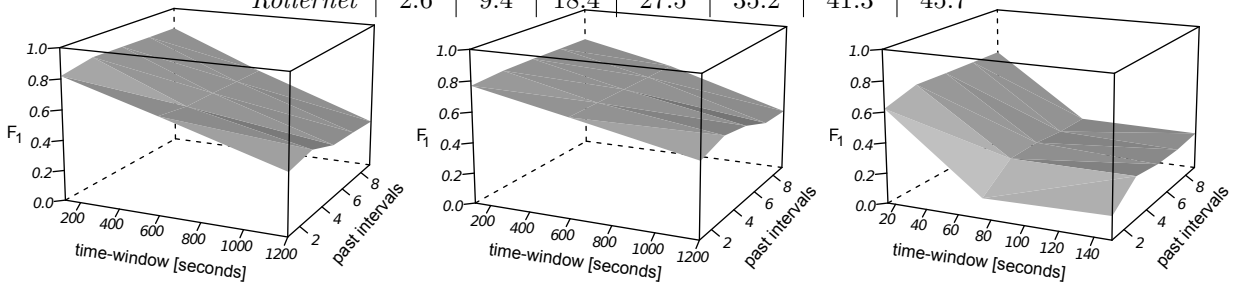
6.2 Predicting κ-contact encounters

6.2.1 Dynamic graph representation

The mobile traces analyzed in this work represent dynamic networks composed of a set of mobile users that sporadically come in contact. We represent this network using a dynamic graph structure, $G_{0,T} =$

Table 10: κ-contact number of intervals ($\times 1,000$).

	κ						
Datasets	1	2	3	4	5	6	7
Infocom05	3.7	14.7	28.9	40.0	46.7	50.3	51.9
Sigcomm09	13.3	49.7	96.9	131.6	152.2	163.4	168.8
Rollernet	2.6	9.4	18.4	27.5	35.2	41.3	45.7



(a) *Infocom05* (b) *Sigcomm09* (c) *Rollernet*

Figure 19: Prediction performance for different time-window durations and by varying the number of training time-windows (past intervals).

$(V, E_{0,T})$, with V the set of mobile users observed during a finite period of time [0, T) and $E_{0,T}$ the set of temporal edges between them. We consider an edge $e_{uv} \in E_{0,T}$ if any two users $u, v \in V$ have been at least once into contact during the period [0, T). To analyze the evolution of this network over time, we split time into fixed time-windows of duration w and represent the dynamic network as a time series of network snapshots G_{t_1}, G_{t_2}, ..., G_{t_n}, with $n = \lceil \frac{T}{w} \rceil$. G_{t_i} represents the aggregate graph G_{t_{i-1},t_i} that records the contacts between mobile users during the period $[t_{i-1}, t_i)$. In a dynamic network, the future changes of the network may depend not only on the most recent state of the network but also on older ones. To model the dynamic evolution and catch possible periodicities in human encounters, the data used as input in the prediction process is represented as a successive series of static snapshots $G_{t_{i-m}}, ..., G_{t_{i-2}}, G_{t_{i-1}}$. Thus, given data from the previous m time-windows our objective is to predict the κ-contacts during the next target period G_{t_i}. We will later discuss how the choice of w and m affect the prediction performance.

6.2.2 κ-contact prediction problem

We formulate the prediction task as a binary classification problem where, given past data recorded until a moment in time t_{i-1}, the goal is to predict if any two mobile nodes will be in κ-contact during the subsequent period $[t_{i-1}, t_i)$.

We rely on two types of information in the prediction model: the frequency of κ-contact occurrences and the structural properties of the connectivity network. The first type of information measures the strength of κ-contact relationships, quantified by the *duration* and the *number of times* any pair of nodes has been in κ-contact in the past. A longer duration and a greater number of κ-contacts can provide stronger evidence that two nodes will be in κ-contact in the future. For the second type of information, to quantify the structural properties of the network, we extract various features that capture the proximity between nodes in the graph of past interactions. These features showed predictive power in various applications such as collaborative filtering and link prediction problems [46, 47, 48]. In this work we use four common proximity measures:

- *Common neighbors (CN).* For each pair of nodes $u, v \in V$, CN represents the number of common neighbors:

Table 11: Notation for the binary classification confusion matrix

		Predicted value	
		predicted = 1	predicted = 0
Actual value	actual = 1	TP	FN
	actual = 0	FP	TN

$$CN_{(u,v)} = \mid \mathcal{V}_1^u \cap \mathcal{V}_1^v \mid. \tag{12}$$

- *Adamic Adar* [49]. This measure extends the notion of common neighbors by weighting each neighbor by the inverse logarithm of its degree centrality:

$$AdamicAdar_{(u,v)} = \sum_{x \in \{\mathcal{V}_1^u \cap \mathcal{V}_1^v\}} \frac{1}{\mid \mathcal{V}_1^x \mid}. \tag{13}$$

- *Katz* [50]. This feature counts all the paths between any pair of nodes, giving a higher weight to shorter paths. If $path_{u,v}^l$ represents the set of paths of length l between two nodes u and v, and β is a damping factor (set to 0.05 in our evaluation), the Katz score is calculated using the following formula:

$$Katz_{(u,v)} = \sum_{l=1}^{\infty} \beta^l \times \mid path_{u,v}^l \mid. \tag{14}$$

- *Preferential attachment* [51]. This feature is built on the premise that the probability of a new contact is correlated with the product of nodes' degree.

$$PA_{(u,v)} = \mid \mathcal{V}_1^u \mid \times \mid \mathcal{V}_1^v \mid \tag{15}$$

The two types of features provide complementary information about nodes' contact patterns. The frequency of interactions catches the persistence of κ-contact relationships but its predictive power is conditioned by the past contact occurrences (using these features one can only predict the reoccurrence of a κ-contact). Topological features, on the other hand, allow us to capture complex data patterns about the structure of the network of interactions. We build the prediction model and report the results using the entire set of features as we observed that taking these features together achieves a higher performance than using them separately.

We adhere to a supervised learning procedure in our evaluation. Each mobile trace is split in two equal-sized temporal parts: the first period is used as the training set and the remaining part serves to report the prediction performance. We examined two classification algorithms: SVM (using LIBSVM library [52]) and logistic regression, under different parameter settings and used a validation set to avoid overfitting. We report the quality of the prediction using the F_1 score (also called F-measure), expressed as the harmonic mean between precision ($\frac{TP}{TP+FP}$) and recall ($\frac{TP}{TP+FN}$) as defined by the confusion matrix (Table 11).

6.2.3 The effect of time-window duration and past data

The prediction performance is influenced by the duration of the time-window. Aggregating data over longer durations may lose useful temporal information about the structure of the dynamic network. Another important aspect is given by the number of training time-windows. Including more information from the past may capture important temporal patterns but also increase the computational cost.

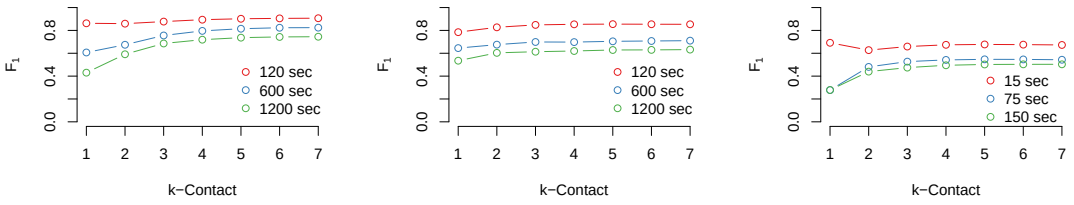
To evaluate the impact of past information in the prediction performance we vary the amount of data used in the prediction model and include information from the previous $\{1, 3, 5, 7, 9\}$ time-windows. We illustrate the results for the 1-contact case as we observed that the remarks made on this value are

consistent with other κ values as well. For the size of the time-window we select the most granular duration (the scanning rate used in the mobile trace) and two other values that represent $5\times$ and $10\times$ this duration. Thus, we consider time-windows of duration {120, 600, 1200} seconds for *Sigcomm09* and *Infocom05* and use {15, 75, 150} seconds for *Rollernet* (which has a more granular frequency).

The results are presented in Fig. 19 by means of 3D plots that represent the F_1 score as a function of the time-window duration and the number past time-windows used in the prediction model. On the x-axis we examine different time-window durations and the y-axis (labeled *past intervals* in Fig. 19) denotes the number of time-windows used in the prediction model. For example, a past interval of length 9 for a time-window of 1200 seconds means that, based on the contacts recorded during the previous 9 intervals of 1200 seconds, we predict contacts during the next time-window.

The figure illustrates that the most recent information plays the most important role in the prediction performance. For all three datasets, using data from the latest three time-periods achieves the highest performance and older information has little predictive power. This indicates that the most recent interactions are the most important in predicting the immediate future. We can also observe that the longer the duration of the time-window, the less accurate the prediction performance. This suggests that aggregating data over longer durations is prone to larger errors. Taking the example of *Infocom05* (Fig. 19(a)), the results show that predicting the contact opportunities during the next 2 minutes shows an F_1 score of 0.8 and the performance drops with 50% when trying to predict what will happen during the next 20 minutes. For *Rollernet*, which represents a more dynamic scenario, the drop of performance is even higher with a 70% decrease when trying to predict the contacts during the next 150 seconds compared to a 15-seconds time-window.

6.2.4 κ-contact prediction results

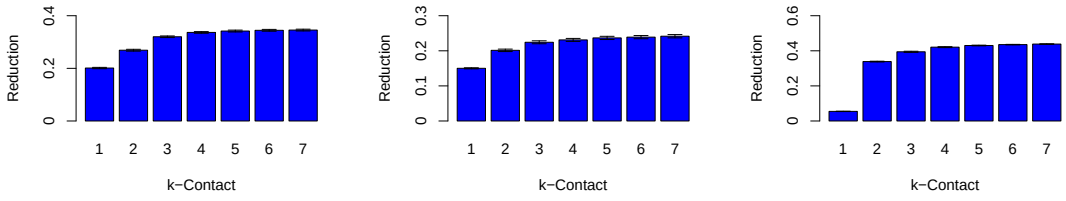


(a) *Infocom05* (b) *Sigcomm09* (c) *Rollernet*

Figure 20: The efficiency of predicting κ-contact relationships for different durations of the time-window. On the y-axis we represent the prediction performance and on the x-axis we vary the value of κ-contact from 1 to 7.

Based on the previous observations of the optimal number of past intervals we assess the performance of predicting κ-contact relationships. We vary the value of κ from 1 to 7 and consider three durations for the time-window: {120, 600, 1200} seconds for *Infocom05* and *Sigcomm09* and {15, 75, 150} seconds for *Rollernet*. The results are illustrated in Fig. 20. First, we observe that predicting that two nodes will be in direct communication range shows particularly poor results in very dynamic mobile settings (e.g. *Rollernet* that describes a rollerblading activity) and for longer durations of the time-window. Thus, in situations that involve important changes in the network topology, predicting that nodes will be in direct contact is prone to large errors.

Relaxing the prediction objective beyond direct contact relationships reveals more accurate predictive power. Overall, the greater the value for κ the more effective the prediction performance. On average (for all mobility traces and different time-window durations) predicting that nodes will be at most at a distance 2, 3, and 4 shows an improvement of 7%, 10%, and 11% compared to the case where we want



(a) *Infocom05* (b) *Sigcomm09* (c) *Rollernet*

Figure 21: The percent of traffic with the infrastructure that can be reduced through κ-contact prediction and mobile opportunistic communications. On the y-axis we represent the traffic reduction compared to the case where content is sent to mobile users using only the infrastructure. On the x-axis we present different values for κ-contact.

to predict direct meetings between mobile users. While the improvement is important for small values of κ we notice that there is little benefit in extending the prediction for a κ greater than 3. The most significant increase, compared to the direct contact case, can be observed for $\kappa = 2$ with an average increase of 10% for *Rollernet*, 7% for *Infocom05*, and 6% for *Sigcomm09*. The benefit is negligible when trying to predict the network change in the immediate horizon but it becomes significant when trying to make predictions over longer periods of time. Taking the case of *Infocom05* for a time-window of 1200 seconds and *Rollernet* for 150 seconds, predicting that nodes will be separated by at most two nodes (κ-contact = 3) reveals an improvement of 60% for *Infocom05* and 74% for *Rollernet* compared to the direct contact prediction case.

While we leave a more detailed analysis of these findings for future work, we give two plausible explanations for these results. First, as we showed in Fig. 18, a non-negligible number of nodes, although never in direct contact, they come at a 2-hop distance. By extending the prediction objective to 2-hop contacts, we include these potential events into consideration, which appear to have a more predictable nature. Then, as showed in Section 6.1 direct contacts between mobile users are scarce and short-lived, which makes them more difficult to predict in very dynamic scenarios and for longer time horizons. This explains the low prediction effectiveness observed with *Rollernet* and for longer time-windows for *Sigcomm09* and *Infocom05*. Thus, extending the notion of contact to κ-contact gives us access to more stable connections (nodes leave direct connectivity but remains in κ-contact for longer durations) that reveal a more predictable nature.

6.3 Practical implications

To capture the possible benefit that κ-contact prediction would bring in practical scenario we propose and evaluate the following use-case example.

We consider a content producer, located on the Internet, that regularly publishes content for a known group of collocated mobile users that communicate with the server using the cellular infrastructure. Content is categorized in topics. Users subscribe to these topics and content is pushed to users upon creation. We also consider that, in order to reduce the amount of cellular traffic caused by content delivery, the content producer collects data about the mobile contact traces and relies on a κ-contact prediction functionality when transmitting information to users. More specifically, at the publication of a content, instead of individually transmitting the content to each subscriber, the content producer optimizes the delivery process based on the predicted κ-contact opportunities. For example, if the server predicts that two users, interested in the same content, will be in κ-contact, a message is sent to only one of these nodes which will opportunistically forward the message to the other node when they will be κ-contact. We also assume that nodes are capable of sensing their κ-vicinity and can detect when a targeted user is in κ-contact. To collect nearby topological knowledge, we assume the existence of a link-state protocol

gathering nearby knowledge under the form of a connectivity graph. The implementation itself is beyond the scope of our study, yet a previous analysis studied the impact of monitoring overhead [32].

We design the experimental setting using ONE simulation environment [53]. In our experiments we set the number of topics to 100. Each mobile node randomly subscribes to 20 up to 100 topics. For the prediction module, we use a time-window of 75 seconds for *Rollernet*, and 600 seconds for *Infocom05* and *Sigcomm09*. Content is uniformly created throughout the duration of the experiments (that covers the duration of a mobility trace) and the results are averaged over 10 simulation runs. We also consider an infinite cache size at the user side and assume that the content is small enough to fit into one message in the communication between content producer and the users and between the mobile users. To measure the impact of κ-contact prediction we report the reduction in the number of messages in the communication between the content producer and the mobile users when using κ-contact prediction module compared to a case where the content is individually sent to each user using the cellular infrastructure.

The results are presented in Fig. 21. First, we observe that the greater the value of κ-contact, the greater the potential of traffic reduction. The biggest improvement of predicting beyond direct neighbors is noticed for $\kappa = 2$, that shows an improvement of 6% for *Sigcomm09*, 7% in *Infocom05*, and 30% for *Rollernet*. The potential traffic reduction is directly affected by the characteristics of the traces: κ-vicinity properties (presented in Fig. 18) and prediction performance (presented in Fig. 20). Taking the example of *Sigcomm09*, even if the effectiveness of the prediction showed little improvement for $\kappa = 2$ compared to $\kappa = 1$ the potential reduction is nevertheless important (6%). This is explained by the significant number of nodes located at a 2-hop distance detected with the κ-contact prediction. The benefit is even more substantial in the case of *Rollernet*. By counting on the pairs of nodes connected at a 2-hop distance (that exceed the number of direct contact opportunities), the traffic reduction attains a performance of 33% compared to 5% when using only direct contact prediction.

6.4 Final remarks

In this section, we addressed the problem of predicting κ-contact opportunities between mobile users – predict if users will find themselves at a distance of at most κ-hops from one another. By analyzing three real-world contact traces, we observed that one can obtain better performances when predicting 2^+-contacts compared to the direct contact case. To assess the impact of these findings in a real-world application, we proposed a simulation experiment in which, by combining mobile opportunistic communications with κ-contact prediction one can reduce the amount of traffic used in the communication of mobile nodes with the infrastructure. Our results suggest that services benefiting from contact predictions can efficiently exploit the predictable nature of κ-contacts.

Research in the area of κ-contact prediction and its applications is an open subject with many possible extensions. First, as the observations are based on specific mobility settings (conferences and rollerblading) more work is needed in order to understand to what extend these observations can be generalized to other mobility scenarios. The quality of the prediction shows promising performances, yet not optimal, and suggests there is still room for improvement. One way to increase the prediction performance is to consider additional features in the prediction model. This includes information about the geographical co-location patterns of nodes (not available in the traces used in this work) that showed strong predictive power in the link prediction problem [44, 48]. Then, in this study, we only consider a one-step ahead prediction problem (we use information received in the previous m time periods to predict what will happen during $m + 1$ period). To cover a larger range of situations the prediction objective can be extended to further time periods, i.e. predict contacts during subsequent time periods $m+1$, $m+2$, ..., $m+n$. This can be particularly favorable in situations where collecting and processing data adds a significant delay, for which immediate prediction can be considered outdated.

Finally, the current evaluation of the applicability of κ-contact prediction in real-world scenarios considers some simplifying assumptions and more work would be needed to assess the benefit of κ-contact prediction in practice. This includes an evaluation of the additional cost of collecting data about nodes' mobility and the feasibility of implementing a mechanism capable of detecting the κ-contact communication opportunities.

7 Contact patterns and security concerns

7.1 The security challenges of a fickle environment

In the communication scheme proposed in MOTO, the network is intrinsically dynamic, that is, users and their roles, the topology, the offloading technology used, etc. are susceptible to change at any moment, while the transport medium is considered open (wireless) and highly fickle (opportunistic).

The contact pattern of this type of communication is mainly driven by the mobility of the MOTO users. Contact patterns in such a scheme, suppose relatively short interaction times between different nodes and no reliability in user's continuity in the network. Users can drop off the network at any moment, both intentionally (e.g. Turning off their UE wireless capabilities) or unintentionally (leaving the range of coverage from the device they are communicating with at a given moment). In such an environment where nodes enter and leave the network at any time, it is likely that they will not be able to provide strong evidence of their identity and intentions.

This, despite being necessary to achieve the expected QoE for the user, the increase of efficiency and the reduction of costs and bandwidth occupation for the operators, sets an enormous challenge for the security. The dynamic yet vulnerable nature of an ad hoc network presents many new security and privacy challenges [54].

Furthermore, lifetime of links and routes is not only determined by user mobility but also by different sources of failures like packet collisions, or interference from any sender emitting at the same frequency band [55].

On the other hand, the specific communication scheme that has been set in MOTO, which can be considered a special type of Mobile Ad Hoc NETworks (MANETS) as there is a centralized management node available (MOTO platform), which is responsible for the definition of a dissemination strategy, brings the necessity for the UE to share its location, in order to enable the management of the content distribution. This sets on its own another challenge for security, as location privacy has become a greater concern than in traditional networks and it is particularly difficult to achieve a satisfactory level of location privacy in situations where nodes rely on location-based services (LBS) [56].

In this sense, the MOTO proposed network can be considered as a hybrid MANET network, where part of the communications are performed through the traditional channel (LTE and Wi-Fi operator's infrastructure), but the bulk of communications rely in the wireless capabilities of the UEs (Ad Hoc). MOTO proposes the migration from an operator infrastructure centred communication, to a hybrid communication, which will rely mostly in user equipment in order to decongest operator's bandwidth.

Therefore, the main aim of the proposed MOTO security approach will be focused on accomplishing users' concerns about security as well as providing the expected security demanded by the rest of the communications actors: legislators, operators and Internet-based service providers.

The main concerns in the adoption of the MOTO proposition from the user point of view are: (i) *confidentiality* regarding both, sensible information stored in their mobile device and information they exchange with other users, (ii) *integrity* perceived as the detection of the manipulation of UE stored or transmitted information, and (iii) *availability*, understood as users being able to perform communication under all circumstances.

This affirmation is reinforced by recognized security institutions such as the NIST (National Institute of Standard and Technology) of the U.S. Department of commerce in their Guidelines for managing and securing mobile devices in enterprises, where the NIST states: Mobile devices, such as smartphones and tablets, typically need to support multiple security objectives:

- Confidentiality: ensure that transmitted and stored data cannot be read by unauthorized parties.
- Integrity: detect any intentional or unintentional changes to transmitted and stored data.
- Availability: ensure that users can access resources using mobile devices whenever needed.

In this same line, ENISA (European Network and Information Security Agency) in its publication "Smartphones secure development guidelines for app developers" recommends that "Adequate protection

should be built in order to minimize the loss of sensitive data on the device". ENISA also advices about the necessity of ensuring that sensitive data transmitted is protected, and warns against the interception of communications through insecure channels.

In the scope of the security approach in MOTO, and attending to user's main concerns, users will have to be guaranteed that all their personal information, the one stored in the UE and the one transmitted through the proposed communication scheme remain secured and that the service is accessible and usable under demand. All these conditions will shape the specific security objectives for the MOTO security approach:

- User's trust related to privacy and integrity of personal data:
  - Content protection end-to-end.
  - Strong Access Control Mechanisms.
- User's trust in the service:
  - Availability in all cases.
  - Accomplishment of Users QoE (maximum delay).

In order to analyse the security in such fickle environment, and identify the main threats that may threaten the achievement of the proposed objectives, it is necessary to first understand the proposed communication strategy proposed.

7.1.1 MOTO content dissemination strategy

In MOTO, content is disseminated in two ways, through the primary channel (LTE or Wi-Fi) or through the alternative channel (Wi-Fi Ad Hoc). MOTO dissemination strategy is based in the delivery of the content to a reduced number of MOTO users, that will afterwards disseminate it within other MOTO users using UE own Wi-Fi capabilities (Ad Hoc) with the objective of leveraging the operator's network bandwidth by reducing congestion. In order to assure the reception of the content by all the intended users, a panic zone (maximum delay time) has been defined, that when reached will trigger the direct delivery of the content through the primary channel to those users who have not sent a positive content tracking feedback.

Regarding to the contact patterns of MOTO, user's mobility will affect mainly the following communications:

1. Authentication of UE to core MOTO services: This communication encompasses the interaction of the UE with the core MOTO service so that the UE can start using MOTO services. Information exchanged in this communication may include credential expedition.
2. Offloading instructions delivery: this communication encompasses the delivery of instructions of the dissemination to be accomplished by MOTO users. The information exchanged can include location and identification of users.
3. Content download: this communication deals with the delivery of content from the Internet or core MOTO services CDM to the UE.
4. Authentication between UEs: this communication defines the exchange of credentials between MOTO users for identification and trust purposes before the exchange of content or other data. Information exchanged in this communication includes credentials.
5. Content offloading: this communication is defined as the transmission of content from one UE to another through the alternative (Ad Hoc/Wi-Fi) channel.

6. Content tracking feedback: this communication is the acknowledgement of a UE for the content received to be sent to the core MOTO services.

7. Topological information: this communication although it is not defined in the use case sequences, is defined in the description of the terminal architecture. It is defined as the delivery of topological information from the topological information building block in the UE to the LOC module of the MOTO architecture in the operator core network.

For the scope of this analysis, it is going to be supposed that communications ongoing through the primary channel are secured, understanding that:

1. In LTE: communications are encrypted with the Authentication and Key Agreement (AKA) protocol, and that integrity and confidentiality are granted through the implementation of the UIA2 and UEA2 algorithms and the underlying primitive SNOW 3G stream cipher.

2. In Wi-Fi: communications are secured using the WPA2 (Wi-Fi Protected Access2) protocol, employing the Counter Cipher mode with Message Authentication Code Protocol (CCMP), and an encryption scheme that uses AES in CCM offering both, message privacy and message authentication.

Under these assumptions, the challenges to achieve the defined security objectives rely on the securization of the alternative channel.

7.1.2 Confidentiality of the alternative channel communications

Confidentiality regarding the transmitted data is a major security concern in MOTO. Taking into account the typology of the communication scheme proposed in MOTO, where ideally the communication bulk relies within UEs, with or without central entity intervention (MOTO platform), it is necessary to propose a security scheme that can assure the confidentiality of the personal information that can be transmitted through any possible communication situation.

It is convenient to define a relation between the communications that occur through the alternative channel and their main threats:

- Authentication between UEs. Authentication between users may pose a threat to confidentiality depending on the authentication mechanism that is used. In the proposed security approach for MOTO, exposed in section 7.2, X509 v3 certificates will be used for authentication between MOTO users, and revocation of these certificates are granted through lightweight OSCP protocols. In such an authentication scheme, as users are likely to get in contact with any other MOTO users, and as certificates are considered publicly available between them, no personal unauthorized information will be disclosed. It is important to take into account that in order to avoid personal information disclosure, MOTO certificates should not contain Personally Identifiable Information (PII) about the MOTO user.[9]

- Content offloading. Content to be transmitted through this channel can encompass Personally Identifiable Information (e.g., emails). Therefore, it is imperative that content only is accessible to the actual intended receiver and that disclosure to unauthorized users is avoided. The main threat to this communication will be that a malicious node can also receipt the message, which is very likely given the openness of the channel and extract information from it. This might be done by listening to all communications in the channel (eavesdropping), impersonating the actual intended receiver or the MOTO platform (e.g. by a man-in-the-middle attack or by manipulating the users applications configuration). In order to avoid this threat, communications through the alternative channel SHOULD be encrypted and users UE should confirm other MOTO user's identities when requesting MOTO certificates revocation to the MOTO platform through the primary channel.

[9] As used in US privacy law and information security, is information that can be used on its own or with other information to identify, contact, or locate a single person, or to identify an individual in context.

7.1.3 Integrity of the alternative channel communications

At the time of this analysis, the communication terminal to terminal protocols and the control and coordination protocols are under development. Therefore, the integrity of the communications between MOTO UEs will be based in the following assumptions:

- MOTO applications in the UE authenticate successfully using the proposed authentication scheme (x509 v3 certificates exchange and lightweight OSCP revocation confirmation) before starting.
- Identity of users is confirmed through the revocation certificate request to the MOTO platform.
- MOTO applications in the UE agree in a session key to be used using asymmetric cryptography.

Under these assumptions, it can be supposed that the integrity of data transmitted through the alternative channel is granted. As the proposed security of this channel encompasses WPA2 connections and the exchange of certificates and asymmetric cryptography, data tampering is unlikely. Data tampering is always detectable in this type of protocols, as Message Authentication Codes are used.

Therefore, for communications through the alternative channel, a major threat for integrity will be the nodes themselves. A misconfiguration or an erroneous implementation in the MOTO application of the security of one of the nodes could habilitate threats for security concerning the communication by reducing the strength of the encryption. Another threat for integrity could be that a malicious node performs a man-in-the-middle attack and impersonates one of the legitimate nodes.

The impact of the corruption of the integrity in this communication could suppose the impossibility of the user to access the real content transmitted, the disclosure of personal information (when accepting the offloading of malware from an mistakenly considered trust user or when sending this information through the communication, or even the disruption of the MOTO services for the user.

7.1.4 Availability of the alternative channel communications

Availability stands for the normal service provision in face of all kinds of attacks. The availability of services through the alternative channel relies on the availability in turn, of several services and assets. In particular, service availability of the alternative channel, relies in UEs access to MOTO's core platform services (offloading instructions, OSCP, etc.), wireless capabilities of MOTO UE and UEs access to other legitimate MOTO UEs.

The main threat for the services availability is related to the openness of the transmission channel, interferences or massive data transmission can collapse the transmission channel, thus disrupting MOTO services. To attempt an attack in this sense, it is necessary to either overload the network with data packet transmission or use a device that interferes, introducing signals in the same frequencies of transmission in order to decrease signal to noise ratio (jamming) [57]. On the other hand, as each of the elements of the UE architecture has their own functionality in the communication scheme, the effect of the disruption of any of them will have an effect to a greater or lesser extent.

In deliverable D2.2.1 (*General Architecture of the Mobile Offloading System Release A*), a detailed explanation about the different building blocks of the architecture is done, which will be used as a base for this analysis. In Table 12, a description of the impact of disruption of each of these building blocks is given.

The main threat that could enable the corruption of the elements listed before is that an attacker introduces some type of malware in the Moto UE that allow him to access the application.

7.2 Proposed MOTO security

Regarding the authentication between users, MOTO considers the secure exchange of credentials and the trust acknowledgement from previous connections.

The MOTO project is investigating the performance of different types of protocols and certificates under different mobility patterns. The initial assumption is that these credentials will be based on the

Table 12: Disruption of building blocks at a MOTO terminal.

Building block	**Description of Impact if disruption**	**Impact range**
Flow identification	If this element of the user's terminal is corrupted, the MOTO application will not be able to distinguish between packages intended for control issues and data packages corresponding to the content. Depending on how the configuration is settled in this element, this could affect severely the service availability.	MOTO application services availability
Topological information	If this element of the user's terminal is corrupted, the terminal might have problems to be located by the core MOTO services. Under these circumstances, other MOTO nodes will not receive offloading instructions to send them the requested content. Therefore, the user will not be able to participate through the alternative channel and will have to receive the content through the primary channel.	Terminal participation in the alternative channel (ad hoc network).
Content tracking feedback	If this element of the user's terminal is corrupted, the MOTO platform will not receive appropriate feedback from the communications of the UE. This may cause that content already received through the alternative channel is sent again to the user through the primary channel, or it can cause the user to send acknowledgement of content when it has not really received the content. This presumably will affect content reception, so it will have an effect on service availability.	Terminal participation in the alternative channel (Ad Hoc network) Content reception duplicity or no reception at all.
Diffusion Instructions	If this element of the user's terminal is corrupted and the information it stores (instructions) is corrupted or deleted, the MOTO application will forward the content erroneously or even do not forward it. It will affect the dissemination strategy stated by the MOTO platform. Presumably, although this will mainly affect content delivery, it will not have an effect on service availability, unless the corruption affects a large number of UE, thus causing a meltdown of the ad hoc network.	Terminal participation in the alternative channel (ad hoc network) Content reception duplicity or no reception at all.
Cache	If this element of the user's terminal is corrupted and the information it stores (content) is corrupted or deleted, the MOTO application will forward an erroneous content or even do not forward it. Presumably, although this will affect content delivery, it will not have effect on service availability for the user.	No effect on Users participation through the alternative channel.
Routing	If this element of the MOTO platform is corrupted, the MOTO application will not perform correctly the dissemination strategy instructions sent by the MOTO platform, which could affect the content forwarding. Presumably, although this will mainly affect content delivery, it will not have effect on service availability for the user.	No effect on user participation through the alternative channel.
Auth	If these element of the UE terminal is corrupted, the MOTO application could be unable to authenticate to another MOTO application, or even to the MOTO platform, or could accept connection from an untrusted user (potentially malicious). All of which could affect the access to the service (availability).	Terminal participation in the alternative channel (Ad Hoc network). Unable to access the MOTO service.

third generation of X.509 certificates, as they provide much more flexibility than the traditional certificates by allowing the introduction of new fields, which would be really useful in MOTO, for instance for the incorporation of users' trust level. This trust level is intended to indicate the updated user's reputation at a certain moment, as it varies dynamically depending on the trust feedbacks received after every communication they are involved in with other users.

Furthermore, MOTO will study the use of certificate's revocation request protocols, like the on demand OCSP protocol (Online Certificate Status Protocol), which determines the status of the digital certificates in a more agile and efficient way than the traditional CRLs (Certificate Revocation Lists), thus leading to a minimization of the communication bandwidth and client-side processing consumption and enabling the identity confirmation of the users.

As it is true that security and trust will have an impact in the overall architecture in terms of opportunistic communication efficiency, the right balance in terms of credential management complexity and communication efficiency is being investigated. For this purpose, the use of flexible security technologies and lightweight protocols is a key target for the definition of the security approach in the MOTO environment.

8 Conclusion

We have presented in this deliverable several contributions that help us better understand the phenomena and possibilities behind contact and intercontact patterns in opportunistic networks. This step is fundamental to evaluate the potential gains that we may obtain by offloading cellular traffic onto device-to-device communications.

Our work put light on a number of elements that seem to strongly influence the way protocols and algorithms could better benefit from contact opportunities. In particular, the compound of individual and collective behaviors seem to be an important condition to this end. Our contributions in the context of the MOTO project makes a step forward to allow opportunistic offloading techniques generate de necessary savings to push network operators to adopt this technique.

References

[1] C. Boldrini and A. Passarella, "HCMM: Modelling spatial and temporal properties of human mobility driven by users' social relationshipss," *Computer Communications*, vol. 33, no. 9, pp. 1056–1074, 2010.

[2] V. Borrel, F. Legendre, M. De Amorim, and S. Fdida, "Simps: Using sociology for personal mobility," *IEEE/ACM Transactions on Networking (TON)*, vol. 17, no. 3, pp. 831–842, 2009.

[3] D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella, "Spot: Representing the social, spatial, and temporal dimensions of human mobility with a unifying framework," *Pervasive and Mobile Computing*, vol. 11, pp. 19–40, Apr. 2014.

[4] M. Newman, "The structure and function of complex networks," *SIAM review*, pp. 167–256, 2003.

[5] J. Silvis, D. Niemeier, and R. D'Souza, "Social networks and travel behavior: Report from an integrated travel diary," in *11th International Conference on Travel Behaviour Reserach, Kyoto*, 2006.

[6] E. Cho, S. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *SIGKDD'11*. ACM, 2011, pp. 1082–1090.

[7] M. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, no. 2, p. 026113, 2004.

[8] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, 2005.

[9] C. Bron and J. Kerbosch, "Algorithm 457: finding all cliques of an undirected graph," *Communications of the ACM*, vol. 16, no. 9, pp. 575–577, 1973.

[10] T. M. Fruchterman and E. M. Reingold, "Graph drawing by force-directed placement," *Software: Practice and experience*, vol. 21, no. 11, pp. 1129–1164, 1991.

[11] M. Jacomy, S. Heymann, T. Venturini, and M. Bastian, "Forceatlas2, a graph layout algorithm for handy network visualization," Gephi Consortium, Tech. Rep., 2011.

[12] T. Hossmann, T. Spyropoulos, and F. Legendre, "Putting contacts into context: Mobility modeling beyond inter-contact times," in *Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 11)*. Paris, France: ACM, May 2011.

[13] T. Hossmann, G. Nomikos, T. Spyropoulos, and F. Legendre, "Collection and analysis of multi-dimensional network data for opportunistic networking research," *Computer Communications*, 2012.

[14] S. Catanese, P. De Meo, E. Ferrara, and G. Fiumara, "Analyzing the facebook friendship graph," 2010.

[15] "Gowalla." [Online]. Available: http://blog.gowalla.com/

[16] "Foursquare." [Online]. Available: https://foursquare.com/about

[17] "Altergeo." [Online]. Available: http://www.crunchbase.com/company/altergeo

[18] E. Lehmann and J. Romano, *Testing statistical hypotheses*. Springer, 2005.

[19] W.-J. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling spatial and temporal dependencies of user mobility in wireless mobile networks," *Networking, IEEE/ACM Transactions on*, vol. 17, no. 5, pp. 1564–1577, 2009.

[20] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.

[21] E. Biondi, C. Boldrini, A. Passarella, and M. Conti, "Duty cycling in opportunistic networks: inter-contact times and energy-delay tradeoff," IIT-CNR, Tech. Rep. 2013-TR-022, 2013.

[22] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*. ACM, 2005, pp. 244–251.

[23] S. Gaito, E. Pagani, and G. P. Rossi, "Strangers help friends to communicate in opportunistic networks," *Computer Networks*, vol. 55, no. 2, pp. 374–385, 2011.

[24] P.-U. Tournoux, J. Leguay, F. Benbadis, J. Whitbeck, V. Conan, and M. D. de Amorim, "Density-aware routing in highly dynamic dtns: The rollernet case," *IEEE Trans. on Mob. Comp.*, vol. 10, no. 12, pp. 1755–1768, 2011.

[25] N. Eagle and A. (Sandy) Pentland, "Reality mining: Sensing complex social systems," *Personal Ubiquitous Comput.*, vol. 10, no. 4, pp. 255–268, 2006.

[26] C. Boldrini, M. Conti, and A. Passarella, "Modelling social-aware forwarding in opportunistic networks," in *Performance Evaluation of Computer and Communication Systems. Milestones and Future Challenges*. Springer, 2011, pp. 141–152.

[27] A. Passarella and M. Conti, "Analysis of individual pair and aggregate intercontact times in heterogeneous opportunistic networks," *IEEE Trans. on Mob. Comp.*, vol. 12, no. 12, pp. 2483–2495, 2013.

[28] H. Tijms and J. Wiley, *A first course in stochastic models*. Wiley Online Library, 2003, vol. 2.

[29] R. Gail and L. Kleinrock, "An invariant property of computer network power," in *Conference Record, International Conference on Communications*, 1981, pp. 63–1.

[30] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Trans. on Mob. Comp.*, vol. 6, no. 6, pp. 606–620, 2007.

[31] T. Phe-Neau, M. Dias de Amorim, and V. Conan, "Vicinity-based DTN Characterization," in *ACM MobiOpp*, Zurich, Switzerland, Mar. 2012.

[32] ——, "The Strength of Vicinity Annexation in Opportunistic Networking," in *IEEE International Workshop on Network Science For Communication Networks (Netscicom)*, Torino, Italy, Apr. 2013.

[33] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 606–620, 2007.

[34] A.-K. Pietiläinen and C. Diot, "Dissemination in opportunistic social networks: the role of temporal communities," in *ACM Mobihoc*, Hilton Head, South Carolina, USA, Jun. 2012.

[35] P.-U. Tournoux, J. Leguay, F. Benbadis, J. Whitbeck, V. Conan, and M. D. de Amorim, "Density-aware routing in highly dynamic DTNs: The rollernet case," *IEEE Transactions on Mobile Computing*, vol. 10, pp. 1755–1768, 2011.

[36] A. Galati and C. Greenhalgh, "Human mobility in shopping mall environments," in *ACM MobiOpp*, Pisa, Italy, 2010, pp. 1–7.

[37] S. Gaito, E. Pagani, and G. P. Rossi, "Fine-Grained Tracking of Human Mobility in Dense Scenarios," in *IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, Rome, Italy, Jun. 2009.

[38] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, "A high-resolution human contact network for infectious disease transmission," *PNAS*, vol. 107, no. 50, pp. pp. 22 020–22 025, 2010.

[39] S. PalChaudhuri, J.-Y. Le Boudec, and M. Vojnovic, "Perfect Simulations for Random Trip Mobility Models," in *IEEE Infocom*, Miami, Florida, USA, Aug. 2005.

[40] M. Musolesi and C. Mascolo, "Designing mobility models based on social network theory," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 11, pp. 59–70, July 2007.

[41] T. Phe-Neau, M. Dias de Amorim, and V. Conan, "Fine-Grained Intercontact Characterization in Disruption-Tolerant Networks," in *IEEE Symposium on Computers and Communication*, Kerkyra, Greece, Jun. 2011.

[42] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson, *Introduction to Algorithms*, 3rd ed. The MIT Press, 2009.

[43] S. Heimlicher and K. Salamatian, "Globs in the Primordial Soup – The Emergence of Connected Crowds in Mobile Wireless Networks," in *ACM Mobihoc*, Chicago, Illinois, USA, Sep. 2010.

[44] M.-H. Zayani, V. Gauthier, and D. Zeghlache, "Improving link prediction in intermittently connected wireless networks by considering link and proximity stabilities," in *WOWMOM*, 2012, pp. 1–10.

[45] A. Tatar, T. Phe-Neau, M. D. de Amorim, V. Conan, and S. Fdida, "Beyond contact predictions in mobile opportunistic networks," in *IEEE/IFIP Annual Conference on Wireless On-demand Network Systems and Services*, Obergurgl, Austria, Apr. 2014.

[46] Z. Huang, X. Li, and H. Chen, "Link prediction approach to collaborative filtering," in *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*. ACM, 2005, pp. 141–142.

[47] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *J. Am. Soc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, May 2007.

[48] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi, "Human mobility, social ties, and link prediction," in *ACM KDD*, 2011, pp. 1100–1108.

[49] L. A. Adamic and E. Adar, "Friends and Neighbors on the Web," *Social Networks*, vol. 25, no. 3, pp. 211–230, 2003.

[50] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39– 43, 1953.

[51] A.-L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations," *Physica A: Statistical Mechanics and its Applications*, vol. 311, no. 3, pp. 590–614, 2002.

[52] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[53] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE Simulator for DTN Protocol Evaluation," in *International Conference on Simulation Tools and Techniques*, Rome, Italy, Mar. 2009.

[54] A. Leung, C. Mitchell, and R. Holloway, "A service discovery threat model for ad hoc networks," in *International Conference on Security and Cryptography*, Setubal, Spain, Aug. 2006.

[55] V. Lenders, J. Wagner, and M. May, "Analyzing the impact of mobility in ad hoc networks," in *ACM/Sigmobile Workshop on Multi-hop Ad Hoc Networks: From theory to reality*, Florence, Italy, May 2006.

[56] S. Zakhary and M. Radenkovic, "Utilizing social links for location privacy in opportunistic delay-tolerant networks," in *IEEE International Conference on Communications*, Ottawa, Canada, Jun. 2012.

[57] R. P. Jover, "Security attacks against the availability of LTE mobility networks: Overview and research directions," in *International Symposium on Wireless Personal Multimedia Communications*, Atlantic City, NJ, USA, Jun. 2013.

A Associated publications

- D. Karamshuk, C. Boldrini, Marco Conti, and Andrea Passarella, "SPoT: Representing the social, spatial, and temporal dimensions of human mobility with a unifying framework", *Pervasive and Mobile Computing*, Volume 11, Pages 19-40, April 2014.
- E. Biondi, C. Boldrini, A. Passarella, and M. Conti, "Duty cycling in opportunistic networks: inter-contact times and energy-delay tradeoff," IIT-CNR, Tech. Rep. 2013-TR-022, 2013.
- T. Phe-Neau, M. Dias de Amorim, and V. Conan, "Vicinity-based DTN Characterization", *ACM MobiOpp 2012*, Zurich, Switzerland, March 2012.
- A. Tatar, T. Phe-Neau, M. Dias de Amorim, V. Conan, and S. Fdida, "Beyond Contact Predictions in Mobile Opportunistic Networks", *IFIP/IEEE WONS*, Obergurgl, Austria, April 2014.

Pervasive and Mobile Computing 11 (2014) 19–40

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

Fast track article

SPoT: Representing the social, spatial, and temporal dimensions of human mobility with a unifying framework

Dmytro Karamshuk, Chiara Boldrini *, Marco Conti, Andrea Passarella

IIT-CNR, Pisa, Italy

ARTICLE INFO

Article history:
Available online 29 July 2013

Keywords:
Human mobility
Opportunistic networks
Complex networks
Stochastic processes

ABSTRACT

Modeling human mobility is crucial in the analysis and simulation of opportunistic networks, where contacts are exploited as opportunities for peer-to-peer message forwarding. The current approach to human mobility modeling has been based on continuously modifying models, trying to embed in them the mobility properties (e.g., visiting patterns to locations or specific distributions of inter-contact times) as they arose from trace analysis. As a consequence, with these models it is difficult, if not impossible, to modify the features of mobility or to control the exact shape of mobility metrics (e.g., modifying the distribution of inter-contact times). For these reasons, in this paper we propose a mobility *framework* rather than a mobility *model*, with the explicit goal of providing a flexible and controllable tool for modeling mathematically and generating simulatively different possible features of human mobility.

Our framework, named SPoT, is able to incorporate the three dimensions – spatial, social, and temporal – of human mobility. The way SPoT does this is by mapping the different social communities of the network into different locations, whose members visit with a configurable temporal pattern. In order to characterize the temporal patterns of user visits to locations and the relative positioning of locations based on their shared users, we analyze the traces of real user movements extracted from three location-based online social networks (Gowalla, Foursquare, and Altergeo). We observe that a Bernoulli process effectively approximates user visits to locations in the majority of cases, and that locations that share many common users visiting them frequently tend to be located close to each other. In addition, we use these traces to test the flexibility of the framework, and we show that SPoT is able to accurately reproduce the mobility behavior observed in traces. Finally, relying on the Bernoulli assumption for arrival processes, we provide a thorough mathematical analysis of the controllability of the framework, deriving the conditions under which heavy-tailed and exponentially-tailed aggregate inter-contact times (often observed in real traces) emerge.

1. Introduction

Due to the widespread diffusion of personal handheld devices such as smartphones and tablets, emerging wireless ad hoc networks are characterized by high user mobility, which ultimately leads to intermittent connectivity and end-to-end paths that are continuously changing or even completely lacking. Reversing the traditional approach, these potentially disconnected networks benefit from the exploitation of user mobility to bridge disconnected users in the network, and for this reason they are often referred to as *opportunistic networks* [1]. In opportunistic networks, messages are routed by

* Corresponding author. Tel.: +39 050 315 3504.
E-mail address: chiara.boldrini@iit.cnr.it (C. Boldrini).

1574-1192/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.pmcj.2013.07.011

the users of the network (which exchange them upon encounters with other users) and are eventually delivered to their destinations. The delay experienced by messages is thus a function of the users' mobility process. In particular, pairwise inter-contact times (i.e., the time intervals between consecutive contacts of a pair of nodes) are very important, since they characterize the temporal distance between two consecutive forwarding opportunities. Inter-contact times are determined by the movement patterns of users: users visiting the same locations will meet more frequently, and their inter-contact time will be shorter. Given the dependence of the delay on inter-contact times, characterizing the inter-contact time is therefore essential for modeling the performance of opportunistic networking protocols.

The first step in modeling human mobility is to understand how users move. Recently, starting from traces of real user movements, there has been a huge research effort in order to characterize the spatio-temporal (i.e., how users travel across locations [2–4]) and social (i.e., how the nature of a social relationship impacts on, for example, inter-contact times between two users [5,6]) properties of human mobility. There is a general agreement that users tend to travel most of the time along short distances while only occasionally following very long paths. In addition, user movements are generally characterized by a high degree of predictability: users tend to visit the same locations frequently, and to appear at them at about the same time. Less clear is how inter-contact times are characterized. Many hypotheses have been made (about them featuring an exponential distribution [7], a Pareto distribution [5], a Pareto with exponential cut-off distribution [8], and a lognormal distribution [6]), but the problem has yet to be solved. The fact is that inter-contact times are by nature heterogeneous, and trace analysis suggests that a one-distribution-fits-all approach is probably wrong.

Building upon the above findings, the current approach to human mobility modeling has so far been based on trying to incorporate in the model the newest features of mobility properties as they arose from trace analysis. Typically, each model focuses on just a few properties of human mobility. The class of location-based mobility models aims to realistically represent user mobility patterns in space. They are typically concerned with the regular reappearance to a set of preferred locations [9] or with the length of paths traveled by the users [10]. Similarly, there are models mostly focused on the accurate representation of the time-varying behavior of users, often relying on very detailed schedules of human activities [11,12]. Finally, the class of social-based mobility models aims to exploit the relation between sociality and movements, and to formalize social interactions as the main driver of human movements [13,14].

The disadvantage of the current approach to modeling human mobility is that the proposed models are intrinsically bound to the current state of the art on trace analysis, and typically need to be redesigned from scratch any time a new discovery is made. In addition, with current mobility models it is typically difficult, if not impossible, to fine tune the mobility properties (e.g., obtaining inter-contact times featuring a probability distribution with controllable parameters). Overall, flexibility and controllability are currently missing from available models of human mobility. *Flexibility* implies allowing for different distributions of mobility properties (e.g., return times to locations or inter-contact times) to be used with the model. The importance of flexibility is twofold. First, it gives the opportunity to evaluate networking protocols in different scenarios, and to test their robustness to different mobility behaviors. Second, it allows for changing the model upon new discoveries from trace analysis without the need to start over from a clean slate. On the other hand, *controllability* relates to the capability of obtaining a predictable output starting from a given input. This can be done only at a coarse granularity with the majority of available mobility models. For example, in social-based mobility, where social relationships determine the shape of inter-contact times, an appropriate configuration can lead to heavy-tailed inter-contact times [13]. However, there is no direct way for quantitatively controlling the parameters characterizing this distribution, and a fine tuning can be attempted only with a trial-and-error approach.

In light of the above discussion, the contribution of this paper is threefold. First, we propose (Section 3) a mobility framework (SPoT—Social, sPatial, and Temporal mobility framework) that incorporates the three dimensions of human mobility, while at the same time being flexible and controllable. SPoT takes as input the social graph representing the social relationships between the users of the network and the stochastic processes characterizing the visiting patterns of users to locations. Based on the input social graph, communities are identified and are assigned to different locations. Thus, people belonging to the same community share a common location where the members of the community meet. Then, users visit these locations over time based on a configurable stochastic process. The proposed framework thus builds a network of users and locations (called the *arrival network*), where a link between a generic user i and a location l characterizes the way user i visits location l. Overall, SPoT aims at being as accurate as possible in matching the real behavior of human movements while at the same time being tractable for mathematical analysis. In addition, the fact that it links together the three dimensions of human mobility provides a complete knowledge on the main mobility drivers, which are often exploited by networking protocols for opportunistic networks. For example, SPoT is superior to the direct generation of inter-contact times, since it also provides information on the social structure of the network. Knowing that a user belongs to a specific community can be very helpful when evaluating the performance of community detection schemes for opportunistic networks or social-aware forwarding protocols.

The second contribution of this work lies in studying the mobility behavior that emerges in real traces of human mobility (Section 4), and in using this information to address two open points in our framework, i.e., how to characterize the way users visit locations and how to position meeting places in the considered scenario. To this aim, we study three datasets of human self-reported whereabouts records obtained from the online location-based social networks (LBSNs) Gowalla [15], Foursquare [16], and Altergeo [17]. LBSN applications, where people can check into places (e.g., restaurants, offices) and share their location with friends, have become incredibly popular with the widespread diffusion of smartphones. From the check-in records we study the time sequences of individual user arrivals at places, and reveal that, for the majority of

user–place pairs, i.e., from 66% to 54% of the pairs, depending on the dataset, they are well approximated by a Bernoulli process, for which the intervals between consecutive arrivals follow a geometric distribution. Similarly, we show that the contact sequences between the majority of user pairs, i.e., from 78% to 87% of the total number of pairs, depending on the dataset, can be approximated by a Bernoulli process. As we show later in the paper, this finding is important, as the Bernoulli process features a number of properties that significantly simplify the mathematical analysis of the framework. We also use the check-in records to study the correlation between the distances between locations and the number of regular visitors they share. This property, to the best of our knowledge, has not been studied in the literature before. We find that locations that share many common users that visit them frequently tend to be located close to each other. We use this result to realistically position meeting places in the area of the modeling scenario.

The third contribution of the paper lies in showing that the proposed framework is at the same time flexible and controllable. More specifically, in Section 5, we show that SPoT is able to accurately reproduce the features of aggregate inter-contact times observed in the Gowalla dataset. This highlights the fact that the framework can be instantiated to a desired general mobility configuration by just changing its input parameters. On the other hand, in Section 6, we focus on the controllability of the framework, i.e., on its capability to generate a predictable output. Building upon the results of the analysis of real mobility data, we represent the way users arrive at locations as Bernoulli processes. Then, first we prove that, when the arrival processes are Bernoulli processes, the contact process between users is also a Bernoulli process, which is well aligned with the corresponding results of the data analysis (see Section 4). Finally, we mathematically derive the conditions under which heavy-tailed and exponentially-tailed aggregate inter-contact times emerge starting from simple, but heterogeneous, Bernoulli arrival processes for user visits to locations. This advances the knowledge on the dependence between aggregate and pairwise mobility statistics (explored for the first time in [18]), and confirms the main result in [18], i.e., that heterogeneity in pairwise statistics can lead to aggregate statistics that are very distant in distribution.

Please note that in this paper we focus on the ability of SPoT to produce a realistic output in terms of inter-contact times. As discussed above, inter-contact times are extremely important for the evaluation of opportunistic networks. For this reason, most network simulators, either public platforms [19] or custom simulators [20,21], are designed to work with contact-based traces as input. Alternatively, especially outside the opportunistic networks domain, network simulators can take as input information about node movements. This spatial output is not the main focus of the paper but, due to its relevance, in Section 7 we discuss how SPoT can be extended to generate a movement-based output. However, we leave the complete evaluation of the properties of this spatial output for future work.

2. Related work

A comprehensive overview of the state of the art in mobility modeling was presented in [22]. The work points out that the main findings in human mobility research can be classified along the three axes of *spatial*, *temporal*, and *connectivity* (or *social*) properties. Spatial properties pertain to the behavior of users in the physical space (e.g., the distance they travel), temporal properties to the time-varying features of human mobility (e.g., the time users spend at specific locations), and connectivity properties to the interactions between users. One of the first significant findings in human mobility, which highlighted the difference between our movements and random motion, was documented by Brockmann et al. [3], who analyzed a huge dataset of records of banknote circulation, interpreting them as a proxy of human movements. They showed that the travel distances, frequently called the *jump sizes*, of individuals follow a power-law distribution. This fits the intuition that we usually move over short distances, whereas occasionally we take rather long trips. Studying data collected tracing mobile phone users, Gonzalez et al. [2] extended the previous finding, showing that the distribution of jumps was that of a power law up to a certain point, after which the decay was exponential. In addition, they showed that individual truncated power-law trajectories co-exist with population-based heterogeneity. Thus, it was shown that the distribution of the radius of gyration – a measure which depicts the characteristic distance traveled by a user – can be approximated by a truncated power law. This suggests that the majority of people usually travel in close vicinity to their home location, while few of them frequently make long journeys.

As for the temporal properties of human movements, Gonzales et al. [2] detected the tendency of people to return to a previously visited location with a frequency proportional to the ranking in popularity of the location with respect to other locations. The authors also computed the *return time* probability distribution (probability of returning at time t to a selected place) and concluded that prominent peaks (at 24, 48, 72, …, h) capture the tendency of humans to return regularly (on a daily basis) to a location they visited before.

Connectivity properties have been extensively studied in the context of opportunistic networks research. In fact, as we have already discussed, the way users interact and get in touch with each other is crucial for message delivery. In particular, the time between two consecutive contacts of two devices contributes to the overall delay, while the duration of the contact bounds the size of the data that can be exchanged at each encounter. Typically, user interactions are measured through human-carried mobile devices, which are assumed to be proxies of real users. However, despite great efforts, a consensus has not been reached yet on how to exactly characterize the connectivity metrics in probabilistic terms.

From a taxonomy standpoint, the three dimensions of human mobility (spatial, temporal, and social) described above can be mapped into three different approaches to modeling human mobility: *maps of preferred locations*, *personal agendas*, and *social graphs*. The models of the first group account for the properties characterizing the regular reappearance of users at a set of preferred locations. Their general approach is to store the maps (i.e., the sets) of preferred places for each user,

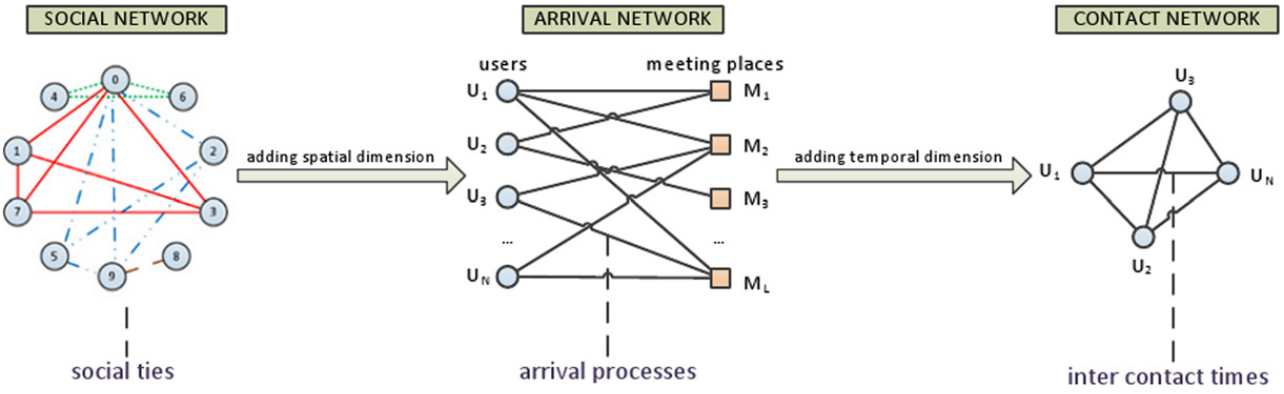


Fig. 1. Framework overview.

and to explore them while deciding on the next destination for his/her walk. The main representatives of this group are SLAW [10] and the model proposed by Song et al. [4]. These models are able to satisfy the main spatial properties of human mobility trajectories, but they do not pay enough attention to the social and temporal aspects of human movement.

The second class of models focuses on reproducing realistic temporal patterns of human mobility explicitly including repeating daily activities in human schedules. The most comprehensive approach of this group is presented in [12]. The model incorporates detailed geographic topology, personal schedules, and motion generators defined for more than 30 different types of activity. Although the model gives an extremely thorough representation of human movements in some specific scenarios, it does not explain the main driving forces of human mobility, and it is too complex for analytical tractability.

The most recent and most rapidly evolving trend in modeling human mobility is based on incorporating sociality into models, thus considering human relations as the main driver of individual movements. The main idea is that the destination for the next movement of a user depends on the position of people with whom the user shares social ties. The first models of this class of approaches were CMM [23] and HCMM [13], although others have recently been developed.

A recent work that is orthogonal to the above classifications is the work by Hossmann et al. [24]. They have found that, regardless of the modeling approach to human mobility, the contact graph (i.e., the graph whose vertices are the nodes of the network and whose edge weights are given by a combination of contact frequency and aggregate contact duration) generated by most synthetic models differs from that obtained from mobility traces. More specifically, traces tend to generate bridging links (only few strong edges connecting communities) in the contact graph, while synthetic models tend to generate bridging nodes (nodes linked to many other nodes). In addition to this result, Hossmann et al. found that contacts happening outside a community location are typically synchronized. In this paper, we do not consider synchronized meetings, in order to keep the framework mathematically tractable. Due to lack of space, we also do not verify whether bridging links are generated.

With respect to the related literature, SPoT covers and links together all the three dimensions of human mobility using a flexible and controllable framework, which can be instantiated to the desired mobility scenario and which is naturally suited for mathematical analysis. This work is an extended version of our previous paper in [25]. Specifically, here we have added the analysis of three relevant datasets extracted from the location-based online social networks Gowalla, Foursquare, and Altergeo. Results from trace analysis provide a strong case for Bernoulli arrivals, which are then used as the reference assumption in the mathematical analysis of the framework. In addition, we use these datasets to test the flexibility of the SPoT framework, showing that the latter is able to reproduce the mobility behavior observed in traces. With respect to [25], we also extend the mathematical analysis of the framework with the derivation of the settings under which exponentially-tailed aggregate inter-contact times (a case frequently encountered in traces) can be obtained. Finally, here we also discuss how SPoT can be extended for producing movement-based output.

3. The proposed mobility framework

In this section, we introduce our SPoT mobility framework, designed around the three main dimensions of human mobility, i.e., social, spatial and temporal (see Fig. 1). The social dimension is explicitly captured in the framework by taking a graph of human social relationships as an input parameter. This graph can be any well-known graph, such as a random graph [26] or a scale-free graph [26], or it can be extracted from real traces. Then, the framework adds the spatial dimension to the social ties by generating an arrival network, which is a bipartite graph that connects users and meeting places. A link between a user and a meeting place in the arrival network implies that the user visits that place during his/her movements. We exploit the fact that the structure of communities in the social graph has a significant impact on human mobility, and thus we assign users to meeting places such that communities of tightly connected users (*cliques*, in complex network terminology) share meeting places.

In order to add the temporal dimension to the model, we describe the way users visit the meeting places to which they are connected in terms of stochastic point processes [27]. A stochastic point process is a stochastic process that characterizes how events (*arrivals* at a location, in our framework) are distributed over time. By sampling from the random variables

Table 1
Algorithm for building the arrival network—Input: social graph G and removal probability α.

1.	Divide input social graph G into a set of overlapping cliques, such that the sizes of the cliques are maximum and each link is assigned to exactly one clique. To this aim, the Bron–Kerbosch algorithm [32] can be used.
2.	To each clique assign a separate meeting place, i.e., create a new meeting place and a set of links between this place and each member of the clique in the arrival network.
3.	Remove randomly each link in the social graph with probability α, inducing emergence of new nested cliques.
4.	Proceed to the next round starting from the first step, until there are no links left in the input graph.

representing the time between consecutive arrivals, we obtain the time sequences of the visits of a user to a given location. Then, the contact network, i.e., the network describing the contacts between nodes, can be obtained by assuming that two nodes are in contact with each other if they happen to be at the same time in the same meeting place.

3.1. The social and spatial dimensions of human mobility

Social interactions between users have emerged as one of the key factors defining human mobile behavior, because individuals belong to *social communities* and their social ties strongly affect their movement decisions [28,29]. As anticipated, in our analysis we consider proximity-based communities, i.e., communities whose members share a common meeting place (e.g., offices, bars, apartments). Since all members of the community visit a shared meeting place, this implies that users are socially connected with all other members of the community, and, therefore, form fully connected components (i.e., cliques) in the social graph.

Such cliques in realistic social networks exhibit *overlapping* and *hierarchical structure* [30,31]. Each user belongs to several overlapping cliques, representing different social circles (e.g., friends, relatives, colleagues). On the other hand, each clique is itself composed of a number of nested cliques, which share additional meeting places that are not common to all the users of a parent clique. For example, a company shares a set of offices visited by all its employees, while each subdivision has its own working places.

As anticipated, we represent the relation between the spatial and the social dimension of human mobility by means of a bipartite graph of users and meeting places, which we call the arrival network. In the algorithm (summarized in Table 1) for generating the arrival network starting from the input social graph we mainly need two components: a clique-finding algorithm (which also detects overlapping cliques) and a way of reproducing hierarchical cliques.

The first component corresponds to steps 1 and 2 in Table 1. In each round, the social graph is divided into a set (called the *cover*) of overlapping cliques, such that each link of the social graph is assigned to exactly one clique. To this purpose, we use the Bron–Kerbosch algorithm [32]. The cover of each round tries to capture the biggest possible cliques. For each of the newly identified cliques, we create a new meeting place and assign all members of the clique to that meeting place. In other words, we create a new meeting place vertex in the arrival network, and we add links between this vertex and all members of the community. As an example, we describe in Fig. 2 how cliques identified in the social graph are reflected into corresponding meeting places.

The second component (step 3 in Table 1) of the algorithm for generating the arrival network allows us to generate nested cliques. More specifically, our algorithm tries to identify cliques of lower size nested into those identified in the previous round. To do so, cliques are split according to a very simple random process, according to which every link in the social graph is randomly removed with a constant configurable probability α (the *removal probability*). This leads to the emergence of smaller cliques, which are indeed nested into the original ones. This simple strategy also has the advantage of allowing for a fine control of the number of meeting places shared by users. In fact, each link participates in a geometrically distributed (with parameter α) number of rounds of meeting place assignments. As each link is assigned to at most one clique per round, also the number of cliques that includes that link will be geometrically distributed. This implies that the pair of users i, j with which this link is associated will share a number L_{ij} of cliques (and thus of meeting places) that is itself geometrically distributed with parameter α.

The algorithm for generating the arrival network stops (step 4 in Table 1) when there are no more links in the social graph to be removed.

3.1.1. From meeting places to geographical locations

The analysis of the algorithm in Table 1 reveals that the number of meeting places generated grows with the number of cliques. Thus, the more cliques in the input social graph, the more meeting places are required. The proliferation of meeting places is not of great concern, as meeting places might correspond to very small geographic areas (e.g., offices). However, in order to improve the realism of the generated scenario, we combine these meeting places into a fixed number L of wider physical locations (e.g., this is equivalent to combining offices into a business center).

To assign meeting places to geographical locations, we explore the observation that, intuitively, the places that share many common frequent visitors should be located geographically close to each other, like in the case of different buildings of a university campus or different offices in a company building. In Section 4.4, we validate and confirm this observation using our datasets extracted from location-based online social networks. In order to quantify the closeness between two

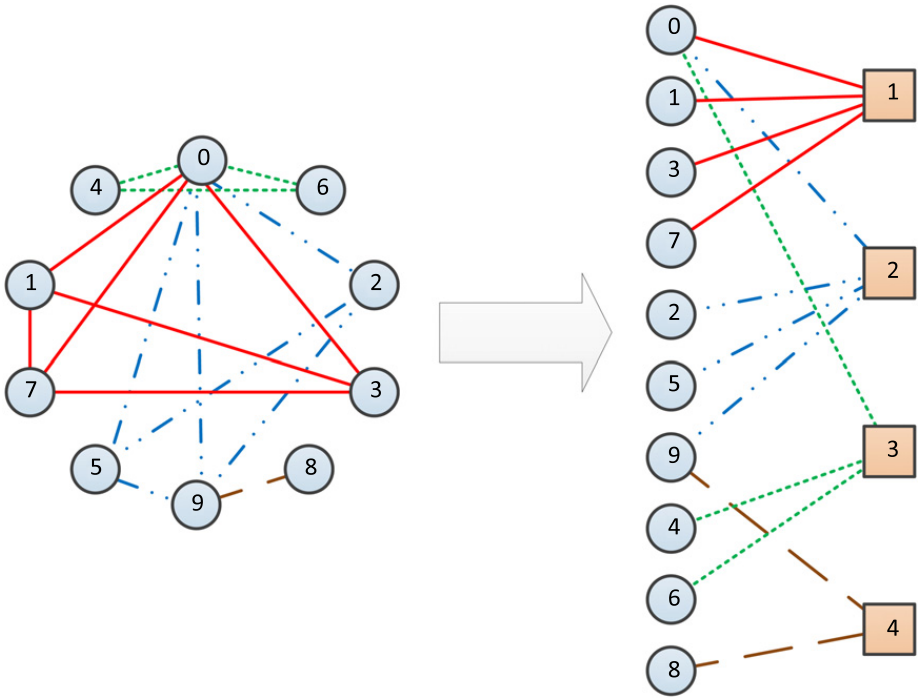


Fig. 2. A round of assigning social cliques to meeting place; cliques are marked with different line styles.

meeting places, we define the strength F_{ij} of ties between a pair of meeting places i and j as the summary co-appearing frequency across all the users the two places share. More formally, we can write F_{ij} as follows:

$$F_{i,j} = \sum_{u \in \mathcal{U}_{i,j}} f_u^i \times f_u^j, \tag{1}$$

where f_u^i is the frequency of user u's visits to location i and $\mathcal{U}_{i,j}$ is the set of all users shared between place i and j. The higher the arrival frequency of user u to both places i and j, the higher the strength between the two places. We anticipate here that the result of the analysis of realistic traces in Section 4.4 suggests that the mean and the median of the distance between two places i and j decreases with the strength F_{ij}. This validate our observation and allows us to exploit it for aggregating meeting places.

Our goal now is to distribute meeting places on the two-dimensional (2D) plane such that pairs of places with stronger ties in terms of shared visitors would be located closer. To this aim, we use a variation of the energy model for graph drawing described in [33,34]. In this model, the places are represented as particles, and particles connected with a link attract each other proportionally to the power of the strength of the link and inversely proportional to the power of the distance between the particles. Similarly, particles that are not connected with a link repulse each other. The final spatial positioning of the meeting places is achieved through simulation, in which the initial positions of the places are selected randomly in a rectangle of size $w \times h$. As a result of applying attractive and repulsive forces to the nodes, the system eventually reaches an equilibrium state in which tightly connected meeting places are situated close to each other, thus achieving our desired goal.

3.2. The temporal dimension of user visits to meeting places

The arrival network that we have built in the previous section tells us which are the meeting places visited by each user. Here we add the temporal properties of such visits. To this aim, we assign to each link in the arrival network a discrete stochastic point process A_i^l that describes the arrivals of user i at a meeting place l over time. In this work, we consider only discrete point processes, leaving the continuous case for future work. In a discrete point process, the time is slotted. During a time slot, each node visits a set of locations, and this set is determined by the evolution of the arrival processes.

In this paper, we assume that the processes A_i^l are independent, i.e., that nodes arrive at locations independently of each other. This implies that also the resulting contact processes are independent. In real traces, contacts can be synchronized [24], but coordination[1] between nodes may drastically complicate the mathematical analysis of mobility frameworks. For this reason, keeping in mind our target of controllability, we decided to limit the scope of the paper to independent arrival

[1] A weaker coordination involving only pairs of nodes has been sometimes assumed in the literature for modeling purposes. With pairwise coordination, pairs of nodes can still meet on purpose, but independently of the other pairs. However, since the tractability of our analytical framework would not benefit from this assumption, we decided to use the strongest independence condition.

processes. The comparison with traces (where these assumptions in general do not hold) presented in Section 5 shows that the accuracy of the model is good, nevertheless.

Once we have characterized the time at which users visit their assigned meeting places, we can build the contact graph of the network (Fig. 1). In fact, a contact between two users happens if the two users appear in the same meeting place at the same time slot. A contact duration is measured as the number of consecutive time slots in which two users have at least one commonly visited location. The contact graph can be fully mathematically characterized (we provide an example of this characterization in Section 6 for the case of arrival processes being heterogeneous Bernoulli processes) or it can be obtained from simulations.

4. Analysis of real user movements

As discussed in the previous section, the SPoT framework takes as input the social graph of the network users and the arrival processes describing how users visit locations. While the properties of the user social graph have been extensively studied in the literature [35,36], thus making their configuration easy, the statistical characterization of user arrivals has been little explored, especially in terms of the individual user-pair behavior. In order to address this open point in the framework (i.e., which arrival process is best indicated to describe how users visit places in reality), in this section we consider three datasets of real user movements, extracted from the location-based online social networks Gowalla [15], Foursquare [16], and Altergeo [17]. In location-based online social networks, users check into places (e.g., restaurants, offices) and share their location with their friends. Thus, the concept of check-ins is very similar to the arrivals considered in the SPoT mobility framework. In fact, both notions represent records of the time at which users visit particular venues. For this reason, we chose to take check-ins as proxies for user arrivals at places, and to use them to measure the temporal characteristics of arrival sequences.

In this section, we also use the three datasets for studying the features of pairwise inter-contact times in this real scenario. The pairwise results we will be later compared against the mathematical results in Section 6, showing that the data and model predictions are totally in agreement. Finally, we also study the geographic distribution of the meeting places sharing common visitors, whose results we used in Section 3.1.1 when defining the algorithm for aggregating meeting places into locations.

4.1. Collecting data

Here, we consider the datasets of check-ins collected from the three online location-based social networks Gowalla, Foursquare, and Altergeo. Each check-in record is stored as a tuple $\omega = (U, V, T) \in \Omega$, where U represents the user, V the venue, and T the time of the check-in. For a pair of user U_i and place V_l we consider a sequence of check-ins $\Omega_i^l = \{(U, V, T) \in \Omega : U = U_i \text{ and } V = V_l\}$, and denote the number of check-ins in a sequence as n_i^l ($n_i^l = |\Omega_i^l|$). We denote the total number of user–place pairs in the dataset with Q. We use venues as proxies of meeting places without performing any aggregation. Even in the case when two venues have similar coordinates, we treat them as two different meeting places.

4.1.1. Gowalla

The first dataset used in this paper comprises check-ins of Gowalla users collected via public API [24]. Launched in 2007, Gowalla was a pioneer location-based social network available via a mobile app for most of the major platforms (Android, iPhone, etc.). The Gowalla service was bought by Facebook in December 2011 and eventually shut down in 2012. The dataset considered in this paper accounts for $|\Omega_{\text{GO}}| = 27\text{M}$ check-in records collected from $|U_{\text{GO}}| = 619\text{k}$ users at $|V_{\text{GO}}| = 2.4\text{M}$ venues in the period of time from 21 January 2009 to 7 July 2011.

4.1.2. Foursquare

Foursquare was launched in 2009, and it has quickly become the most popular location-based service, with more than 35 million users as of January 2013 [16]. Similarly to Gowalla, Foursquare users receive bonuses for check-ins at places. Recently, Foursquare has become increasingly focused on being a tool for exploring nearby places, e.g., finding a restaurant, hotel, nightclub etc. Per user Foursquare check-in data are not directly accessible. However, users can opt to share their check-ins publicly on Twitter. Using Twitter's streaming API, it was possible to crawl publicly available check-ins [37]. Note that we can only access those check-ins that users explicitly choose to share on Twitter, although users have the possibility to set this option as default. In this paper, we consider a dataset of $|\Omega_{\text{FS}}| = 23\text{M}$ check-in records collected from $|U_{\text{FS}}| = 494\text{k}$ users at $|V_{\text{FS}}| = 2.3\text{M}$ venues across the United States in the period of time from 16 November 2010 to 19 September 2011.

4.1.3. Altergeo

Altergeo is an alternative online location-based social networking service focused on Russian-speaking countries. Launched in 2008, Altergeo has recently reported the audience of 1M+ users, mostly from big cities of Russia and Ukraine [17]. Similarly to Foursquare and Gowalla, the Altergeo service is available to its users as a check-in app. The service also explores check-in data for personalized food recommendation in a mobile phone app called Gvidi [38]. In this paper, we explore the dataset of $|\Omega_{\text{AG}}| = 700\text{k}$ check-ins that we collected from $|U_{\text{AG}}| = 49\text{k}$ Altergeo users at $|V_{\text{AG}}| = 94\text{k}$ places in the period of time between 12 February 2010 and 12 February 2012.

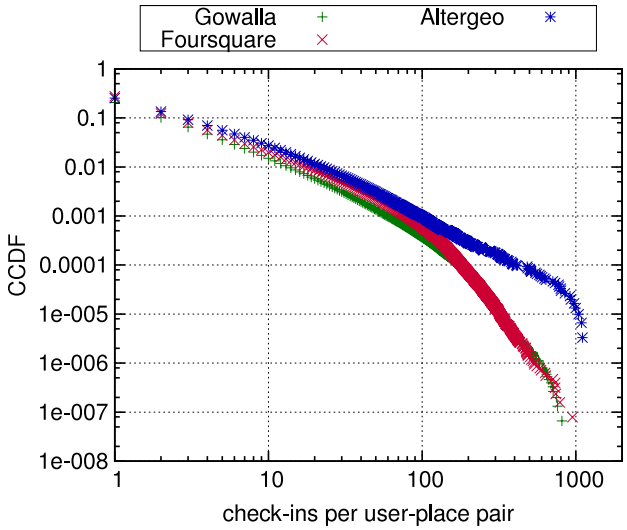


Fig. 3. Distribution of number of check-ins n_i^l per user–place pair for the three considered datasets.

Table 2
Statistics for the three considered datasets.

Characteristic	Gowalla	Foursquare	Altergeo
Check-ins, $\|\Omega\|$	27M	23M	0.7M
Users, $\|U\|$	619k	494k	49k
Places, $\|V\|$	2.4M	2.3M	94k
User–place pairs, Q	15M	13M	0.3M
Arrival sequences, Q'	94k	90k	2.4k
Contact pairs, C	46k	34k	998

4.2. Data preprocessing

In Fig. 3, we plot the distribution of the number n_i^l of check-ins per user–place pair across all considered datasets. As the plot suggests, the distribution of the number of check-ins for individual user–place pairs is extremely heterogeneous: while 80% of check-ins in places are never repeated, i.e., $n_i^l = 1$, there exist user–place pairs with a number of repeated check-ins higher than 800, i.e., $n_i^l > 800$. In order to deliver a reliable analysis from the statistical standpoint, we discarded those pairs with a small number of check-ins. Thus, we explore the datasets of $Q'_{\mathrm{GO}} = 94\mathrm{k}$, $Q'_{\mathrm{FS}} = 90\mathrm{k}$ and $Q'_{\mathrm{AG}} = 2.4\mathrm{k}$ user–place pairs each containing at least 20 check-in records, i.e., $n_i^l \geq 20$. We note that the resulting datasets account for $C_{\mathrm{GO}} = 46\mathrm{k}$, $C_{\mathrm{FS}} = 34\mathrm{k}$ and $C_{\mathrm{AG}} = 998$ contact pairs correspondingly. The summary statistics for all three considered datasets are given in Table 2.

In order to be able to obtain results that can be applied to a discrete-time mobility framework like the one we have defined in Section 3, we need to fix the duration of the time slot we consider. Since it has been shown [2,9] that users tend to appear at previously visited locations with a period of about 24 h, we decided to focus on time slots of 24 h length. This choice allows us to capture daily dynamics of user movements: people regularly commuting between home and work, working out at the gym, etc. The interested reader can refer to [39] for a detailed analysis under different granularities (8 h, 12 h). It suffices to mention here that in all cases the geometric hypothesis is not rejected for around 50% of pairs for arrivals and 75% of pairs for inter-contacts. As a general finding, the match between the model and traces tends to worsen with the reduced granularity. This effect is probably due to the lack of stationarity of user movements within the same day. In fact, while the days tend all to be similar to each other (apart from some deviations registered during weekends [9]), the different parts of the day tend to differ significantly (e.g., morning versus evening activities), even if the same user usually visits the same location at about the same time [2].

4.3. Analysis of individual inter-arrival times and inter-contact times

We now describe the methodology that we exploit to characterize the distribution of individual inter-arrival times and individual inter-contact times. From a preliminary analysis, we observed that across a significant population of user–place pairs the distribution of inter-arrival times has the shape of a straight line in lin–log scale, which roughly corresponds to a geometric distribution in the discrete case. Similarly, a preliminary observation of the pairwise inter-contact time yielded again a geometric distribution. We aim to validate this hypothesis by fitting individual inter-arrival time and inter-contact time distribution to a geometric distribution and evaluating the goodness of fit across all user–place pairs and user–user pairs, respectively, in the dataset.

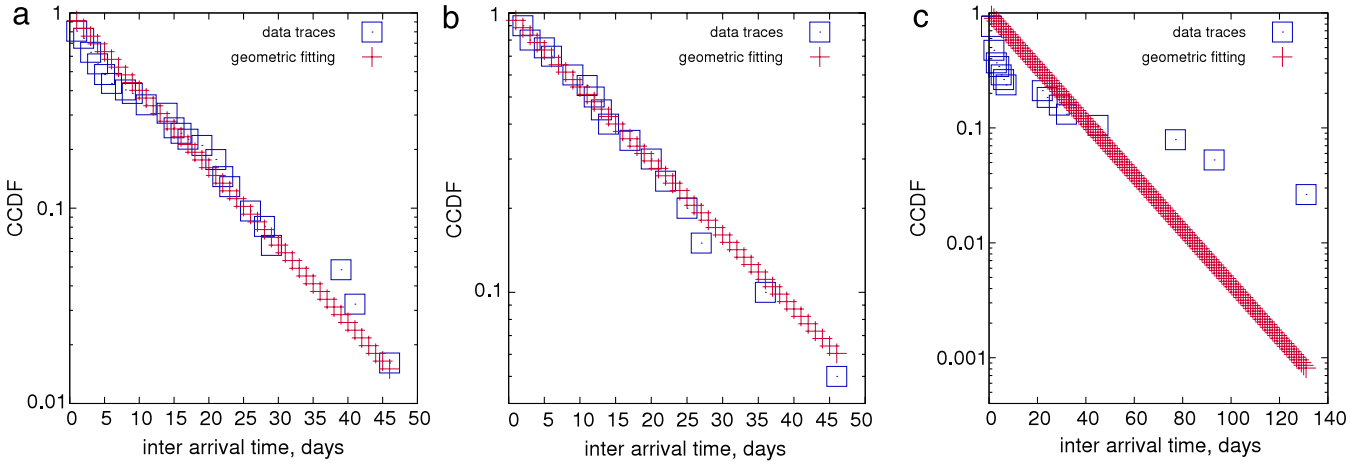


Fig. 4. Individual inter-arrival time distribution from traces (blue) versus geometric fitting (red) (a), (b) for the cases when the assumption of geometric distribution is not rejected (c) for the case when that assumption is rejected.

The fitting is performed using maximum likelihood estimation [40], which, in the case of a geometric distribution with success probability ρ, yields an estimator $\hat{\rho} = \frac{n_i^l}{\sum_{k=1}^{n_i^l} \tau_k}$ (where $\tau_1, \tau_2, \ldots, \tau_{n_i^l}$ are the n_i^l observations in the sample). Once we have fitted our data to a geometric distribution, we test whether it is plausible that our data come in fact from such fitted distribution. To this aim, we rely on one of the most popular goodness of fit tests, Pearson's chi-squared test [40], which works well for discrete distributions. In Pearson's chi-squared test, the test statistic TS is calculated as the sum of differences between observed and expected outcome frequencies (that is, counts of observations), each squared and divided by the expectation:

$$\mathrm{TS} = \sum_{k=1}^{n} \frac{(O_k - E_k)^2}{E_k}, \tag{2}$$

where n is the number of observations, O_k is an observed frequency for a bin k of values, and E_k is an expected frequency for a bin k. The test statistic TS follows, approximately, a chi-squared distribution with $K = (n - c - 1)$ degrees of freedom (i.e., $\mathrm{TS} \sim \chi^2_K$), where n is the number of non-empty bins and c is the number of estimated parameters for the distribution. In the case of a geometric distribution, $c = 1$; thus it follows that $K = n - 2$. If we denote with $q_{\chi^2_K, 1-\alpha}$ the $1 - \alpha$ quantile of χ^2_K, then the test rejects the geometric hypothesis at level α when $\mathrm{TS} > q_{\chi^2_K, 1-\alpha}$. In our analysis, we set α to 0.01 (similar to reference works in the literature [18,41]), which corresponds to a 0.01 probability of making a Type I error (i.e., rejecting the hypothesis when it is actually true). However, for the sake of completeness, we also report the results for $\alpha = 0.05$ and $\alpha = 0.001$ (Tables 3 and 4). Please note that, with all significance levels, for all datasets the percentage of pairs for which the geometric hypothesis is not rejected remains above 49%. For each pair, the number of bins is computed as the maximum number of bins that allows us to have at least five expected occurrences in each bin (the standard rule of thumb used with the chi-squared test). To estimate the goodness of fit across the population of individual inter-arrival times and inter-contact times, we calculate the percentage Q^{geom} of the user–place pairs and user–user pairs, respectively, for which the hypothesis of geometric distribution is not rejected.

4.3.1. Characterizing individual inter-arrival times

In Fig. 4, we plot the inter-arrival time distribution (blue dots) for three characteristic user–place pairs along with the corresponding fitted geometric distributions (red crosses) as estimated with the methodology described above. In the first two cases, the chi-squared test brings no evidence against the assumption of a geometric distribution of the inter-arrival times, as the calculated chi-squared statistics $\mathrm{TS}_{(a)} = 14.76$ and $\mathrm{TS}_{(b)} = 0.11$ are smaller than the corresponding quantiles for the chi-squared distribution $q_{\chi^2_{K_{(a)}}, 1-\alpha} = 18.48$ and $q_{\chi^2_{K_{(b)}}, 1-\alpha} = 9.21$, with $K_{(a)} = 7$ and $K_{(b)} = 2$ degrees of freedom and statistical significance level $\alpha = 0.01$. In contrast, in the latter case, the assumption is rejected, since $\mathrm{TS}_{(c)} = 72.43$ is bigger than the corresponding quantile $q_{\chi^2_{K_{(c)}}, 1-\alpha} = 9.21$ for the chi-squared distribution with $K_{(c)} = 2$ degrees of freedom.

We further calculate the percentage of user–place pairs for which the assumption of a geometric distribution of inter-arrival times is not rejected at different significance levels (Table 3). Thus, we observe that, for the majority of pairs across all datasets, i.e., $Q^{\mathrm{geom}}_{\mathrm{GO}} = 60\%$, $Q^{\mathrm{geom}}_{\mathrm{FS}} = 66\%$, $Q^{\mathrm{geom}}_{\mathrm{AG}} = 54\%$, the inter-arrival time distribution follows a geometric distribution. This result is important, as a geometric distribution of inter-arrival times can be modeled with a simple Bernoulli arrival process, which, as we discuss in Section 6, is very convenient for mathematical analysis. The implication behind Bernoulli arrivals is that users tend to visit places with a fixed rate. This matches the common finding [9] that users tend be quite regular in their movements.

Table 3
Percentage of pairs for which the geometric distribution hypothesis for arrivals is not rejected, at different significance levels.

α	Gowalla (%)	Foursquare (%)	Altergeo (%)
0.001	0.71	0.77	0.57
0.01	0.60	0.66	0.54
0.05	0.50	0.51	0.49

Table 4
Percentage of pairs for which the geometric distribution hypothesis for inter-contact times is not rejected, at different significance levels.

α	Gowalla (%)	Foursquare (%)	Altergeo (%)
0.001	0.89	0.94	0.84
0.01	0.80	0.87	0.78
0.05	0.51	0.74	0.56

4.3.2. Characterizing individual inter-contact times

We now analyze the pairwise inter-contact time sequences measured between consecutive contacts of the users in our datasets. In order to have statistically reliable results, we discarded pairs that have fewer than 20 contacts. The main obstacle in computing inter-contact times in our datasets is that there are no check-out records, i.e., no records of the time when users leave places. For this reason, we have to make some assumptions about the duration of the sojourn time at a location. In [24], the inter-contact times for the Gowalla trace (the exact same trace that we consider in this work) were measured assuming that a contact between two users happen if they have checked in less than 1 h apart at the same place. The rationale for this choice lies behind the nature of location-based online social networks like Gowalla, Foursquare, and Altergeo. In fact, these applications capture mostly users going out for eating or entertainment, for which the 1 h choice appears reasonable. Thus, in this work we also keep the 1 h assumption.

The plots for three significant pairs (blue dots) and the corresponding fitted geometric distributions (red crosses) are shown in Fig. 5. As we can guess from the plot, in the first two cases the assumption about geometric distribution of the inter-contact times is not rejected. In fact, in this case the chi-squared statistics $\mathrm{TS}_{(a)} = 5.91$ and $\mathrm{TS}_{(b)} = 0.51$ are smaller than the corresponding chi-squared distribution's quantiles $q_{\chi^2_{K_{(a)}},1-\alpha} = 20.09$ and $q_{\chi^2_{K_{(b)}},1-\alpha} = 9.21$, with degrees of freedom $K_{(a)} = 8$ and $K_{(b)} = 2$ and statistical significance $\alpha = 0.01$. In the third case (Fig. 5(c)), instead, the assumption of a geometric distribution of individual inter-contact times is rejected, since $\mathrm{TS}_{(c)} = 107.46$ is bigger than the corresponding quantile $q_{\chi^2_{K_{(c)}},1-\alpha} = 21.67$ from the chi-squared distribution with $K_{(c)} = 9$ degrees of freedom. Summarizing, the chi-squared test does not reject the geometric hypothesis for $Q^{\mathrm{geom}}_{\mathrm{GO}} = 80\%$, $Q^{\mathrm{geom}}_{\mathrm{FS}} = 87\%$, $Q^{\mathrm{geom}}_{\mathrm{AG}} = 78\%$ of pairs in our datasets.

4.4. Relative positioning of meeting places sharing common visitors

We now study the relationship between the relative positions of meeting places and the frequency of user visits to those places. More specifically, we investigate whether places that share many common users that visit them frequently happen to be located close to each other. To this end, we measure the social strength between pairs of meeting places in our dataset, exploiting the definition of social strength that we provided in Section 3.1.1. Please recall that the social strength between places i and j measures the co-appearing frequency across all users the two places share, and it is defined as $F_{i,j} = \sum_{u \in \mathcal{U}_{i,j}} f^i_u \times f^j_u$, where f^i_u is the frequency of user u's visits to location i and $\mathcal{U}_{i,j}$ is the set of users shared between places i and j. Intuitively, the social strength is higher if two places share a lot of common users that frequently visit both places.

In Fig. 6 we plot the median and mean values of the geographic distance for different values of social strength $F_{i,j}$ between venues in the biggest cities of each dataset, i.e., Austin for Gowalla, New York for Foursquare, and Moscow for Altergeo. As we can see from the plot, the distances between places tend to decrease with the strength, therefore suggesting that the places that share a lot of frequent users tend to be located close to each other. For instance, half of the venues with social strength between 2^3 and 2^4 are situated more than 6 km away from each other in Austin and Moscow and more than 4 km away in New York, whereas half of pairs of venues with very high social strengths of 2^{15}–2^{16} are placed not farther than 1–2 km away from each other across all datasets.

5. Testing the framework flexibility

While in previous sections we have introduced the SPoT framework and we have used real data from location-based online social networks to address the open points in the framework, in this section we start the evaluation of SPoT. More specifically, here we study the flexibility of SPoT, i.e., its capability to reproduce a desired, general, mobility behavior, while in Section 6 we test its controllability.

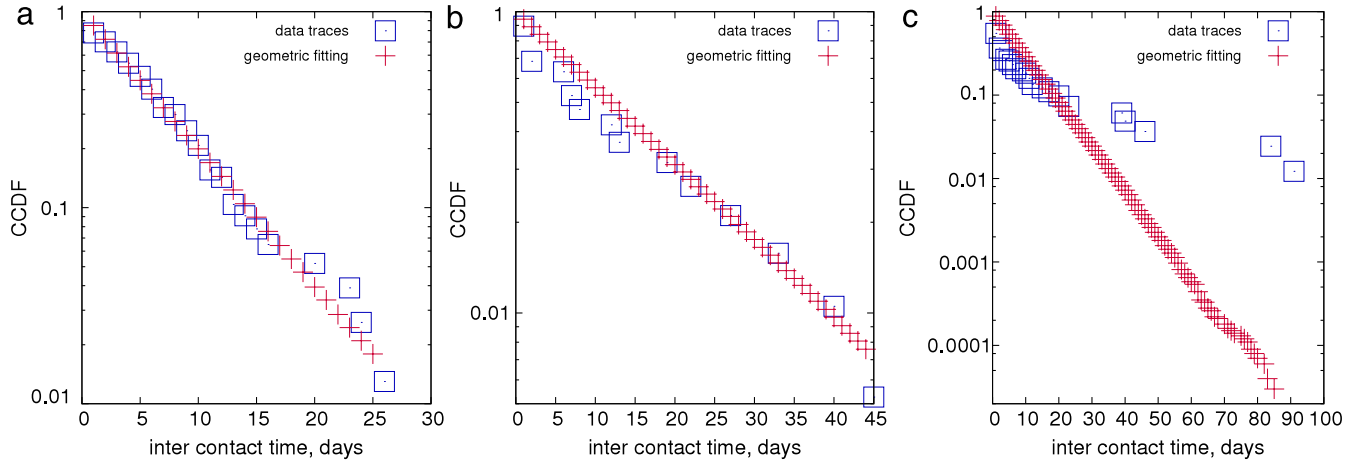


Fig. 5. Individual inter-contact time distribution from the data traces (blue) versus geometric fitting (red) (a), (b) for the cases when the assumption of geometric distribution is not rejected (c) for the case when that assumption is rejected.

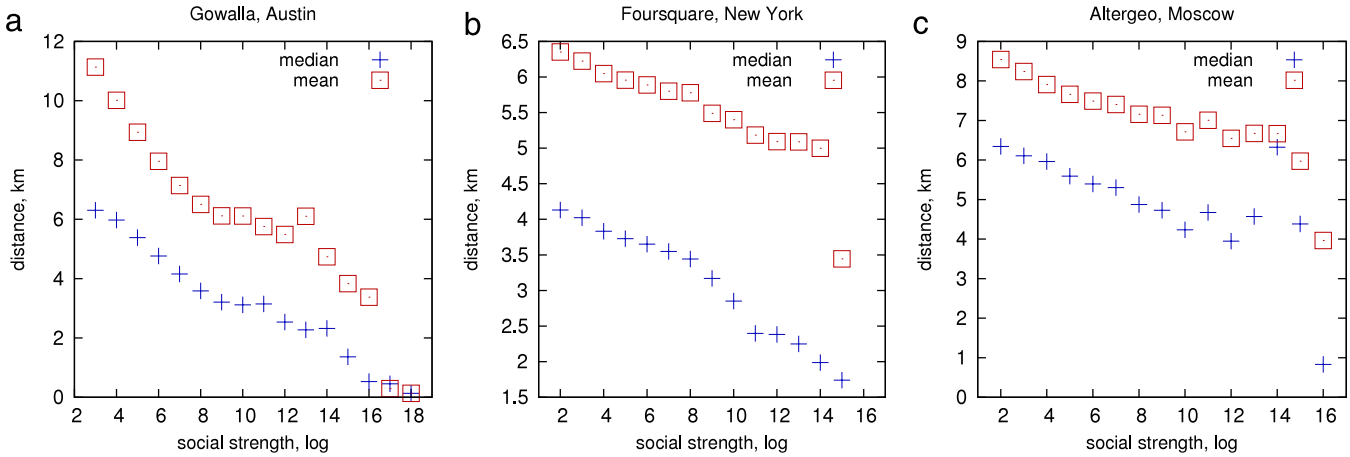


Fig. 6. Median and mean values of the geographic distance between venues in the city for different values of social strength $F_{i,j}$ grouped in logarithm bins for Gowalla, Austin (left), Foursquare, New York (center), and Altergeo, Moscow (right) datasets.

Our goal in this section is to show that the framework, once configured for the settings observed in a real mobility trace, generates the same aggregate characteristics (aggregate inter-contact times, specifically) as those seen in traces. We chose the aggregate over the pairwise statistics in this case because, from the mathematical characterization of the framework (Section 6), we know that SPoT, once configured with Bernoulli arrivals, will generate geometric inter-contact times. Since we also know from trace analysis (Section 4.3.2) that inter-contact times in the dataset can be approximated with a geometric distribution for a large fraction of pairs, a match between the inter-contact times generated by the framework and those seen in traces would be quite expected. Less obvious, however, is the capability of also reproducing a realistic aggregate behavior starting from pairwise controlled parameters. Please note that in the following we are not validating those aspects that we directly derived from traces (e.g., Bernoulli arrivals). Instead, we aim to evaluate if the proposed generative algorithm based on the creation of the arrival network is able to produce an output that matches the distribution derived from traces.

Aggregate inter-contact times are an important metric often used in the related literature [5,2]. While it has been shown [18] that their relation with pairwise inter-contact times might not always be as straightforward as was thought initially (i.e., in general, aggregate and pairwise inter-contact times do not follow the same distribution), aggregate inter-contact times are useful for several reasons. First, as discussed in [18], while it is true that in several cases the aggregate distribution does not contain enough information to assess the performance of opportunistic networks (which always depends on the pairwise distribution), there are other cases where looking at the aggregate distribution is sufficient to assess some performance aspects. For example, in addition to the cases where the aggregate and individual distributions belong to the same family, highlighted in [18], we have also shown in [42] that, when the aggregate distribution does not present a heavy tail, we can be sure that no pairwise distribution will have a heavy tail and thus, for example, no convergence problems will arise for randomized opportunistic routing protocols. Moreover, due to the increasing privacy concerns that are emerging in relation to collecting datasets with potentially sensitive information about people, in the future we can reasonably expect that it will be easier to find and distribute datasets with samples of the aggregate distribution rather than of pairwise distributions, as the former in general do not disclose the behavior of the individuals. So, aggregate information might be the only information available to researchers. In this case, having a mobility framework able to control not only

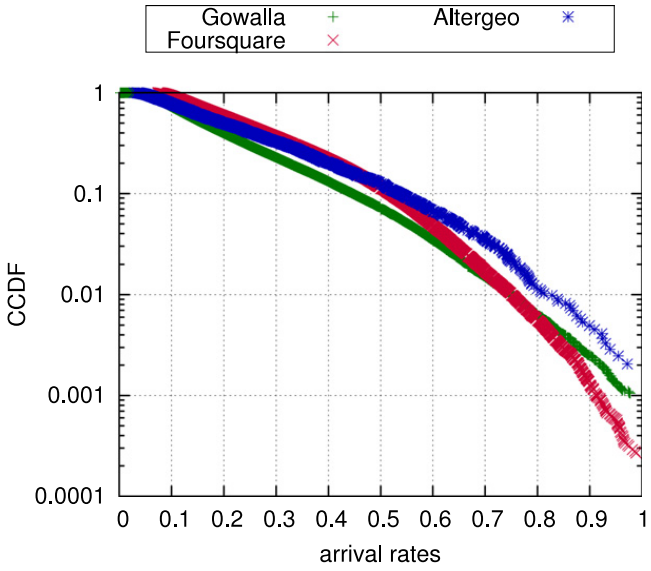


Fig. 7. Distribution of arrival rates in the Gowalla, Foursquare, and Altergeo traces.

the pairwise but also the aggregate statistics will be important. Finally, for any possible dataset, the statistical confidence of fitting the aggregate distribution is (much) higher than for the individual pairwise distributions (due to the greater number of samples available, by definition, for the former). Therefore, considering the aggregate distribution may be the only way to obtain statistically relevant fittings.

In order to use the framework, we need to configure the following quantities: the social graph G, the removal probability α, and the arrival processes A_i^l for each user i visiting location l. We extract this information from the data traces themselves, relying on the same subset of users (those with at least 20 check-ins) that we have used in Section 4. We take users and friendship records from the dataset to construct the social graph G. In order to estimate the removal probability α from the trace, we recall that this probability is the reciprocal of the average number of places L_{ij} shared between a pair of users i and j, i.e., $\alpha = \frac{1}{E[L_{ij}]}$ (see Section 3 for more details). From the analysis of the traces we compute the sample mean: $\hat{E}[L_{ij}^{\mathrm{GO}}] = 1.22$ for Gowalla, $\hat{E}[L_{ij}^{\mathrm{FS}}] = 1.60$ for Foursquare, and $\hat{E}[L_{ij}^{\mathrm{FS}}] = 1.64$ for Altergeo. From these we calculate the corresponding removal probabilities $\hat{\alpha}_{\mathrm{GO}} = 0.82$, $\hat{\alpha}_{\mathrm{FS}} = 0.63$, and $\hat{\alpha}_{\mathrm{AG}} = 0.61$. In order to configure the arrival processes A_i^l we exploit the result from Section 4, and we set them to be Bernoulli processes. We configure the rates of such processes so that they match the empirical rate distribution derived from the trace (Fig. 7). More specifically, the rate distribution in Fig. 7 is obtained by aggregating the arrival rates for each user–location pair extracted from the traces. The use of these rate distributions allows us to maintain the statistical properties of the arrival process, regardless of the actual number of users or locations that we actually simulate.

In Fig. 8, we show the aggregate inter-contact time generated by SPoT against those observed in the traces. As we can see from the plot, the aggregate behavior observed in the traces (red squares) is in good agreement with the corresponding results from the simulation (blue crosses). This confirms the flexibility of the framework to capture a desired realistic behavior seen in real traces.

6. Testing the framework controllability

In this section, we show mathematically how the SPoT framework is able to produce different controllable outputs depending on its initial configuration. To this aim, we exploit the data analysis results and we focus on Bernoulli arrivals, which we have shown in Section 4 to represent the behavior of the majority of user–place pairs. Using the Bernoulli assumption, in this section we fully characterize the pairwise dynamics of the framework and we also analytically derive the conditions under which heavy-tailed and exponentially-tailed aggregate inter-contact times, two cases often observed in real traces, emerge.

In our analysis we use the term *contact process* to describe how users meet with each other. Assuming that two users U_i and U_j can meet at L_{ij} distinct meeting places, the contact process between users i and j comprises all contacts happening at all L_{ij} shared meeting places. The time between consecutive contacts in the contact process defines the inter-contact times between the pair of nodes. In the following, we also characterize the *single-place contact process*, as the contact process between users U_i and U_j limited to a specific meeting place M_l.

As anticipated, in this analysis we model arrival processes as Bernoulli processes, since they feature geometric inter-arrivals like those seen in traces (Section 4.3.1). In a Bernoulli process, the probability of an arrival (a success, in Bernoulli processes terminology) in a given time slot is constant and corresponds to the rate of the process, defined as the expected

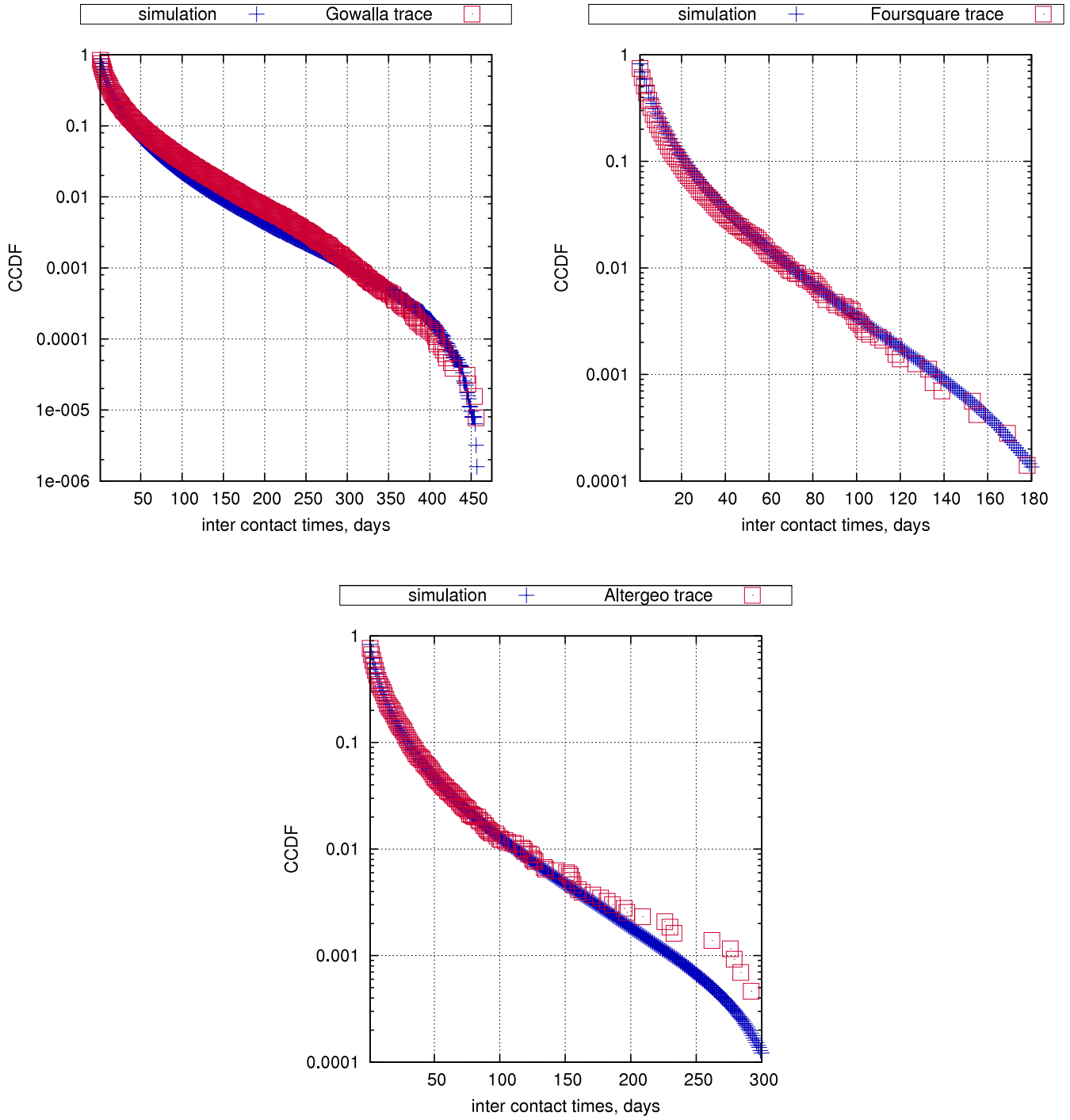


Fig. 8. Aggregate inter-contact times obtained from traces (red squares) and from simulations (blue crosses).

frequency of arrivals.[2] Hence, in the following we will use the terms probability and rate interchangeably. We show that, if the individual arrival processes are Bernoulli processes, then the contact process and the single-place contact process are also Bernoulli processes for any pair of users. As the inter-arrival times for a Bernoulli process follow a geometric distribution, we obtain that from geometric inter-arrival times at specific meeting places (corresponding to Bernoulli arrivals) a geometric distribution of pairwise inter-contact times follows, exactly as seen in the traces.

Additionally, we show that the rates of the contact processes depend on the rates of the arrival processes. Starting from this dependence, we are able also to derive analytically the aggregate inter-contact times as a function of the arrival rates

[2] It is straightforward to prove the equivalency between the rate ρ and the success probability p of the Bernoulli process. The rate can in fact be computed as the expected number of successes in n trials divided by the number of trials. Since the expected number of successes in n trials can be computed as the expectation of the binomial distribution with parameters p and n, we get that $\rho = \frac{np}{n} = p$. Please note also that the rate of the Bernoulli process takes thus values in [0, 1].

Table 5
Table of notation.

N	Number of users in the arrival network
L	Number of meeting places in the arrival network
U_i	User i
M_l	Meeting place l
L_{ij}	Number of shared meeting places between users U_i and U_j
A_i^l	Arrival process of user U_i at meeting place M_l
C_{ij}^l	Single-place contact process between users U_i and U_j at meeting place M_l
C_{ij}	Contact process between users U_i and U_j
$\rho_{A_i^l}$	Rate of arrival process A_i^l
$\rho_{C_{ij}^l}$	Rate of single-place contact process C_{ij}^l
$\rho_{C_{ij}}$	Rate of contact process C_{ij}
$E[P]$	Expectation of the rate of pairwise inter-contact times
$F_\rho(\tau)$	CCDF of individual inter-contact times τ between a pair of nodes whose rate is equal to ρ
$f_P(\rho)$	PDF of the rates of individual inter-contact times
$F(\tau)$	CCDF of the aggregated inter-contact times

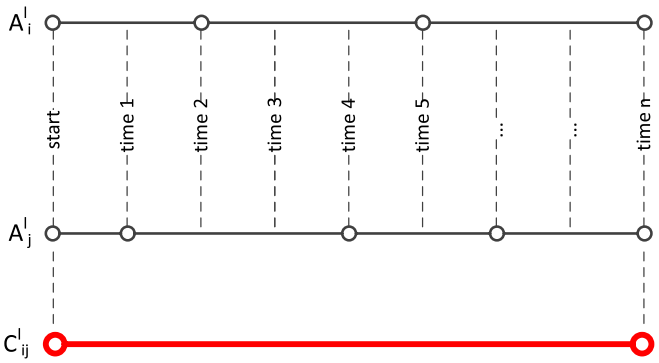


Fig. 9. The single-place contact process as an intersection of arrival processes.

of users at meeting places. Although this dependence is not trivial in the general case, we show that different shapes of the aggregate inter-contact distribution can be obtained starting from simple Bernoulli arrival processes. More specifically, we focus on the two cases frequently reported in the related literature, namely, when the aggregate inter-contact time has a power-law or an exponential tail. We show that the latter emerges in homogeneous networks when all the rates of individual Bernoulli processes are equal, and the former when the rates follow a specific distribution.

Before proceeding to the details of our analysis, we first introduce the notation used throughout the section. We consider an arrival network made up of N users and L meeting places. We assume that each user U_i visits place M_l according to a Bernoulli process A_i^l with rate $\rho_{A_i^l}$. For each meeting place M_l, and for each pair of users U_i and U_j, we characterize the single-place contact process C_{ij}^l (of rate $\rho_{C_{ij}^l}$) and the contact process C_{ij} of rate $\rho_{C_{ij}}$, aggregated over the L_{ij} shared meeting places. The latter defines the distribution of pairwise inter-contact times. We denote the complementary cumulative distribution function (CCDF) of the pairwise inter-contact times of rate ρ with $F_\rho(\tau)$, and that of the aggregate inter-contact times with $F(\tau)$. $F(\tau)$ is obtained as a function of the probability density function (PDF) of the rates of individual inter-contact times $f_P(\rho)$. The notation is summarized in Table 5. The complete proofs for the results shown in this section, when not provided inline, can be found in the associated technical report [39].

6.1. Contact process for a pair of users

In this section, assuming Bernoulli arrivals at locations, we analytically characterize the contact process between a pair of users. To this aim, consider two Bernoulli processes A_i^l and A_j^l, describing arrivals of users U_i and U_j at a shared place M_l. For a Bernoulli process, the probability $0 < \rho \leq 1$ of an arrival in a time slot τ is constant (i.e., does not depend on τ), and is called the *parameter* or the *rate* of the process. Moreover, the time intervals between arrivals are independent geometrically distributed random variables.

We assume that individual arrival processes are independent, and that a contact between two users happens if both decide to visit place M_l in the same time slot. Thus, the single-place contact process C_{ij}^l between user pair U_i, U_j at meeting place M_l can be obtained from the intersection of the individual Bernoulli arrival processes of users U_i and U_j at meeting place M_l. An example of the intersection of individual arrival processes is provided in Fig. 9. In the following lemma, we prove that the single-place contact process C_{ij}^l is also a Bernoulli point process.

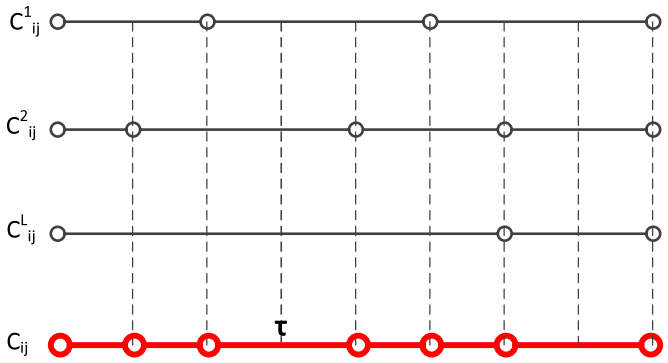


Fig. 10. The compound contact process as a merging of single-place contact processes.

Lemma 1 (*Single-Place Contact Process*)**.** *The single-place contact process C_{ij}^l resulting from independent Bernoulli arrival processes A_i^l and A_j^l, of rates $\rho_{A_i^l}$ and $\rho_{A_j^l}$ respectively, is a Bernoulli process of rate $\rho_{C_{ij}^l} = \rho_{A_i^l} \times \rho_{A_j^l}$.*

Proof. The probability of a contact at meeting place M_l is equal to the probability that both users are at meeting place M_l in the same time slot. This can be obtained as the product $\rho_{A_i^l} \times \rho_{A_j^l}$, recalling that, for a Bernoulli process, the rate of the process is equal to the probability of an arrival in a time slot. A discrete stochastic process in which arrivals happen with constant probability $\rho_{C_{ij}^l} = \rho_{A_i^l} \times \rho_{A_j^l}$ is again a Bernoulli process of rate $\rho_{C_{ij}^l}$. □

In the following, we focus on the contact process between a pair of users U_i, U_j, i.e., on their contacts in the L_{ij} shared meeting places. A contact happens between the two users in a given time slot if they meet in at least one of the L_{ij} meeting places that they share. Thus, the contact process between users U_i and U_j can be obtained by merging (as shown in [27]) their single-place contact processes (Fig. 10). In the following theorem, we show that, if the single-place contact processes are Bernoulli processes, then the contact process is also a Bernoulli process.

Theorem 1 (*Contact Process*)**.** *The contact process C_{ij} between contacts resulting from a number L_{ij} of individual place contact processes C_{ij}^l, which, in their turn, emerge from Bernoulli arrival processes A_i^l and A_j^l of rates $\rho_{A_i^l}$ and $\rho_{A_j^l}$, is a Bernoulli process of rate $\rho_{C_{ij}} = 1 - \prod_{l=1}^{L_{ij}}(1 - \rho_{A_i^l} \times \rho_{A_j^l})$.*

Proof. The probability of at least one contact in a time slot can be computed as one minus the probability of no contact in that time slot. The probability of no contact in the time slot is equal to the probability that the two users do not meet in any of their shared meeting places. As follows from Lemma 1, the probability of a contact in a single shared place is constant and equal to $\rho_{A_i^l} \times \rho_{A_j^l}$. Therefore, the probability of at least one contact in a time slot is also constant and equal to $\rho_{C_{ij}} = 1 - \prod_{l=1}^{L_{ij}}(1 - \rho_{A_i^l} \times \rho_{A_j^l})$. It then follows that the sequence of time slots with at least one contact forms a Bernoulli process of rate $\rho_{C_{ij}}$. □

The contact process described in Theorem 1 also defines the time intervals between consecutive contacts of a pair of users. Specifically, for a Bernoulli process the distribution of inter-contact times is geometric. We summarize this result in the following corollary.

Corollary 1 (*Pairwise Inter-Contact Times*)**.** *The inter-contact time distribution between a pair of users U_i and U_j, meeting at a number L_{ij} of meeting places, and whose arrivals at these meeting places are described as Bernoulli arrival processes A_i^l and A_j^l of rates $\rho_{A_i^l}$ and $\rho_{A_j^l}$, is geometric, with the following rate:*

$$\rho = 1 - \prod_{l=1}^{L_{ij}}(1 - \rho_{A_i^l} \times \rho_{A_j^l}). \tag{3}$$

Please note that the above result is perfectly in agreement with what we have seen in traces (Section 4).

6.2. Aggregate contact process

In this section, we describe how to derive the aggregate inter-contact times starting from pairwise inter-contact times featuring a geometric distribution. More specifically, we solve two cases by providing the conditions on the Bernoulli arrival processes of users at locations such that the resulting aggregate inter-contact time distribution is either heavy tailed or exponential. The two cases are important, as they have often emerged from the analysis of real mobility traces [5,7]. Our

derivation shows how these different aggregate behaviors can result from simple heterogeneous Bernoulli arrival processes, which are very convenient to deal with for mathematical analysis. This result also confirms the main finding of [18]: very different aggregate statistics can emerge from the heterogeneity of simple pairwise statistics.

In order to derive the aggregate inter-contact times, we exploit the result in [18], which describes the dependence between the aggregate inter-contact time distribution and the inter-contact time distributions of individual pairs of users. Specifically, the authors consider a heterogeneous scenario, in which the pairwise inter-contact time distributions are all of the same type (e.g., exponential), but whose parameters (the rates, in the exponential example) are unknown a priori. The rates of the individual contact sequences are drawn from a given distribution, which, therefore, determines the specific parameters of each pair's inter-contact times. The model described in [18] shows that both the distribution of the rates and the distributions of pairwise inter-contact times impact on the aggregate distribution. For the convenience of the reader we recall this result in Theorem 2.

Theorem 2. *In a network where the rates of pairwise inter-contact times are distributed according to a continuous random variable P with density $f_P(\rho)$, the CCDF of the aggregate inter-contact time is as follows:*

$$F(\tau) = \frac{1}{E[P]} \int_0^\infty \rho f_P(\rho) F_\rho(\tau) \mathrm{d}\rho, \tag{4}$$

where $F_\rho(\tau)$ denotes the CCDF of the inter-contact times between a pair of nodes whose rate is equal to ρ.

Please note that, while originally derived for inter-contact times featuring a continuous distribution, Theorem 2 can be used also for discrete inter-contact times. In fact, the integral in Eq. (4) depends on ρ, which was continuous in [18], and it is still continuous here. Thus, discrete inter-contact times do not change the expression for $F(\tau)$, except that now Eq. (4) only holds for discrete values of τ.

In Corollary 2, we extend the finding in Theorem 2 to our network of interest, where pairwise inter-contact times depend on their corresponding arrival processes. We have shown in Corollary 1 that, for the case of independent Bernoulli arrival processes, the distribution of individual inter-contact times is geometric. In other words, the shape of the pairwise inter-contact time distribution $F_\rho(\tau)$ is fixed in our model and, thus, the resulting aggregate inter-contact time characteristic is controlled by the distribution of the rates of individual inter-contact times $f_P(\rho)$. This distribution, in turn, depends on the distribution of the corresponding arrival rates. This dependence may not be trivial in the general case.

In order to apply Theorem 2 to our case of pairwise inter-contact times featuring a geometric distribution, we note that a discrete random variable X featuring a geometric distribution with rate ρ can be expressed in terms of a discrete random variable Y featuring a discrete exponential distribution. More specifically, the CCDF[3] of the geometric distribution of the pairwise inter-contact times, i.e., $F_\rho(\tau) = (1-\rho)^\tau$, $\tau \in \{1, 2, 3, \ldots\}$, can be rewritten in a discrete exponential form, i.e., $F_\lambda(\tau) = e^{-\lambda\tau}$, $\tau \in \{1, 2, 3, \ldots\}$, by substituting $\rho = 1 - e^{-\lambda}$, where $\lambda \in (0, \infty)$. Variables X and Y are thus exactly the same, but *written in a different form*. Using this substitution, we derive the following corollary of Theorem 2.

Corollary 2. *In a network in which the pairwise inter-contact times follow a geometric distribution with rate ρ, or, equivalently, a discrete exponential distribution with parameter $\lambda = -\ln(1-\rho)$, the CCDF of the aggregate inter-contact time is given by the following:*

$$F(\tau) = \frac{\int_0^\infty (1-e^{-\lambda}) e^{-\lambda\tau} f_\Lambda(\lambda) \mathrm{d}\lambda}{\int_0^\infty (1-e^{-\lambda}) f_\Lambda(\lambda) \mathrm{d}\lambda}. \tag{5}$$

In the above equation, function $f_\Lambda(\lambda)$ denotes the density of the parameters of pairwise inter-contact times.

In the remainder of this section, we show under which arrival rate distribution it is possible to obtain heavy-tailed and exponentially-tailed aggregate inter-contact times, two specific cases frequently reported in the literature.

6.2.1. Modeling a heavy-tailed distribution of aggregate inter-contact times

In this section, we study under which arrival rate distribution heavy-tailed aggregate inter-contact times are obtained. To this aim, using Corollary 2, we first derive in Lemma 2 the pairwise contact rate distribution that leads to heavy-tailed aggregate inter-contact times.

Lemma 2. *In a network in which the pairwise inter-contact times have a discrete exponential distribution of the form $F_\lambda(\tau) = e^{-\lambda\tau}$, $\tau \in \{1, 2, 3, \ldots\}$, and the parameters λ are drawn from an exponential distribution with rate a, the aggregate inter-contact time distribution is as follows:*

$$F(\tau) = \frac{a + a^2}{(\tau + a)(\tau + a + 1)} \quad \left(x \to \infty \Rightarrow F(\tau) \sim 1/\tau^2\right). \tag{6}$$

The complete proof for the above lemma and for all results introduced below can be found in [39].

[3] Please note that the corresponding probability mass function is given by $e^{-\lambda\tau}\left(1 - e^{-\lambda}\right)$ and adds up to one, thus showing that the discrete exponential is a properly defined distribution.

Lemma 2 says that the aggregate inter-contact time distribution decays proportionally to the power $\gamma = -2$ of τ, i.e., $F(\tau) \sim 1/\tau^2$, if the distribution of the parameters λ of the individual inter-contact times is exponential. In the rest of this section we develop this case, and show how the exponential distribution of the parameters of the individual inter-contact times emerges in an arrival network with independent Bernoulli arrival processes.

As we have already shown, the distribution of the parameters of pairwise inter-contact times depends on the distribution of the corresponding arrival rates. This dependence is described by Eq. (3), which, after substitution of $\rho_{C_{ij}}$ with λ, according to what we discussed above, takes the form $\lambda = \sum_{l=1}^{L_{ij}} -\ln(1 - \rho_{A_i^l} \times \rho_{A_j^l})$. From this dependence, we find a distribution of arrival rates $\rho_{A_i^l}$ such that the conditions of Lemma 2 are satisfied, i.e., the distribution of the parameters λ of the individual inter-contact times is exponential. To this aim, we prove the following lemma.

Lemma 3. *If the individual arrival processes are independent Bernoulli point processes, the rates $\rho_{A_i^l}$ of the processes are drawn such that $\rho_{A_i^l} = e^{-\frac{1}{2}Y^2}$, where Y is a standard normal random variable, and the number of shared meeting places L_{ij} between pairs of users is a geometric random variable with parameter α, then the resulting pairwise inter-contact time parameters λ are exponentially distributed with parameter α.*

A condition for Lemma 3 to be applicable is that the number of shared meeting places between pairs of users is geometrically distributed. Recall that this type of distribution is secured by the arrival network generating algorithm described in Section 3. Therefore, the result of Lemma 3 can be applied to the networks generated by the mobility framework. Finally, we combine the results of Lemmas 2 and 3 in the following theorem.

Theorem 3 (*Heavy-Tailed Aggregate Inter-Contact Times*)**.** *If the individual arrival processes are independent Bernoulli point processes, the rates $\rho_{A_i^l}$ of the processes are drawn such that $\rho_{A_i^l} = e^{-\frac{1}{2}Y^2}$, where Y is a standard normal random variable, and the number of shared meeting places L_{ij} between pairs of users is a geometric random variable with parameter α, the CCDF of the aggregated inter-contact times is given by Eq. (6).*

6.2.2. Modeling an exponential distribution of aggregate inter-contact times

In this section, we show that the aggregate inter-contact time distribution has an exponential decay if the arrival processes are homogeneous. To this end, we first consider the case when the number of shared meeting places L_{ij} between pairs of users is constant, and prove that in these conditions the aggregate inter-contact times result in a discrete exponential (i.e., geometric) distribution. Formally, we get the following result (the proof can be found in [39]).

Theorem 4 (*Exponential Aggregate Inter-Contact Times*)**.** *If the individual arrival processes are independent Bernoulli point processes with homogeneous rates $\rho_{A_i^l} = \beta$ and the number of shared meeting places L_{ij} between pairs of users is constant, i.e., $L_{ij} = L$, then the aggregated inter-contact times follow a discrete exponential (i.e., geometric) distribution with CCDF*

$$F(\tau) = e^{-\gamma\tau}, \tag{7}$$

where $\gamma = -L\ln(1 - \beta^2)$.

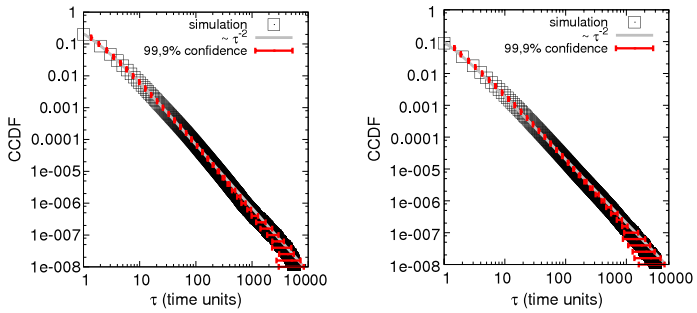
In the above case, we have shown that an exponential inter-contact time distribution emerges if we put additional constraints on the number of shared meeting places L_{ij}, i.e., we assume that L_{ij} is constant across all pairs of users. Below, we consider a more general scenario in which the number of shared meeting places L_{ij} between pairs of users is a geometric random variable with parameter α. Recall that this case is secured by the arrival-network-generating algorithm described in Section 3. In the following theorem (the proof can be found in [39]), we show that for this case the aggregated statistic also has an exponential decay in the tail of the distribution.

Theorem 5 (*Exponentially-Tailed Aggregate Inter-Contact Time*)**.** *If the individual arrival processes are independent Bernoulli point processes with homogeneous rates $\rho_{A_i^l} = \beta$ and the number of shared meeting places L_{ij} between pairs of users is a geometric random variable with parameter α, then the CCDF of the aggregated inter-contact times has an exponential tail, i.e.,*

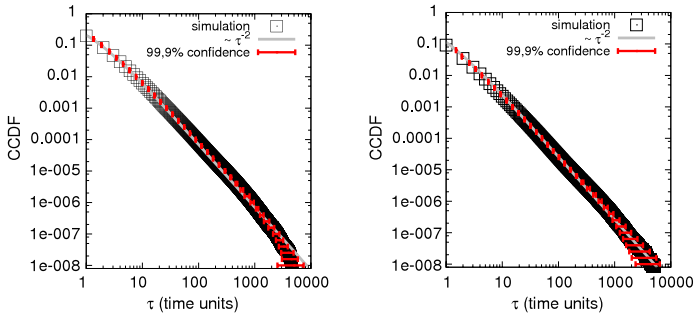
$$F(\tau) \sim e^{-\delta\tau}, \quad \tau \to \infty, \tag{8}$$

where $\delta = -\ln(1 - \beta^2)$.

In this section, we have studied arrival networks in which the links between users and places correspond to Bernoulli processes. We have shown that the pairwise contact sequences in such networks are described by Bernoulli processes, for which the inter-contact times follow a geometric distribution. We have also shown that the rate of the resulting inter-contact times distribution can be derived from the rates of the arrival processes. Thus, the pairwise inter-contact times in such networks, first, follow a geometric distribution, and second, have rate distributions that are controllable by the distribution of arrival rates. As both components, i.e., the individual inter-contact time distribution and the distribution of inter-contact time rates, have been shown [18] to have an impact on the aggregate inter-contact time distribution, we were able to derive different forms of the latter from different distributions of the arrival rates.



(a) $n = 500\, m = 50\, \alpha = 0.5$. (b) $n = 500\, m = 50\, \alpha = 0.2$.



(c) $n = 1000\, m = 30\, \alpha = 0.5$. (d) $n = 1000\, m = 30\, \alpha = 0.2$.

Fig. 11. The aggregate inter-contact time distributions for different arrival networks.

6.3. Validation

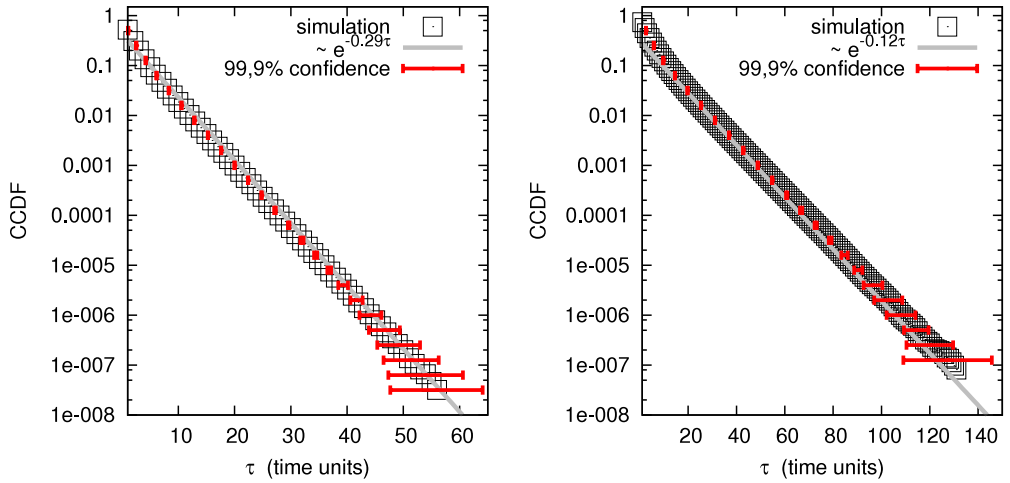
In this section, we support the results obtained above by comparing analytical predictions against simulation results. Please note that this validation is needed since Theorems 3 and 5 provide an approximation for the tail of the distribution of inter-contact times, not an exact analytical prediction.

In order to instantiate the proposed framework, we need to define its input parameters: the social graph G, the removal probability α, and the arrival processes A_i^l for each user i visiting location l. We use the state-of-the-art Barabási–Albert model [43] to generate input social graphs with realistic characteristics (e.g., scale-free degree distribution, short average path length). Thus we consider the two graphs G_{n_1,m_1} and G_{n_2,m_2} of $n_1 = 500$ and $n_2 = 1000$ users and growth parameters $m_1 = 50$ and $m_2 = 30$. The graph-generating algorithm starts with m randomly connected nodes and adds nodes to the network one at a time. Each new node is connected to m existing nodes with a probability that is proportional to the number of links that the existing nodes already have. As a result, heavily linked nodes tend to accumulate even more links, while nodes with only a few links are unlikely to attract a lot of new links. This mechanism of "preferential attachment" has been shown to govern the evolution of realistic social networks [43].

We evaluate both graphs G_{n_1,m_1} and G_{n_2,m_2} when the removal probability used by the algorithm for generating the arrival network is $\alpha_1 = 0.5$ and $\alpha_2 = 0.2$. These settings correspond to an average number of locations shared by a pair of users (which are geometrically distributed) equal to $1/\alpha_1 = 2$ and $1/\alpha_2 = 5$, correspondingly. As a result, we obtain four arrival networks with different structural parameters which we explore in simulations. For each of these arrival networks, we study the resulting inter-contact times obtained by changing the characteristics of the arrival processes A_i^l of users at meeting places. More specifically, we focus on two cases discussed in the previous sections, namely, when the arrival processes are homogeneous and when the arrival rates feature a specific distribution that leads to heavy-tailed aggregate inter-contact times. Simulations are run for 10 000 time units of simulated time, and results are shown with a confidence level of 99.9%.

We assign rates $\rho_{A_i^l}$ of the Bernoulli arrival processes such that $\rho_{A_i^l} = e^{-\frac{1}{2}Y^2}$, where Y is a standard normal random variable. These settings correspond to the case which we mathematically characterized in Section 6.2.1. Fig. 11 depicts the result of simulations for each of the arrival networks. For instance, Fig. 11(a) depicts simulation results for the network with parameters $n = 500$, $m = 50$, and $a = 0.5$. As we can see from the figure, the resulting aggregate inter-contact time CCDF for this network decays as a power law with exponent $\gamma = -2$, i.e., $F(\tau) \sim \tau^{-2}$. In the other arrival networks we observe similar results, which are in agreement with the theoretical predictions from Theorem 3.

In the second experiment, we simulate arrival networks in which the arrival processes are Bernoulli processes, like in the first experiment, but this time with identical rates. These settings correspond to the case which we mathematically characterize in Section 6.2.2. More specifically, we model two networks with the same parameters $\{n = 500, m = 50, a = 0.5\}$, in which all the rates of the arrival processes are identical and equal to $\rho^{(1)}_{A_i^l} = 1/2$ for the first network, and $\rho^{(2)}_{A_i^l} = 1/3$



(a) $\rho = 0.5$. (b) $\rho = 1/3$.

Fig. 12. The aggregate inter-contact time distributions for an arrival network with identical arrival rates.

for the second. Recall that the rate of the arrival process is the reciprocal of the average of the inter-arrival times. Therefore, the first case corresponds to a network in which the average inter-arrival time for all processes is equal to $1/\rho_{A_i^l}^{(1)} = 2$ time units, and the second case to a network with average inter-arrival time for all processes equal to $1/\rho_{A_i^l}^{(2)} = 3$ time units. From Fig. 12, we can see that the resulting distribution of the aggregate inter-contact times decays as an exponential function with exponent $\delta_{(1)} = 0.29$ in the first case and $\delta_{(2)} = 0.12$ in the second. This result is in agreement with the theoretical prediction ($\delta = -\ln(1-\rho^2)$, where ρ is the rate of the arrival process) from Theorem 5.

7. Extending SPoT to generate a spatial output

The main focus of the previous sections was on the ability of SPoT to produce a realistic output in terms of inter-contact times. As previously discussed, inter-contact times are extremely important for the evaluation of an opportunistic network and, for this reason, most network simulators (general simulation platforms [19] or custom simulators [20,21]) are designed to work with contact-based traces as input. Outside the opportunistic networks domain, network simulators [44] often take as input information about node movements instead of (inter-)contact times. In order to make SPoT more general, in this section we discuss how it can be extended to generate a movement-based output. We do not intend to provide an exhaustive analysis of the problem, but just to sketch the main steps for generating a movement-based output. Due to lack of space, we leave the complete evaluation of the properties of this spatial output for future work.

7.1. Generating user trajectories from arrival sequences

In order to obtain a movement-based output, we need to derive trajectories from arrival sequences. To explain the mechanism of transformation, we consider a scenario in which a user U_i visits a set of places $\{M_1, M_2, \ldots, M_l\}$ in a time slot T. The order in which user U_i visits individual locations M_j can be defined through a sequence of arrival times $\{T_1, T_2, \ldots, T_l\}$, where T_j is the time inside time slot T when the user arrives at location M_j. Then, the trajectory of user movements can be reconstructed by connecting places in the order defined by the sequence of arrival times $\{T_j\}$.

Clearly, there are many possible orders in which U_i can visit places $\{M_j\}$ and, therefore, many possible instantiations of the sequence $\{T_j\}$. By design, the SPoT framework assumes that all pairs of users who arrive at time slot T at a place M_j meet with each other. This means that all visitors to M_j should be at M_j during a common time interval. The problem of scheduling the meetings such that everyone attends but also visits other places on their agendas can be transformed into a graph-coloring problem [45]. There are numerous algorithms available in the literature to solve graph-coloring problems efficiently [46–48]. Here, we consider a graph as composed of meeting places and links between those pairs of places that appear in the agenda of at least one user (see Fig. 13). The goal of a graph-coloring (or meeting-scheduling) algorithm is to assign a color to each vertex, i.e., an arrival time T_j at a place M_j, such that the vertices at the ends of each edge are assigned different colors, i.e., meetings do not overlap in time. In this way, meetings at places with the same colors must be scheduled at the same time, whereas meetings at places with different colors (i.e., sharing common visitors) must be scheduled at different times. More schematically, the trajectory-generating algorithm proceeds as described in Table 6.

Please note that the coloring process does not include any notion of "sequence", i.e., taking for example two places with different colors, the coloring algorithm does not tell us anything about whether the first place should be visited before the other one, or vice versa. One possible option is to preserve the same order of visits across all time slots, thus producing

Table 6
The steps of the trajectory-generating algorithm.

1.	The pairs of places that are on the agenda for time slot T (an example is shown on the left in Fig. 13) of at least one common user are connected with links in the graph of places (on the right in Fig. 13).
2.	A graph-coloring algorithm assigns different colors to vertices that share a common link.
3.	Arrival times T_j are assigned to places M_j, such that meeting places with the same color are assigned the same T_j.
4.	Individual trajectories are generated by connecting places on individual user agendas according to the order defined by the sequence $\{T_j\}$.

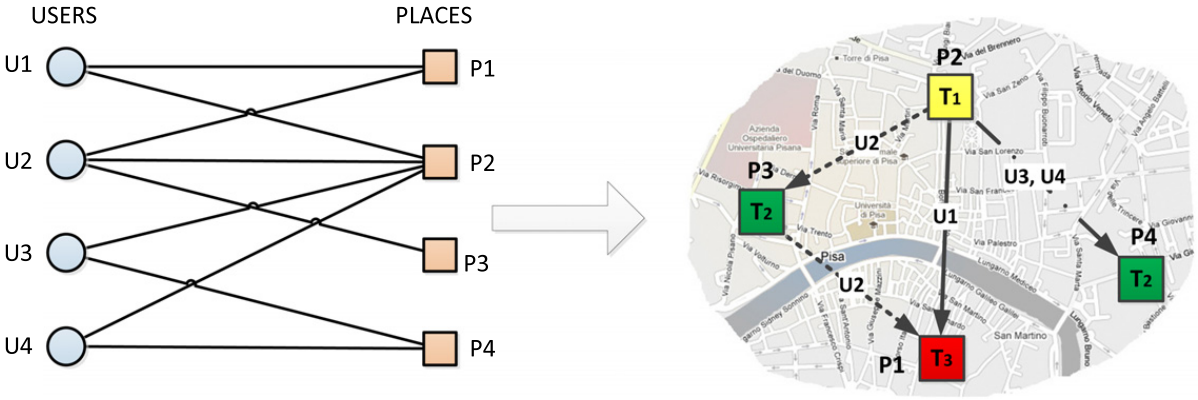


Fig. 13. An example of the transformation from arrival sequences generated by the framework to trajectories of movements. The bipartite graph on the left describes user arrivals at places at a time slot T. The resulting trajectories of users are shown on the right with different arrow styles: $P_2 \rightarrow P_3 \rightarrow P_1$ for user U_2, $P_2 \rightarrow P_1$ for user U_1, and $P_2 \rightarrow P_4$ for users U_3 and U_4.

repeating sequences in the way people visit different locations (a property that has been observed in real traces [49]). This can be simply achieved by tagging all meeting places in the arrival network with numeric IDs, and ordering colors at each time slot T in increasing (or decreasing) order of IDs. Clearly, this is just one of the possible ways for assigning visiting times to meeting places, and we leave to future work a more extensive evaluation of the problem.

7.2. Discussion

The realism of the movements generated by the approach proposed in this section to a large extent depends on the parameters of the arrival network. For instance, the number of places that a user visits per time slot depends on the number of places he/she is connected to and on the arrival rates at those places. The former, in turns, depends on the structure of the initial social graph, whereas the latter depends on the distribution of arrival rates $f_P(\rho)$. In general, a user cannot visit all meeting places of the network (unless he/she is a member of all the cliques identified when running the algorithm for generating the arrival network, an event that is extremely unlikely in realistic scenarios), but only a subset. When selecting the locations to be visited in a time slot, each user takes a subset of the set of meeting places he/she can visit (according to the outcome, for example, of the Bernoulli selection process). The size of this subset depends on the arrival rates defined for users in the arrival network, since, the higher the rates, the greater the number of places selected to be visited in a time slot. We note that the majority of rates in the real traces which we have considered in Section 5 have small values (i.e., less than 0.4 for more than 80% of user–place pairs) and, thus, in general, each user visits a small number of places in each time slot.

The fact that users are bound to visit in a time slot all the selected meeting places introduces a consistency problem: at what speed should the users move to visit all these meeting places, and is this speed realistic? There are two main parameters that can be tuned to guarantee realistic user movements: the duration of the time slot T and the size of the scenario considered. As for the latter, it is clear that in a city-wide area visiting multiple locations does not pose great challenges as these multiple locations can be reached, in the worst case, at bus speed in quite a short time (e.g., an hour or so). For larger scenarios (which are typically not considered for opportunistic networks), obtaining realistic movements can be more challenging, and further investigation is required to address this point. The time slot T can also be helpful. In fact, the larger the time slot, the higher the chances that multiple meeting places can be reached using a realistic speed.

8. Conclusion

In this paper we have proposed SPoT, a mobility framework that incorporates the spatial, social, and temporal dimensions of human mobility. The social and spatial dimensions are added imposing that people belonging to the same social community are assigned to the same location, which is where the people of that community meet. Then, the way users visit their assigned locations over time (corresponding to the temporal aspects of mobility) is described by means of a stochastic process.

In order to provide a realistic instantiation of two building blocks of the framework, namely, the arrival process of users at meeting places and the aggregation of meeting places into larger locations, we have analyzed three datasets containing

traces of human check-ins at real locations, extracted from the online location-based social networks Gowalla, Foursquare, and Altergeo. The analysis of these datasets has revealed that human arrivals at places can be reasonably approximated, for the majority of user–place pairs, by Bernoulli processes. In addition, we have found that meeting places sharing a lot of common users visiting them with high frequency are typically located close to each other (thus, they should be aggregated).

In the third part of the paper we have focused on the flexibility and controllability of the framework. First, we have shown that the SPoT framework can be easily instantiated to accurately reproduce the mobility behavior seen in the Gowalla, Foursquare, and Altergeo traces. Second, as far as the controllability is concerned, we have analytically derived the conditions under which aggregate heavy-tailed and exponentially-tailed inter-contact times emerge, and we have shown that these analytical predictions are totally in agreement with simulation results.

SPoT produces as output a contact-based trace, which can be fed to the vast majority of simulators for opportunistic networks. In the last part of the paper we have discussed how SPoT can be extended to generate a movement-based trace, which can be useful for using SPoT together with simulators such as NS3. In the future, we plan to fully investigate the properties and constraints of this spatial output.

Acknowledgments

We would like to thank Theus Hossmann for generously sharing the Gowalla dataset. We are also grateful to Anastasios Noulas, Chloë Brown, and Cecilia Mascolo for kindly providing us with the Foursquare dataset.

This work was partially funded by the European Commission under the SCAMPI (FP7-FIRE 258414), RECOGNITION (FET-AWARENESS 257756), EINS (FP7-FIRE 288021), MOTO (FP7 317959), and EIT ICT Labs MONC (Business Plan 2013) projects.

References

[1] L. Pelusi, A. Passarella, M. Conti, Opportunistic networking: data forwarding in disconnected mobile ad hoc networks, IEEE Communications Magazine 44 (11) (2006) 134–141.
[2] M. Gonzalez, C. Hidalgo, A. Barabási, Understanding individual human mobility patterns, Nature 453 (7196) (2008) 779–782.
[3] D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel, Nature 439 (7075) (2006) 462–465.
[4] C. Song, T. Koren, P. Wang, A. Barabási, Modelling the scaling properties of human mobility, Nature Physics 6 (10) (2010) 818–823.
[5] 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 (2007) 606–620.
[6] V. Conan, J. Leguay, T. Friedman, Characterizing pairwise inter-contact patterns in delay tolerant networks, in: Proceedings of the Autonomics'07, 2007.
[7] W. Gao, Q. Li, B. Zhao, G. Cao, Multicasting in delay tolerant networks: a social network perspective, in: MobiHoc'09, ACM, 2009, pp. 299–308.
[8] T. Karagiannis, J. Le Boudec, M. Vojnovic, Power law and exponential decay of intercontact times between mobile devices, IEEE Transactions on Mobile Computing 9 (10) (2010) 1377–1390.
[9] W.-J. Hsu, T. Spyropoulos, K. Psounis, A. Helmy, Modeling spatial and temporal dependencies of user mobility in wireless mobile networks, IEEE/ACM Transactions on Networking 17 (5) (2009) 1564–1577.
[10] K. Lee, S. Hong, S. Kim, I. Rhee, S. Chong, Slaw: a new mobility model for human walks, in: IEEE INFOCOM 2009, IEEE, 2009, pp. 855–863.
[11] F. Ekman, A. Keränen, J. Karvo, J. Ott, Working day movement model, in: Proceeding of the 1st ACM SIGMOBILE Workshop on Mobility Models, ACM, 2008, pp. 33–40.
[12] Q. Zheng, X. Hong, J. Liu, D. Cordes, Agenda driven mobility modelling, International Journal of Ad Hoc and Ubiquitous Computing 5 (1) (2010) 22–36.
[13] C. Boldrini, A. Passarella, HCMM: modelling spatial and temporal properties of human mobility driven by users' social relationships, Computer Communications 33 (9) (2010) 1056–1074.
[14] V. Borrel, F. Legendre, M. De Amorim, S. Fdida, Simps: using sociology for personal mobility, IEEE/ACM Transactions on Networking (TON) 17 (3) (2009) 831–842.
[15] Gowalla. URL: http://blog.gowalla.com/.
[16] Foursquare. URL: https://foursquare.com/about.
[17] Altergeo. URL: http://www.crunchbase.com/company/altergeo.
[18] A. Passarella, M. Conti, Analysis of individual pair and aggregate inter-contact times in heterogeneous opportunistic networks, IEEE Transactions on Mobile Computing (2012) http://dx.doi.org/10.1109/TMC.2012.213 [Available online].
[19] A. Keränen, J. Ott, T. Kärkkäinen, The ONE simulator for DTN protocol evaluation, in: SIMUTools'09, 2009.
[20] P. Hui, J. Crowcroft, E. Yoneki, Bubble rap: social-based forwarding in delay-tolerant networks, IEEE Transactions on Mobile Computing 10 (11) (2011) 1576–1589.
[21] C. Boldrini, M. Conti, A. Passarella, Exploiting users social relations to forward data in opportunistic networks: the HiBOp solution, Pervasive and Mobile Computing 4 (5) (2008) 633–657.
[22] D. Karamshuk, C. Boldrini, M. Conti, A. Passarella, Human mobility models for opportunistic networks, IEEE Communications Magazine 49 (12) (2011) 157–165.
[23] M. Musolesi, C. Mascolo, Designing mobility models based on social network theory, ACM SIGMOBILE Mobile Computing and Communications Review 11 (3) (2007) 59–70.
[24] T. Hossmann, T. Spyropoulos, F. Legendre, Putting contacts into context: mobility modeling beyond inter-contact times, in: Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc 11, ACM, Paris, France, 2011.
[25] D. Karamshuk, C. Boldrini, M. Conti, A. Passarella, An arrival-based framework for human mobility modeling, in: 2012 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMoM, IEEE, 2012, pp. 1–9.
[26] M. Newman, The structure and function of complex networks, SIAM Review (2003) 167–256.
[27] M. Zukerman, An introduction to queueing theory and stochastic teletraffic models, 2007.
[28] J. Silvis, D. Niemeier, R. D'Souza, Social networks and travel behavior: report from an integrated travel diary, in: 11th International Conference on Travel Behaviour Reserach, Kyoto, 2006.
[29] E. Cho, S. Myers, J. Leskovec, Friendship and mobility: user movement in location-based social networks, in: SIGKDD'11, ACM, 2011, pp. 1082–1090.
[30] M. Newman, M. Girvan, Finding and evaluating community structure in networks, Physical Review E 69 (2) (2004) 026113.
[31] G. Palla, I. Derényi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature 435 (7043) (2005) 814–818.
[32] C. Bron, J. Kerbosch, Algorithm 457: finding all cliques of an undirected graph, Communications of the ACM 16 (9) (1973) 575–577.

[33] T.M. Fruchterman, E.M. Reingold, Graph drawing by force-directed placement, Software—Practice and Experience 21 (11) (1991) 1129–1164.
[34] M. Jacomy, S. Heymann, T. Venturini, M. Bastian, Forceatlas2, A Graph Layout Algorithm for Handy Network Visualization, Tech. Rep., Gephi Consortium, 2011.
[35] T. Hossmann, G. Nomikos, T. Spyropoulos, F. Legendre, Collection and analysis of multi-dimensional network data for opportunistic networking research, Computer Communications (2013).
[36] S. Catanese, P. De Meo, E. Ferrara, G. Fiumara, Analyzing the Facebook friendship graph.
[37] C. Brown, V. Nicosia, S. Scellato, A. Noulas, C. Mascolo, Where online friends meet: Social communities in location-based networks, in: Proc. Sixth International AAAI Conference on Weblogs and Social Media, ICWSM 2012, Dublin, Ireland, 2012.
[38] Gvidi. URL: http://gvidi.ru.
[39] D. Karamshuk, C. Boldrini, M. Conti, A. Passarella, Spot: Representing The Social, Spatial, and Temporal Dimensions of Human Mobility with a Unifying Framework, Tech. Rep., IIT-CNR, 2013 (available online at: http://cnd.iit.cnr.it/chiara/techrep/spotPMC13TR.pdf).
[40] E. Lehmann, J. Romano, Testing Statistical Hypotheses, Springer, 2005.
[41] P.-U. Tournoux, J. Leguay, F. Benbadis, J. Whitbeck, V. Conan, M.D. de Amorim, Density-aware routing in highly dynamic DTNs: the rollernet case, IEEE Transactions on Mobile Computing 10 (12) (2011) 1755–1768.
[42] A. Passarella, R.I. Dunbar, M. Conti, F. Pezzoni, Ego network models for future Internet social networking environments, Computer Communications 35 (18) (2012) 2201–2217.
[43] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Reviews of Modern Physics 74 (1) (2002) 47.
[44] Ns3. URL: http://www.nsnam.org/.
[45] T.R. Jensen, B. Toft, Graph Coloring Problems, Vol. 39, Wiley-Interscience, 1994.
[46] H. Kierstead, A simple competitive graph coloring algorithm, Journal of Combinatorial Theory. Series B 78 (1) (2000) 57–68.
[47] D. Brélaz, New methods to color the vertices of a graph, Communications of the ACM 22 (4) (1979) 251–256.
[48] P. Galinier, J.-K. Hao, Hybrid evolutionary algorithms for graph coloring, Journal of Combinatorial Optimization 3 (4) (1999) 379–397.
[49] C. Song, Z. Qu, N. Blumm, A.-L. Barabási, Limits of predictability in human mobility, Science 327 (5968) (2010) 1018–1021.

Duty cycling in opportunistic networks: intercontact times and energy-delay tradeoff

Elisabetta Biondi, Chiara Boldrini, Andrea Passarella, and Marco Conti
IIT-CNR, Via G. Moruzzi 1, 56124 Pisa, Italy
Email: first.last@iit.cnr.it

***Abstract*—Portable mobile devices like smartphones and tablets are the enablers for communications in mobile ad hoc networks. In order to optimise their energy usage, one of the most popular techniques is to implement a duty cycling policy, which periodically puts the user device in a energy saving mode (e.g., Bluetooth *inquiry scan* phase or turning off the WiFi interface) for a certain amount of time. Clearly, this strategy increases the battery lifetime, but it also has the net effect of reducing the number of usable contacts for delivering messages, increasing intercontact times and delays. In order to understand the effect of duty cycling in opportunistic networks, in this paper we propose a general model for deriving the pairwise intercontact times modified by a duty cycling policy. Then, we specialise this model when the original intercontact times are exponential (an assumption popular in the literature), and we show that, in this case, the intercontact times measured after duty cycling are, approximately, again exponential, but with a rate proportional to the inverse of the duty cycle. Once we have the distribution of the intercontact times after duty cycling, we use it for analysing how duty cycling affects the delay of message forwarding and the network lifetime.**

I. Introduction

The widespread availability of smart, handheld devices like smartphones and tablet has stimulated the discussion and research about the possibility of extending the communication opportunities between users. Particularly appealing, towards this direction, is the opportunistic networking paradigm, in which messages arrive to their final destination through consecutive pairwise exchanges between users that are in radio contact with each other. Thus, user mobility, and especially user encounters, are the key enablers of opportunistic communications. This networking paradigm reverses the approach of traditional Mobile Ad Hoc Networks: where user mobility was previously seen as an accident, it is now one of the enablers of message circulation.

One aspect that is hindering opportunistic communications from being widely implemented is the lack of technologies allowing for transparent and efficient ad hoc communications in off-the-shelf smartphones. Severely critical are particularly the energy requirements of such protocols. In fact, it is a well-known problem [1] that WiFi in ad hoc mode is extremely energy hungry, even if idle, and also Bluetooth, generally considered energy-efficient, suffers from high energy consumption during the *inquiry* phase. Clearly, no user will be willing to participate to an opportunistic network if they risk to see their battery drained in a few hours. Due to the energy issues of all the ad hoc communications technologies available, there have been some attempts in the literature to improve the energy efficiency of opportunistic communications. To the best of our knowledge all power saving schemes designed for opportunistic networks focus on contact probing (or, equivalently, neighbour discovery) mechanisms, for two main reasons. First, as discussed in Section II, the neighbour discovery phase is energy hungry for both WiFi and Bluetooth. The second reason is that, in opportunistic networks, nodes are typically not in contact with anyone for a large fraction of time. This implies that, while detecting contacts is crucial and has to be done for the whole network lifetime, continuously scanning without interruptions can be not only energy inefficient but also a waste of time, since often no contact would be detected. Thus, the contact probing phase lends itself to many improvements from the perspective of designing a strategy that misses just a few contacts but provides a high energy gain.

While research on other aspects of opportunistic communications (e.g., routing, data dissemination) has been thriving in the recent years, energy issues in opportunistic networks have still to be fully addressed. Even worse, while there are some works discussing new smart power saving strategies for opportunistic networks, there are very few contributions that study the effect of power saving mechanisms on intercontact times, which are the prominent metric of user contact dynamics and therefore one of the key elements determining forwarding performance. The intercontact time is defined as the time interval between two consecutive encounters between the same pair of nodes. Intercontact times are considered to be the main bottleneck in opportunistic communications, as they are typically one or two orders of magnitude greater than contact times [2]. Being messages exchanged between nodes when they meet, it is clear that the main contribution to message delay is determined by intercontact times. For this reason, we believe it is of paramount importance to understand how power saving techniques, which may effectively reduce the number of usable contacts, affect the intercontact time.

In light of the above discussion, the contribution of this paper is threefold. First, we derive an analytical model of the actual inter-contact times between nodes *after* duty cycling is factored in, i.e. by taking into account that some contacts may be missed because at least one of the devices may be in a low-energy state that does not allow it to detect the contact. In the following, we refer to this figure as *detected* intercontact times. While deriving an exact characterisation of the detected intercontact times is in general too complex from an analytical standpoint, we are able to derive their first two moments. As it is well-known, this is sufficient to approximate the distribution of the detected intercontact times using hyper- or hypo- exponential distributions, using standard techniques [3]. Thus, using this model, we are able to obtain

an approximated representation of intercontact times measured when a duty cycling policy is in place under virtually any distribution.

The second contribution of the paper is the solution of the above model for the case of exponential intercontact times, which is a popular assumption in the related literature [4] [5] (even if a general consensus on which is the best distribution for representing realistic intercontact times has yet to be achieved). With exponential intercontact times, the proposed model can be solved approximately in closed form and, under a specific condition that we derive, the detected intercontact times are still exponential, but with a rate proportional to the inverse of the duty cycle. In addition, we show that the condition under which our approximation holds is satisfied by the most popular traces available in the literature. This result tells us that models (e.g., of the delay) that assume exponential intercontact times (that are typically tractable and thus very popular in the literature) can still be used when a duty cycling policy is in place, as long as the original rates are scaled proportionally to the inverse of the duty cycle. So far, this aspect (i.e., the fact that duty cycling can affect the detected pairwise contacts) has been largely ignored in the literature. In fact, while periodic probing is implemented in almost all contact datasets available in the literature [6] [7] [8], both when studying the statistical properties of intercontact times in the datasets and when using the rates estimated from these traces in order to validate analytical models, the effect of the periodic rather than continuous duty cycling has been ignored.

The third and final contribution of this paper lies in studying how the detected intercontact times affect the delay experienced by messages and network lifetime. As expected, we find that the delay (both first and second moments) increases, but no additional variability is introduced (i.e., the coefficient of variation remains unchanged). Also the network lifetime (i.e., the time until nodes run out of battery, given a certain energy budget) evidently increases. Being the network lifetime longer, a larger volume of traffic is handled by the network when a duty cycling policy is in place. Finally, we focus on the problem of finding an optimal duty cycle value that would allows us to minimise the loss from the delay standpoint and maximise the gain in terms of traffic handled by the network. We show that, when intercontact times are exponential, it is not possible to achieve both goals at the same time and the operator has to decide whether delay or volume of traffic (or equivalently, network lifetime) is of concern.

II. Related work

Ad hoc communications in opportunistic networks typically go through either the WiFi or Bluetooth interface. While for 802.11 cards used in infrastructure mode energy consumption in the idle state (previously comparable to that used in the transmitting/receiving state) has been drastically reduced by the introduction of the Power Saving Mode (PSM) [1], ad hoc 802.11 is still highly inefficient [1] and PSM for ad hoc is typically not implemented in smartphones' 802.11 interfaces. Thus, in the ad hoc case, it is still true that the energy for being idle is comparable to the energy consumed for sending and receiving messages, hence turning off the network cards when possible is the simplest yet the only viable approach for increasing battery lifetime. Measurements of Bluetooth energy consumption are more consistent across past and recent studies. Of all Bluetooth phases, scanning is found to be the most energy hungry across different studies [1], [9]. For this reason, it is often advised to reduce the scanning frequency in order to save battery. The Bluetooth discoverable state is generally considered energy cheap (as stated in [10], [1]) but keeping the phone in the discoverable state indefinitely in the long run incurs in a significant energy consumption as well. In a recent work, Trifunovic et al. [10] directly compare Bluetooth and two innovative WiFi-based ad hoc communications modes, namely, WiFi Direct [11] and WLAN-Opp [12]. As far as neighbor discovery is concerned, Bluetooth is found two times more energy efficient than WLAN-Opp and about five times more energy efficient than WiFi Direct. As expected, the energy consumption for all three protocols heavily depends on the scanning interval, i.e., how often they scan their neighbourhood for new devices. For scanning intervals smaller than $\sim 100s$, the power consumption increases drastically. Looking at energy consumption at a finer detail, we see that, as discussed before, in Bluetooth discovery, the higher amount of energy is consumed during the scanning phase. Instead, for WLAN-Opp both being discoverable (i.e., when the device acts as AP) and discovering are expensive operations. In WiFi Direct the situation is even worse, since for detecting each others both devices have to be scanning at the same time.

A taxonomy proposed by Jun et al. [13] for traditional MANETs classifies duty cycling policies into four categories. With the *synchronous* mechanisms (the most popular of which is 802.11 Power Saving Mode) nodes all wake up/go to sleep at the same time. The drawback of this approach is that it requires a global synchronisation among nodes and a fully connected network[13]. In *asynchronous* strategies, neighbouring nodes select their wake-up slot so that they overlap with one another. In *cell-base* schemes the network is divided into non-overlapping cells, within which only a few nodes are awake at a given time. In *on demand* strategies nodes are equipped with an additional low-power radio interface that is used for signalling wake up or sleep commands. Unfortunately, all the above strategies were designed for a dense and connected network, which is quite different from the typical conditions in which opportunistic networks operate.

As discussed in Section I, all papers dealing with power saving issues in opportunistic networks have focused on the contact probing phase. There are quite a few papers in the literature aiming at designing smart contact probing strategies for DTN/opportunistic networks. Contact probing schemes can be classified into fixed, when the ON/OFF duration of the duty cycle is established at the beginning and never changed [13], or adaptive, when the frequency of probing is increased or decreased according to some policy [14]. Both fixed and adaptive strategies can be context-oblivious [13], if they do not exploit information on user past behaviour or position, or context-aware [15], [16] otherwise. In order to address the lack of duty cycling mechanisms in sparse networks, [13] proposes three schemes based on different availability of information at nodes. In the oracle case, nodes can predict their meetings, so they are able to wake up and get to sleep exactly at the beginning and end of a contact. This strategy is clearly used for benchmarking only, since real nodes are not omniscient. At the opposite range, the authors identify the zero knowledge

strategy, in which nodes have no means to estimate their future contacts, so basically enter a periodic, fixed, beaming period in which neighbour discover is performed. Between these two extremes, the authors propose the partial knowledge strategy, in which they assume that nodes are able to estimate the mean and variance of the intercontact times and the contact times, information that is used in order to predict when to scan more intensely for neighbours (intuitively, not right after a contact has ended).

The work by Choi et al. [14] considers a DTN in which nodes wake up at synchronised intervals and sends beacons in order to discover new neighbours. If no neighbour is discovered, the next beaconing period is set at double the previous interval and nodes sleep until then. This exponential increasing is performed until a maximum value is reached. The main contribution of the paper lies in designing an algorithm (AEB) in order to compute this maximum value in order minimise power consumption and maximise the probability of detecting a contact. This is done exploiting the CDF of contact duration. Unfortunately, this algorithm is targeted at random mobility models (in which contact and intercontact times are i.i.d.), hence it is arguable whether it is applicable to real networks.

An energy-efficient variation of Bluetooth discovery mode is proposed in [15]. This work starts from the observation that wherever a neighbour is encountered, other contacts are likely to happen as well. So the authors propose two schemes: one in which the discovery mode is triggered by indications of recent activities and one in which previous contacts at a specific location (access to a positioning system is assumed) are used to switch the device to a more aggressive scanning mode.

Wang et al. [16], assuming that nodes scan their surroundings every T and that contact duration for a pair of nodes is stationary and i.i.d. are able to prove that among all social-oblivious strategies with the same average contact probing interval, the strategy exploiting fixed probing intervals is optimal for minimising the probability of missing a contact. This is an interesting result, supporting the use of simple fixed contact probing strategies over more elaborate ones. Then, the authors propose some heuristic social-aware adaptive probing algorithm, of which the STAR algorithm, which adapts to the contact arrival process exploiting a self-similar argument, is shown to outperform the others.

The work in [17] is a theoretical work that extends the fluid model of two-hop forwarding with the fact that nodes, with a certain rate, go from the inactive state to the active state (in which they start beaconing until they receive a message copy) and use this model to solve the optimal activation problem. The model is tailored to a homogeneous environment (i.e., nodes with contact process i.i.d.) and, as discussed by the authors themselves, cannot be used as it is to describe the heterogeneous environments seen in real mobility traces. Moreover, the model is not directly applicable to duty-cycling-style power saving, since node activation is considered monotone, i.e., once activated nodes remain active until their battery is depleted or the deadline is reached.

All the above contributions aimed at deriving a power saving strategy for DTNs, be it fixed [13] or adaptive [14], social-oblivious [13] or social-aware [15], [16]. This is not our goal. Instead, in this paper, we assume that a fixed contact probing scheme is given (whether social-oblivious or social-aware is not important, as long as its parameters are kept constant). Our goal is to investigate the effects of such duty cycle in neighbour discovery as far as intercontact times are concerned.

In the literature, the works closest to ours are [18] [19]. Zhou et al. [18] focus on the RWP mobility model and show how the number of contacts detected is affected when neighbour discovery is performed every T seconds and what is the effect on the energy consumption (defined as $1/T$). Our contribution is more general, as it is not bound to the RWP model but it can be applied to any well known distribution for intercontact times, if numerical solutions are sufficient, or it can be solved in closed form for the exponential case. In addition, we also provide an analysis of how the delay is affected by the duty cycling. Moreover, despite its simplicity, our duty cycling function allows for more flexibility than the simple scanning every T seconds, as discussed in the next section.

Qin et al. [19] perform a study that is exactly orthogonal to this work. In fact, they evaluate how link duration (or contact duration, in our terminology) is affected by the contact probing interval. Assuming a fixed, social-oblivious probing scheme similar to the one that we consider, Qin et al. study the effective duration of a contact given that the contact is not discovered immediately when it starts due to the contact probing process in place. They provide a formula for the PDF of the link duration under generic contact duration and study the trade-off between energy (in terms of probing frequency) and throughput. The focus of our work is exactly the opposite. We investigate the effect of contact probing (which can be easily translated into a duty cycling problem) on the intercontact time rather than on the contact time. The motivation for this choice is that intercontact times are typically much larger than contact times in real human mobility, thus the delay in opportunistic networks is mainly determined by the intercontact time. Intercontact times are larger when contacts are not detected immediately or missed and thus it is important to understand how they increase and how this affects the delay experienced by messages.

III. Problem statement

We use duty cycling in a general sense here, meaning any power saving mechanism that hinders the possibility of a continuous scan of the devices in the neighbourhood. So, we assume that nodes can be either in the ON or OFF state. In the ON state, nodes are able to detect contacts with other devices. The semantic of the OFF state depends on the scenario that we want to represent and we will discuss it later in this section. We hereafter focus on a duty cycling technique in which a generic node A is ON for τ_A seconds and OFF for the other $T_A - \tau_A$ seconds, where T_A is the period of the duty cycling scheme. This duty cycling model closely approximates typical duty cycles used in the literature (neglecting the small random back-off time introduced in practice after nodes wake up in order to avoid contention, see e.g., [6]). For example, the proposed duty cycling model is able to model approximately the scanning processes of Bluetooth and WLAN-Opp discussed in [10].

Assuming that the network wakes up at $t = 0$, the first ON interval can be placed anywhere between $[0, T_A)$. We denote with $s_0^{(A)}$ and $s_1^{(A)}$ the time instants at which the ON interval starts and ends, respectively, for which $s_1^{(A)} - s_0^{(A)} = \tau_A$ holds. The duty cycling policy for node A is thus represented by the following function:

$$d_A(t) = \begin{cases} 1 & \text{if } t \bmod T_A \in [s_0^{(A)}, s_1^{(A)}] \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where $\bmod$ denotes the modulo operation.

Starting from this duty cycling policy on single nodes, the semantic of the discovery process depends on the specific scenario under study. For example, in the case Bluetooth, assuming that nodes alternate between the *inquiry* and *inquiry scan* phase (corresponding in our model, respectively, to the ON and OFF state) as described in [20], and that ON intervals do not overlap[1], we have that each node is able to detect a contact with nodes in radio range when it is in the ON state. Vice versa, for all scenarios in which for contacts to be detected and used for communications both nodes A and B must be ON at the same time (e.g., as in WiFi Direct), we have to consider the joint duty cycling function[2] d, defined as $d_A * d_B$ (we assume that all nodes have the same period T). The first ON interval of d starts at $s_0 = \max\{s_0^{(A)}, s_0^{(B)}\}$ and ends at $s_1 = \min\{s_1^{(A)}, s_1^{(B)}\}$. We denote the fraction of time nodes spend in the ON period as Δ ($= \frac{\tau}{T}$, where $\tau = s_1 - s_0$ is the length of the ON interval under d and T the period of d).

$$d(t) = \begin{cases} 1 & \text{if } t \bmod T \in [s_0, s_1] \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

If the ON intervals of nodes A and B never overlap, the two nodes will not be able to communicate, because they never see each other. In order to avoid this trivial case, we assume that duty cycles preserve the ability of nodes to communicate with each other. In the best case, nodes are synchronised, so they are able to fully exploit the length $\min\{\tau_A, \tau_B\}$ of their ON intervals. In the following we will use notation d to generically denote the duty cycling function, be it the single function d_A when only node A needs to be in the ON state for detection or the joint function d.

Function d determines the way contacts are discovered. In the following, we focus on the intercontact process between a generic pair of nodes, and, to make the analysis more tractable, we assume that a contact event is detected only if it starts during an ON period. This assumption is reasonable when the duration of a contact is smaller than the duration of the OFF interval. In fact, in this case the probability of the contact lasting until the next ON interval is negligible. While it is difficult to verify whether this assumption holds for the datasets in which duty cycling is implemented (because the measured contact durations already factor in the duty cycle effect), the work in [2], which relies on special devices with extremely short scanning intervals of $1s$, states that for more than 50% of contact duration samples the duration is smaller than $48s$. Trifunovic et al. [10] have derived that, with Bluetooth and WiFi, scanning intervals greater that $100s$ perform significantly better energy-wise. Actually, typical duty cycling policies used to collect traces have a period of several minutes (e.g., 5 minutes in the Reality Mining dataset [8]), with ON periods in the order of few seconds. Therefore, we can consider this assumption as reasonable.

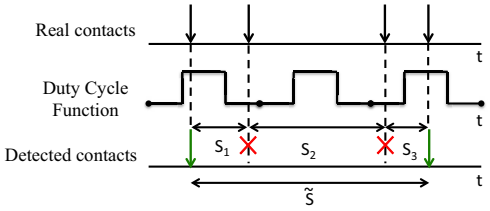


Fig. 1. Contact process with duty cycling

We now focus on the contact process. We assume that the time between two consecutive contacts between the same pair of nodes can be modelled as a continuous random variable S, and that intercontact times between a given pair of nodes are independent and identically distributed (while they can follow different distributions for different pairs). Hence, the contact process can be modelled as a renewal process, where $S_i \sim S$ denotes the time between the i-th and the $(i+1)$ contact event. Similarly, we denote with $\tilde{S}$ the random variable representing the detected intercontact times, and with $\tilde{S}_i \sim \tilde{S}$ the time between the i-th and the $(i+1)$ detected contact event (Figure 1). In the following, without loss of generality, we assume that there is a contact event at t_0 during the first ON period after $t = 0$. Consider the case in which $i-1$ contacts are missed after the one happening at t_0 and the i-th is detected. If we neglect contact duration (recall that this is reasonable, since it is typically one-two orders of magnitude smaller than the intercontact time [6] [2]), it is clear that the time between the two detected events is given by the sum of the interarrival times of the events up to the i-th. Thus, for $\tilde{S}$, the following definition holds.

Definition 1: The detected intercontact time $\tilde{S}$ can be obtained as $\tilde{S} = \sum_{i=1}^{N} S_i$, where N is the random variable describing the number of contacts needed to get one detected.

Therefore, $\tilde{S}$ is a random sum of i.i.d. variables. This sum has some nice properties, which we will exploit in Section V in order to derive the first two moments of $\tilde{S}$. Please also note that Definition 1 is general, i.e., holds for any type of continuous intercontact time distribution and for any type of duty cycling policy. In the next sections we show how the duty cycling model and the contact process can be studied together in order to uncover the features of the detected intercontact times. For the convenience of the reader, the notation used throughout the paper is summarized in Table I.

IV. Deriving the distribution of N

In this section we study the probability distribution of N, defined as the number of contacts needed in order to detect the first one. We aim, in particular, at deriving its first two moments, which are essential, as we will see later in the paper, for the computation of $\tilde{S}$. The rationale behind its derivation is pretty intuitive. In fact, $N = 1$ corresponds to the case of a detection right after the last detected contact. So, if the last

[1]Please note that with Bluetooth two nodes inquiring at the same time do not detect each other. In a real implementation, overlapping inquiry intervals are avoided by means of the random backoff time.

[2]When unambiguous, we drop the subscript indicating the specific node pair considered, as in this case.

TABLE I. NOTATION

f_X, F_X	PDF and CDF of random variable X
S	intercontact time
$\hat{S}$	detected intercontact time
N	number of ICTs taking places before the next contact is detected
T, τ	period of duty cycle and duration of the ON period
Δ	fraction of time nodes are in ON
t_0	time of the last detected contact
d	duty cycle function
s_0, s_1	start and end time of the first ON period ($s_1 - s_0 = \tau$)
λ_{ij}	rate of intercontact times between nodes i and j
D, D_Δ	delay without duty cycling and delay with duty cycling Δ

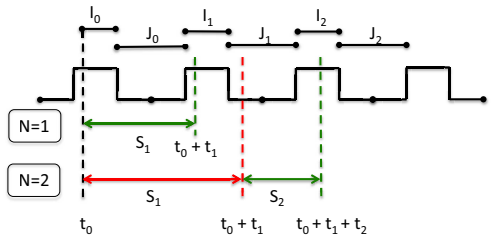


Fig. 2. Example for case $N = 1$ and $N = 2$ (red for missed contact, green for detected contact)

detected contact took place at t_0, with $N = 1$ the next contact can happen in any of the ON intervals after t_0. For case $N = 2$ to happen, the first contact after the one detected at t_0 has to fall in any of the OFF intervals after t_0 ad the second contact in any of the ON intervals after the OFF interval in which the first contact has been missed. The derivation of the PDF of N quantifies the probabilities of each of these events.

Before moving into the detail of the derivation, we first introduce the notation that we use in this section. Let us assume that the last detected contact (we refer to it as the *zero contact*) took place at time t_0. Starting from t_0, we denote the i-th ON interval after t_0 as I_i and the i-th OFF interval after t_0 as J_i. All ON intervals but the first one are of type $[s_0+iT, s_1+iT]$, while all the OFF intervals are defined as $[s_1 + iT, s_0 + (i + 1)T]$. The first ON interval is special, because it is constrained to start after t_0, thus we define it as $(t_0, s1]$. Figure 2 describes this scenario. Then, when N=1, I_{n_1} denotes the ON interval in which the first contact is detected, when N=2 I_{n_2} denotes the ON interval in which the second contact is detected, and so on. We do the same for OFF intervals, thus J_{n_j} denotes the OFF interval in which the j-th contact is missed. We will often use ON and OFF intervals shifted by a time t, which we denote as $I_i - t$ and $J_i - t$. Moreover, we assume that the time t_0 at which the zero contact happens within interval $[s_0, s_1]$ is described by random variable $\hat{S}_0$. Finally, we denote with f_{S_i} the PDF of S_i.

Using the above notation, the PDF of N can be computed as described in Theorem 1 below (the proof can be found in Appendix A). Theorem 1 provides an accurate approximation of the PDF of N when the probability of two undetected contacts falling in the same OFF interval is very low. In Section IV-A we derive the conditions under which this assumption is reasonable and we show that these conditions are satisfied by the most popular traces of human contacts.

Theorem 1 (Distribution of N): The discrete probability density of N can be approximated by the following:

$$P\{N=1\} = \sum_{n_1=0}^{\infty} \int_{s_0}^{s_1} f_{\hat{S}_0}(t_0) \int_{I_{n_1}-t_0} f_{S_1}(t_1) dt_1 dt_0 \quad (3)$$

$$P\{N=2\} = \sum_{n_1=0}^{\infty} \sum_{n_2=n_1+1}^{\infty} \int_{s_0}^{s_1} f_{\hat{S}_0}(t_0) \int_{J_{n_1}-t_0} f_{S_1}(t_1) \cdot \cdot \int_{I_{n_2}-t_0-t_1} f_{S_2}(t_2) dt_0 dt_1 dt_2 \quad (4)$$

$$\vdots$$

$$P\{N=k\} = \sum_{n_1=0}^{\infty} \cdots \sum_{n_k=n_{k-1}+1}^{\infty} \int_{s_0}^{s_1} f_{\hat{S}_0}(t_0) \int_{t_1 \in J_{n_1}-t_0} f_{S_1}(t_1) \cdot \cdots \int_{t_{k-1} \in J_{n_{k-1}}-t_0-\cdots-t_{k-2}} f_{S_{k-1}}(t_{k-1}) \cdot \int_{t_k \in I_{n_k}-t_0-\cdots-t_{k-1}} f_{S_k}(t_k) dt_k dt_{k-1} \ldots dt_1 dt_0. \quad (5)$$

Finding a closed form for the distribution of N in Theorem 1 might be prohibitive in the general case, and we have been able to obtain only numerical solutions. However, when intercontact times are exponential, a closed form solution is available (Corollary 1 below, proof available in Appendix A) and the first two moments of N can be computed (Appendix A).

Corollary 1 (N with exponential intercontact times): When real intercontact times S_i are exponential with rate[3] λ, the probability density of N is given by:

$$\begin{cases} P\{N=1\} = 1 + \frac{e^{-\lambda\tau}-1}{\lambda\tau} + \frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{\lambda\tau(e^{\lambda T}-1)} \\ P\{N=k\} = e^{\lambda\tau}\frac{(1-e^{-\lambda\tau})^2}{\lambda\tau(1-e^{-\lambda T})}\left[\frac{\lambda(T-\tau)}{e^{\lambda T}-1}\right]^{k-1}, \quad k \geq 2 \end{cases} \quad (6)$$

A. Quantifying the error

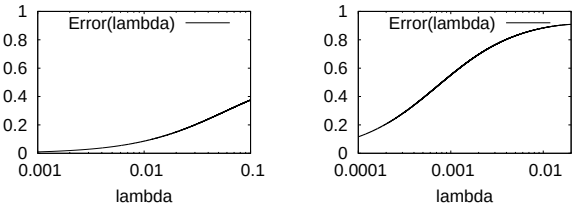
In Theorem 1 we have provided a general formulation for the distribution of N which holds when the probability of two consecutive contacts happening during the same OFF period is small. We now discuss under which settings the approximation introduced is reasonable, taking as reference the exponential case. Exploiting the result in Corollary 1, it is possibile to compute the error function, as follows.

Definition 2: The error $\mathcal{E}$ (defined in $[0, 1]$) introduced by the approximation in Corollary 1 can be expressed as:

$$\mathcal{E}(\tau, T, \lambda) = \frac{1-e^{-\lambda\tau}}{\lambda\tau} - \frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{\lambda\tau(e^{\lambda T}-1)} + - \frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{1-e^{-\lambda T}} \frac{T-\tau}{\tau(e^{\lambda T}-1-\lambda T+\lambda\tau)}. \quad (7)$$

The above expression can be simply obtained as the difference between 1 and $\sum_k P\{N = k\}$, after noting that in our approximation we are neglecting events that belong to the sample space, hence the total probability that we obtain is

[3]Again, for ease of notation, we omit subscript A, B for λ, since it is unambiguous that we are referring to the tagged node pair A, B. Please note, however, that the network model we are referring to is still heterogenous.



(a) $\tau = 5$ and $T = 15$ (b) $\tau = 5$ and $T = 120$

Fig. 3. $\mathcal{E}$ varying λ (logarithmic scale for the x-axis)

less than 1. The distance between $\sum_k P\{N = k\}$ and 1 thus gives us the quality of our approximation, in which the closer $\mathcal{E}$ to 0 the better.

Intuitively, from a physical standpoint, our approximation holds when the average intercontact time is significantly larger than the duration of the OFF interval. Under this condition, the OFF interval is, on average, too short to accommodate two intercontact times. Thus, the condition when we expect that our approximation is very precise is $E[S] \gg T - \tau$, or, equivalently, $\frac{1}{\lambda} \gg T - \tau$. The worst case for this condition happens when τ is very small. In fact, when $\tau \to 0$ the OFF interval can be as big as possible (i.e., occupying almost all T) and, consequently, the probability that two undetected contacts fall into it maximum. Thus, if we ensure that $\frac{1}{\lambda} \gg T$ our condition will hold. We have proved mathematically this result in Appendix A, which is summarised in the following lemma.

Lemma 1: When $\lambda T \ll 1$, the error $\mathcal{E}$ introduced by the approximation of Corollary 1 approaches zero.

Let us now explore the parameter space τ, T, λ in order to characterise how $\mathcal{E}$ goes to zero when the condition introduced above is satisfied. At first we set $\tau = 5$ and $T = 15$ (as in the RollerNet experiment – see Section V-A) and we plot $\mathcal{E}$ varying λ (Figure 3(a)). As $\frac{1}{\lambda}$ represents the mean intercontact time $E[S]$, the smaller λ the bigger the mean intercontact time. And in fact, the error goes to zero as $\lambda \ll \frac{1}{T-\tau} = \frac{1}{10}$ and our condition is confirmed. We now keep the same τ value and increase T, setting it to $120s$, which is the value used for the Infocom experiment (Section V-A). We expect from condition (i) that the error increases for small λ values with respect to the previous case, and this is confirmed in Figure 3(b).

V. The duty cycling effect on exponential intercontact times

Exploiting the results derived in the previous section, here we discuss how to compute the first and second moment of the detected intercontact time $\tilde{S}$ for a generic node pair A, B. The relation between S and $\tilde{S}$ is stated by Definition 1, i.e., $\tilde{S} = \sum_{i=1}^{N} S_i$. Thus, $\tilde{S}$ is a random sum of random variables, and we can exploit well-known properties to compute compute its first and second moment (in Appendix A we specialise this result for the exponential case).

Proposition 1: The first and second moment of $\tilde{S}$ are given by $E[\tilde{S}] = E[N]E[S]$ and $E[\tilde{S}^2] = E[N^2]E[X]^2 + E[N]E[X^2] - E[N]E[X]^2$.

While the above formula holds in general, in the case of exponential intercontact times it is possible to derive an even stronger result, described in Theorem 2 below. This result is the key derivation of this work, and it tells us that, under condition $\lambda T \ll 1$, exponential intercontact times are modified by duty cycling only in terms of the parameter of their distribution but they still remain exponential.

Theorem 2: When $\lambda T \ll 1$, the detected intercontact times $\tilde{S}$ follow approximately an exponential distribution with rate $\lambda\Delta$.

Proof: We can calculate the moment generating function (MGF) of $\tilde{S}$ using the expression described at the beginning of the section, i.e., $M_{\tilde{S}}(s) = M_N(M_S(s))$. First of all, we have to calculate the MGF of N, that we can obtain from Equation (6). In fact, recalling $\tau = T\Delta$, we have

$$\begin{aligned} M_N(s) &= \sum_{k=0}^{\infty} s^k \cdot P\{N = k\} = \\ &= s\left[1 - \frac{1 - e^{-\lambda T\Delta}}{\lambda T\Delta} + e^{\lambda T\Delta}\frac{(1 - e^{-\lambda T\Delta})^2}{\lambda T\Delta(e^{\lambda T} - 1)}\right] + \\ &\quad + \sum_{k=2}^{\infty} s^k \frac{e^{\lambda T\Delta}(1 - e^{-\lambda T\Delta})^2}{\lambda T\Delta(e^{\lambda T} - 1)}\left[\frac{\lambda T(1-\Delta)}{e^{\lambda T} - 1}\right]^{k-1} \\ &= s\left[1 - \frac{1 - e^{-\lambda T\Delta}}{\lambda T\Delta} + e^{\lambda T\Delta}\frac{(1 - e^{-\lambda T\Delta})^2}{\lambda T\Delta(e^{\lambda T} - 1)}\right] + \\ &\quad + s^2 \frac{e^{\lambda T\Delta}(1 - e^{-\lambda T\Delta})^2}{\Delta(e^{\lambda T} - 1)} \frac{1 - \Delta}{e^{\lambda T} - 1 - \lambda T s(1 - \Delta)}. \end{aligned}$$

As we are in the hypothesis $\lambda T \ll 1$, we can use the Taylor expansion to find that $M_N(s) = s\Delta + s^2 \frac{\Delta(1-\Delta)}{1-s(1-\Delta)} + \mathrm{o}(1)$. As the MGF of an exponential distribution S is given by $M_S(s) = \frac{\lambda}{\lambda - s}$, we obtain the following:

$$\begin{aligned} M_{\tilde{S}}(s) &= M_N(M_S(s)) = \\ &= \frac{\lambda\Delta}{\lambda - s} + \frac{\lambda^2}{(\lambda - s)^2} \cdot \frac{\Delta(1-\Delta)}{1 - \frac{\lambda}{\lambda - s}(1 - \Delta)} + \mathrm{o}(1) = \\ &= \frac{\lambda\Delta}{\lambda\Delta - s} + \mathrm{o}(1). \end{aligned} \tag{8}$$

Since the above equation corresponds, approximately, to the MGF of an exponential random variable with rate $\lambda\Delta$, we conclude that $\tilde{S}$ can be approximated as an exponential random variable with rate $\lambda\Delta$. A longer version of this proof is provided in Appendix A. ■

A. Validation

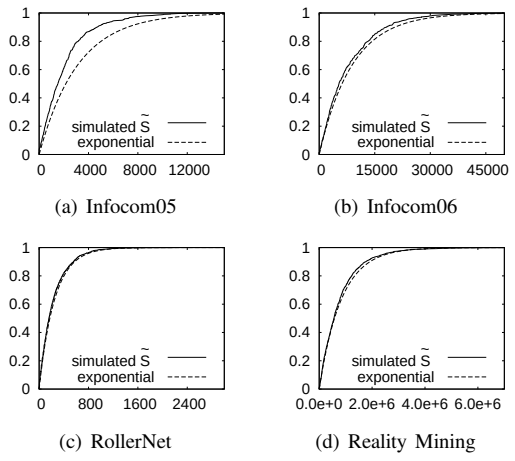
In this section, we consider the average rate measured in real datasets of human mobility and we verify whether our assumption $\lambda T \ll 1$ is reasonable and whether our model correctly predicts the distributions of the detected intercontact times. For opportunistic networks, there are several publicly available datasets obtained from experiments monitoring contacts between device pairs. Pairwise contacts are typically detected through Bluetooth [6], [7], [8] by means of special devices like iMotes. As we have discussed in Section IV, our model is suitable for this kind of scenarios. Here we consider four popular datasets often used in the related literature: Infocom05 [6], Infocom06 [6], RollerNet [7], and Reality Mining [8]. In Table II we report the average pairwise rate [21] extracted from these traces (corresponding to $\tilde{\lambda}$, i.e., to rates after duty cycling) and the associated λ values obtained applying Theorem 2, under the assumption that ICT are exponential. While this assumption is very strong in general,

TABLE II. QUALITY OF APPROXIMATION IN POPULAR DATASETS

Dataset	T	τ	$\tilde{\lambda}$	λ	λT
Infocom05	120	5	$3.2 \cdot 10^{-4}$	$7.7 \cdot 10^{-3}$	0.92
Infocom06	120	5	$1.13 \cdot 10^{-4}$	$2.7 \cdot 10^{-3}$	0.33
RollerNet	15	5	$4.07 \cdot 10^{-3}$	$1.2 \cdot 10^{-2}$	0.18
Reality	300	5	$1.2 \cdot 10^{-6}$	$7.2 \cdot 10^{-5}$	0.02

Tournoux et al. [7] have shown that it is acceptable for a significant percentage of pairs in some of these datasets. In addition, in Table II we also highlight the duration of the ON period used (corresponding to the duration of the Bluetooth phase in which a device scans for neighbours) and the period of the contact probing process. Using these parameters, we are able to compute λT and check whether our approximation holds (due to space reasons, here we perform an average analysis. A pairwise analysis is provided in Appendix A). As it can be seen in Table II, λT is smaller than 1 in all cases. Clearly, the farther from 1 the better, since we require $\lambda T \ll 1$. Thus, we expect the approximation that we make to be quite good for all datasets except for the Infocom05.

In order to complement the theoretical analysis in Section V, here we verify that our prediction for the distribution of $\tilde{S}$ actually matches simulation results exploiting the parameters of real experiments. Specifically, we take a tagged node pair and we assume that the meeting rate of this pair corresponds to the average meeting rate (average across all pairs of nodes in the traces) measured from the traces in Table II. With this approach we are able to represent the behavior of the average node pair. Then, we draw 10000 samples (100000 for the Reality Mining case, due to the long duty cycle period which led to fewer detections) from an exponential distribution, configured with the parameters λ in Table II. The sequence of these samples corresponds to the contact process between the tagged node pair. To this contact process we apply a duty cycling function with $\Delta = \frac{\tau}{T}$, where τ and T are taken again from Table II. Then we measure $\tilde{S}$ after each detected contact and we plot its CDF for all the four datasets in Figure 4. As expected, for the Infocom06 scenario, there are discrepancies between the actual and predicted values. For the other scenarios, in which the product λT is closer to zero, model prediction are very close to simulation results.



(a) Infocom05 (b) Infocom06

(c) RollerNet (d) Reality Mining

Fig. 4. CDF of $\tilde{S}$ in the different scenarios

VI. THE EFFECT OF DUTY CYCLING ON THE DELAY

In this section we exploit the results on the detected intercontact times derived above in order to compute the first two moments of the delay for a set of representative forwarding strategies designed for opportunistic networks. The first step in this direction is to use the first two moments of the detected pairwise intercontact times for approximating the distribution of the intercontact time itself. Please note that while the analysis in the previous section focused on a tagged pair of nodes, in this section we study the whole network. So, assuming exponential intercontact times, we denote their rates as λ_{ij} for node pair i, j. We have shown that, when assumption $\lambda T \ll 1$ holds, the detected intercontact times follow approximately an exponential distribution. Using this approximation, in the following we solve the analytical model proposed in [22] for both real intercontact times and detected intercontact times. The goal of this evaluation is to study how the first two moments of the delay are affected by energy saving techniques.

The forwarding model that we exploit represents the forwarding process in terms of a Continuous Time Markov Chain (CTMC)[23]. The chain has as many states as the nodes of the network and transitions between states depend both on the meeting process between nodes (i.e., their intercontact times) and on the forwarding protocol in use. Denoting the delay of messages from a generic node i to a tagged node d as D_i, and using standard Markov chain theory, it is possibile to derive the first two moments of D_i as in [22]. We report this result below for the convenience of the reader.

Lemma 2 (Delay's first and second moment): The first and second moment of the delay D_i for a message generated by node i and addressed to node d can be obtained from the minimal non-negative solutions, if they exists, to the following systems, respectively:

$$\forall i \neq d : E[D_i] = \frac{1}{\sum_{j \in \mathcal{R}_i} \lambda_{ij}} + \sum_{j \in \mathcal{R}_s - \{d\}} \frac{\lambda_{ij}}{\sum_{z \in \mathcal{R}_i} \lambda_{iz}} E[D_j] \tag{9}$$

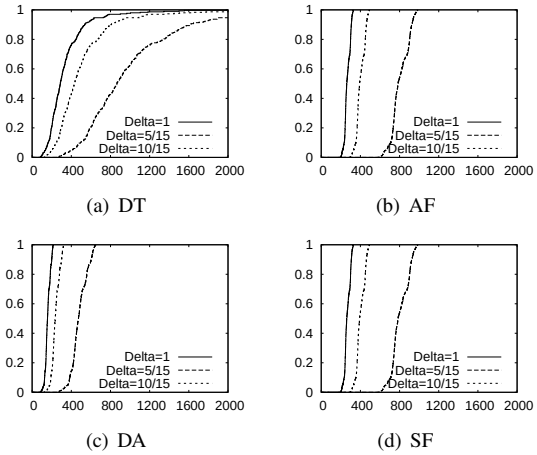
$$\forall i \neq d :$$
$$E[(D_i)^2] = \frac{2}{[\sum_{j \in \mathcal{R}_i} \lambda_{ij}]^2} + \sum_{j \in \mathcal{R}_s - \{d\}} \frac{\lambda_{ij}}{\sum_{z \in \mathcal{R}_i} \lambda_{iz}} E[D_i^2] + $$
$$+2 \sum_{j \in \mathcal{R}_s - \{d\}} \frac{\lambda_{ij}}{\sum_{z \in \mathcal{R}_i} \lambda_{iz}} \frac{1}{\sum_{j \in \mathcal{R}_i} \lambda_{ij}} E[D_i] \tag{10}$$

where we denote with $\mathcal{R}_i$ the set of possible relays towards destination d when the message is on node i (this set depends on the forwarding strategy in use). Please note that this model (as most analytical models in the literature) assumes buffers and bandwidth large enough for accommodating all messages.

In [22], we have defined a set of abstract policies able to capture significant aspects of popular state-of-the-art forwarding strategies. In the following we will focus on tthese policies. Under the Direct Transmission (DT) forwarding scheme, the source of the message is only allowed to hand it over to the destination itself, if ever encountered. With the Always Forward (AF) policy, the message is handed over by the

source, and the following relays to the first nodes encountered. Both DT and AF are social-oblivious (also known as context-oblivious or randomized) policies, i.e., they do not exploit information on node social relationships and contact behavior. In [22] two social-aware policies were also defined. In social-aware policies, each intermediate forwarder hands over the message to nodes that have a higher probability of bringing the message closer to the destination, according to some predefined forwarding metrics. The first of these policies is Direct Acquaintance (DA), in which the forwarding metric is the contact rate with the destination ($\frac{1}{E[S_{id}]}$): a better forwarder is one with a higher contact rate with respect to the node currently holding the message. The second policy is Social Forwarding (SF), for which the forwarding metric is $\beta\frac{1}{E[S_{i,d}]} + (1-\beta)\sum_{j\in\mathcal{P}_i} w_{ij}\frac{1}{E[S_{jd}]}$, where $w_{ij} = \frac{\lambda_{ij}}{\sum_{j\in\mathcal{P}_i}\lambda_{ij}}$. With respect to the DA policy, which only captures direct meetings with the destination, SF is also able to detect indirect meetings, allowing nodes to select relays that not only meet the destination frequently but also meet nodes that meet the destination frequently.

In the following we assume that nodes intercontact times are exponential. The fitting analysis presented in [23] has shown that contact rates in the traces already considered in Section V-A follow a Gamma distribution. Below, we focus on the distribution parameters for the RollerNet scenario reported in [21], i.e., shape $\xi = 4.43$, rate $r = 1088$. We consider a network made up of 25 nodes and we solve the forwarding model described above in the case of duty cycle equal to $\frac{5}{15}$, $\frac{10}{15}$ and 1 (no duty cycling). Figures 5-6 show the CDF of the moments of the delay in this case. As expected, both the first and second moment become larger as we reduce the ON interval in the duty cycle. In fact, as discussed before, the neat effect of duty cycling is to effectively reduce the number of usable contacts to only those happening during an ON period. The shorter the ON period, the fewer the usable contacts every T, the longer the delay.

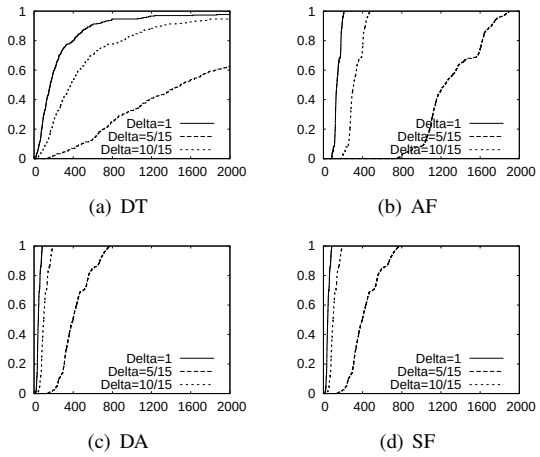


(a) DT (b) AF

(c) DA (d) SF

Fig. 5. CDF of the first moment of the delays for the different forwarding algorithms

Let us now see what happens to the coefficient of variation of the delay (Figure 7). For all pairs, all forwarding strategies and all duty cycling values, the coefficient of variation is bigger than one. This means that the delay can be approximated with

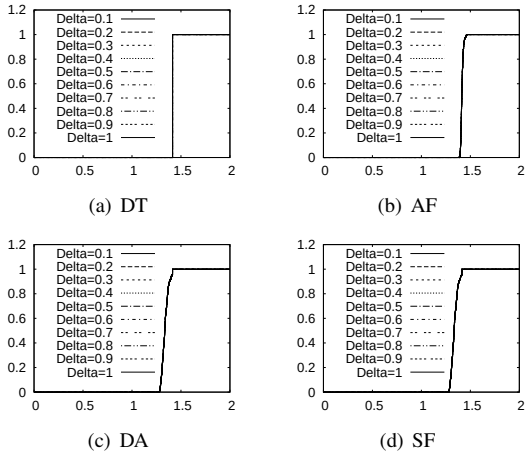


(a) DT (b) AF

(c) DA (d) SF

Fig. 6. CDF of the second moment of the delays for the different forwarding algorithms

an hyper-exponential distribution. In the next section, we will use this representation of the delay in terms of the hyper-exponential distribution in order to compute the volume of traffic carried by the network. Another interesting observation from Figure 7 is that the coefficient of variation does not depend on the duty cycle Δ (in fact, all curves overlap). This means that, in the case of exponential intercontact times, the duty cycle does not introduce variability in the delay experienced by messages.



(a) DT (b) AF

(c) DA (d) SF

Fig. 7. CDF of the coefficient of variation of the delay with different Δ

VII. Energy, traffic, and network lifetime

In this section we investigate the benefits of implementing a duty cycling policy on the nodes of the network. These benefits are in terms of energy saved by nodes and, consequently, increased network lifetime. In fact, assuming that nodes have an energy budget L (expressed as the amount of time they can be on when no duty cycling is implemented), by definition nodes' lifetime is extended to $\frac{L}{\Delta}$ when energy saving strategies are in place[4]. At the same time, however, as we have seen in the

[4]We consider only the part of the energy budget related to networking activities.

previous section, the gain in terms of energy is counterbalanced by a loss from the delay standpoint. In fact, the expected delay increases as Δ decreases, so the network lifetime is longer but nodes also need more time to deliver messages.

In the following we want to study the following three aspects. First, what is the relationship between the energy consumed with and without duty cycling. Second, what is the volume of traffic carried by the network with and without duty cycling. Third, whether there exists an optimal duty cycle value for which the loss in terms of delay is minimum and the gain in terms of traffic carried by the network is maximum.

A. Energy with and without duty cycling

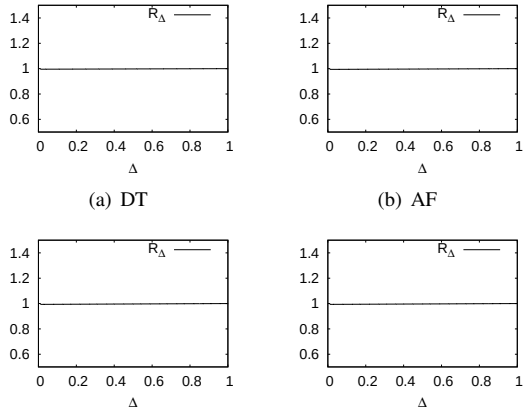
We first assume that messages are all generated at time $t = 0$ and that L is very large. The goal here is to understand how much energy is saved by duty cycling, without considering the limited network lifetime, i.e., just taking into account the delivery of standalone messages without temporal limitations. Throughout the section we use a simple energy model in which nodes consume a certain power w (measured in watts) during ON intervals and zero otherwise. In reality, depending on the semantic of the OFF interval, nodes may or may not consume energy. If, for example, the OFF interval corresponds to the *inquiry scan* state of Bluetooth (when the devices listens and responds to inquiries but do not issues inquiries itself) some energy is used (though quite low, see Section II), if it corresponds to devices actually turned off there is instead no consumption. In addition, real devices also consume a significant amount of energy during transmission and reception. Here, however, we have chosen to neglect the consumption due to tx/rx phases for the following reason. Under our approximation, the intercontact rates after duty cycling are equal to the rates before duty cycling scaled by $\frac{1}{\Delta}$. This means that the relative ranking between rates remains the same, and hence the forwarding decisions will remain the same as well. This implies that the number of hops that messages go through with or without duty cycling, in the exponential case, are statistically the same. Since tx/rx energy is associated with each hop, the amount of tx/rx energy consumed with or without duty cycling is the same. Being this amount of energy constant and independent of Δ we have chosen to neglect it and focus instead on the variable consumption.

Below we focus on the expected delay across the whole network, for having a compact representation. Please note however that the behaviour is the same for the single pairs of nodes. We measure the energy consumed as the product between power w and the length of the time interval for which the network is ON. Without duty cycling, the network is ON for the whole time it takes to deliver a message (hence, for $E[D]$), while in case of duty cycling, the network is ON only for a fraction Δ of the time ($E[D_\Delta]$, where we denote with D_Δ the delay under duty cycling Δ) it takes to complete the delivery. In order to measure the relationship between the two quantities, we study the following:

$$R_\Delta = \frac{wE[D]}{w\Delta E[D_\Delta]} = \frac{E[D]}{\Delta E[D_\Delta]}. \tag{11}$$

If the above ratio is around 1, it means that when a duty cycling policy is in place the amount of energy spent is the same but instead of being used all at once it is fractioned and alternated with intervals in which none is used. When the above ratio is greater than 1 we have that the energy required without duty cycling is higher. Vice versa, when the ratio is smaller than 1, the energy needed with duty cycling is higher.

Figure 8 (derived for the same parameters of the RollerNet scenario used above) tells us that the ratio stays around 1 independently of the specific duty cycle value Δ. This result is very interesting, because it shows that under exponential intercontact times the energy consumed for a standalone message is the same, regardless of the value of Δ. With $\Delta = 1$ the energy budget needed for sending the message is simply used all at once, while with $\Delta < 1$ this budget is spread across different intervals of duration T, during which the nodes of network are partially turned off. This is consistent with the assumption of unlimited network lifetime and standalone messages (all generated at $t = 0$) that we have made in this section. In the next section, we discuss instead what happens when nodes have a limited energy budget L and messages are continuously generated by nodes.



(a) DT (b) AF

(c) DA (d) SF

Fig. 8. R_Δ varying Δ in the different forwarding algorithms

B. Traffic carried by the network

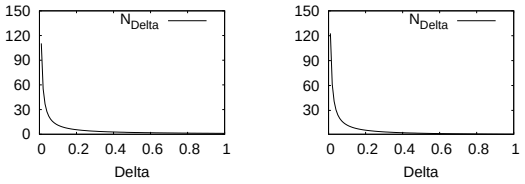
For a more realistic evaluation, we hereafter assume that messages are generated according to a Poisson process with rate μ. By definition, the number of messages created in disjoint timeslots are independent [23]. Thus, the number of messages arriving during a time interval of length dt is given by μdt. When $\Delta = 1$ these messages keep arriving until L, after which the network has exhausted all its energy budget and turns off indefinitely. Instead, when $\Delta < 1$ the network takes longer to consume all its energy budget, thus stays on until $\frac{L}{\Delta}$. In the following we study the volume of messages delivered by the network with and without duty cycling, measured as the number of messages delivered in $\frac{L}{\Delta}$. Then, the following result holds, whose proof can be found in Appendix A.

Theorem 3: The volume N_Δ of messages delivered by the system under duty cycling Δ is given by:

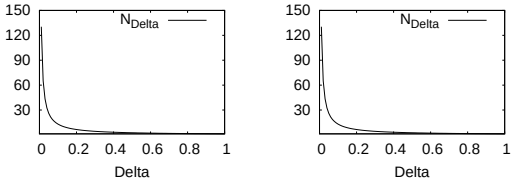
$$N_\Delta = \frac{\mu L}{\Delta} - \mu E[D_\Delta]\left[1 - \frac{1}{2}e^{\frac{-L}{E[D_\Delta]\Delta}}\left(e^{\left(1+\sqrt{\frac{c^2-1}{c^2+1}}\right)} + e^{\left(1-\sqrt{\frac{c^2-1}{c^2+1}}\right)}\right)\right], \tag{12}$$

where c is the coefficient of variation of the delay.

Basically, N_Δ is given by the number of messages generated during the network lifetime (the first term in Equation (12)) minus the number of messages that are not delivered before the energy budget is depleted. The latter quantity is a function of the expected value of the delay and of its variability. We now exploit Theorem 3 in order to study N_Δ. Specifically, in Figure 9 below we show how N_Δ varies with different duty cycles, where we assume that each node generates 1 message every 10 minutes (so $\mu = \frac{1}{600}$). The plot is drawn for a tagged node pair for the sake of readability, but the same results hold for the other pairs. We see that the volume of traffic carried by the network (i.e., the number of messages delivered on average during network lifetime) increases as the duty cycle Δ decreases. So, as expected, increasing the network lifetime more messages get a chance of being delivered but the price to pay, as seen in Section VI, is a larger expected delay.



(a) DT (b) AF



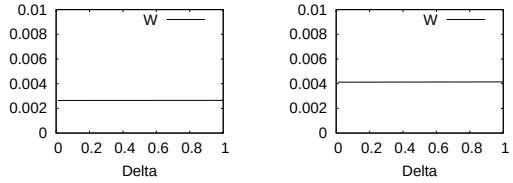
(c) DA (d) SF

Fig. 9. N_Δ varying Δ with different forwarding algorithms

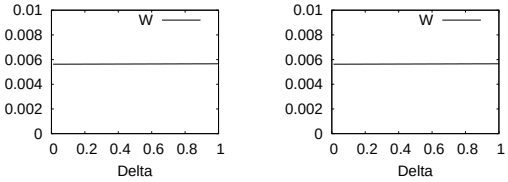
C. Trade-off between delay and volume of traffic

In the final part of our evaluation, we study whether it exists an optimal working point that minimises the expected delay and maximises the volume of messages exchanged. We borrow the definition of *power of the network* (which we denote with $\mathcal{W}$) from traditional queueing theory [24]. Quantity $\mathcal{W}$ measures the trade-off between the traffic N_Δ carried by the network (function of the message injection rate μ) and the expected delay $E[D_\Delta]$. The power is then defined as $\mathcal{W} = \frac{N_\Delta}{D_\Delta}$. In traditional queueing systems, the above trade-off was regulated by contention. In fact, under limited resources, we could not increase indefinitely the quantity of messages successfully delivered without affecting the resulting expected delay (because, e.g., under heavy traffic, packets start to be discarded from buffers). In our case, we do not have contention, since we assume that there are no limitations on buffers and bandwidth. Our knob is instead the duty cycle. When Δ approaches 1, delays are as short as possible given the underlying mobility, but a lot of energy is consumed and the network lifetime is shorter. If we want to increase network lifetime, we have to sacrifice the expected delay.

Figure 10 shows how $\mathcal{W}$ varies depending on Δ. It can be clearly seen that $\mathcal{W}$ remains practically constant, which implies that whatever one gains in network lifetime is immediately lost in expected delay. Thus, under exponential intercontact



(a) DT (b) AF



(c) DA (d) SF

Fig. 10. $\mathcal{W}$ varying Δ in the different forwarding algorithms ($\mu = \frac{1}{600}$)

times, there is no optimal choice of Δ and all working points are equivalent. When configuring a duty cycling policy, the operator can thus only maximise one metric at a time. Note, anyway, that if the total number of messages to be delivered is of primary concern, at the cost of additional delay, then reducing the duty cycling is clearly effective.

VIII. Conclusion

In this work we have investigated the effects of duty cycling on intercontact times, delay, and energy consumption in opportunistic networks. To the best of our knowledge, this is the first contribution that evaluates the actual effects of duty cycling on the forwarding opportunities between nodes. To this aim, we have provided a general formula for the derivation of the intercontact times under duty cycling, and we have specialised this formula obtaining a closed-form expression for the case of exponential intercontact times. Surprisingly enough, under condition $\lambda T \ll 1$ satisfied by most popular contact datasets, the intercontact times after duty cycling can be approximated as exponentially distributed with a rate scaled by a factor $\frac{1}{\Delta}$. Exploiting this result, we have then studied the first two moments of the delay under duty cycling, showing that these moments both increase as Δ becomes smaller. Finally, we have focused on how the network lifetime is impacted by duty cycling, highlighting the fact that a larger volume of traffic is handled by the network when a duty cycling policy is in place, because the network lifetime is increased even if at the expense of the delay experienced by messages. In addition, we have found that it is not possible to derive an optimal duty cycle value that maximises the network lifetime while at the same time minimally impacting the expected delay.

As future work, two main different directions can be pursued. First, we plan to extend the analysis to intercontact times featuring a distribution different from the exponential, e.g., the Pareto intercontact times case, which also represents a popular hypothesis for the intercontact times considered in the literature [25]. Second, we also plan to evaluate different duty cycling policies, e.g., some in which the duration of ON and OFF intervals is not fixed but varies according to a specified distribution.

ACKNOWLEDGMENT

This work was partially funded by the European Commission under the EINS (FP7-FIRE 288021), MOTO (FP7 317959), and EIT ICT Labs MONC Project (Business Plan 2013) projects.

REFERENCES

[1] R. Friedman, A. Kogan, and Y. Krivolapov, "On power and throughput tradeoffs of wifi and bluetooth in smartphones," *IEEE Trans. on Mob. Comp.*, vol. 12, no. 7, pp. 1363–1376, 2013.

[2] S. Gaito, E. Pagani, and G. P. Rossi, "Strangers help friends to communicate in opportunistic networks," *Computer Networks*, vol. 55, no. 2, pp. 374–385, 2011.

[3] H. Tijms and J. Wiley, *A first course in stochastic models*. Wiley Online Library, 2003, vol. 2.

[4] A. Picu, T. Spyropoulos, and T. Hossmann, "An analysis of the information spreading delay in heterogeneous mobility dtns," in *IEEE WoWMoM*, 2012, pp. 1–10.

[5] W. Gao and G. Cao, "User-centric data dissemination in disruption tolerant networks," in *IEEE INFOCOM*, 2011, pp. 3119–3127.

[6] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*. ACM, 2005, pp. 244–251.

[7] P.-U. Tournoux, J. Leguay, F. Benbadis, J. Whitbeck, V. Conan, and M. D. de Amorim, "Density-aware routing in highly dynamic dtns: The rollernet case," *IEEE Trans. on Mob. Comp.*, vol. 10, no. 12, pp. 1755–1768, 2011.

[8] N. Eagle and A. (Sandy) Pentland, "Reality mining: Sensing complex social systems," *Personal Ubiquitous Comput.*, vol. 10, no. 4, pp. 255–268, 2006.

[9] J.-C. Cano, J.-M. Cano, E. González, C. Calafate, and P. Manzoni, "Power characterization of a bluetooth-based wireless node for ubiquitous computing," in *Wireless and Mobile Communications, 2006. ICWMC'06. International Conference on*. IEEE, 2006, pp. 13–13.

[10] S. Trifunovic, A. Picu, T. Hossmann, and K. A. Hummel, "Slicing the battery pie: fair and efficient energy usage in device-to-device communication via role switching," in *Proceedings of the 8th ACM MobiCom workshop on Challenged networks*. ACM, 2013, pp. 31–36.

[11] "WiFi Direct." [Online]. Available: http://www.wi-fi.org/discover-and-learn/wi-fi-direct

[12] S. Trifunovic, B. Distl, D. Schatzmann, and F. Legendre, "Wifi-opp: ad-hoc-less opportunistic networking," in *Proceedings of the 6th ACM workshop on Challenged networks*. ACM, 2011, pp. 37–42.

[13] H. Jun, M. H. Ammar, and E. W. Zegura, "Power management in delay tolerant networks: a framework and knowledge-based mechanisms." in *IEEE SECON*, vol. 5, 2005, pp. 418–429.

[14] B. J. Choi and X. Shen, "Adaptive exponential beacon period protocol for power saving in delay tolerant networks," in *Communications, 2009. ICC'09. IEEE International Conference on*. IEEE, 2009, pp. 1–6.

[15] C. Drula, C. Amza, F. Rousseau, and A. Duda, "Adaptive energy conserving algorithms for neighbor discovery in opportunistic bluetooth networks," *IEEE JSAC*, vol. 25, no. 1, pp. 96–107, 2007.

[16] W. Wang, M. Motani, and V. Srinivasan, "Opportunistic energy-efficient contact probing in delay-tolerant applications," *IEEE/ACM Transactions on Networking (TON)*, vol. 17, no. 5, pp. 1592–1605, 2009.

[17] E. Altman, A. Azad, T. Başar, and F. De Pellegrini, "Combined optimal control of activation and transmission in delay-tolerant networks," *Networking, IEEE/ACM Transactions on*, vol. 21, no. 2, pp. 482–494, 2013.

[18] H. Zhou, H. Zheng, J. Wu, and J. Chen, "Energy-efficient contact probing in opportunistic mobile networks," in *Proc. of the 22nd International Conference on Computer Communications and Networks*, 2013.

[19] S. Qin, G. Feng, and Y. Zhang, "How the contact-probing mechanism affects the transmission capacity of delay-tolerant networks," *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 4, pp. 1825–1834, 2011.

[20] D. Bohman, M. Frank, P. Martini, and C. Scholz, "Performance of symmetric neighbor discovery in bluetooth ad hoc networks." in *GI Jahrestagung (1)*, 2004, pp. 138–142.

[21] A. Passarella and M. Conti, "Analysis of individual pair and aggregate intercontact times in heterogeneous opportunistic networks," *IEEE Trans. on Mob. Comp.*, vol. 12, no. 12, pp. 2483–2495, 2013.

[22] C. Boldrini, M. Conti, and A. Passarella, "Modelling social-aware forwarding in opportunistic networks," in *Performance Evaluation of Computer and Communication Systems. Milestones and Future Challenges*. Springer, 2011, pp. 141–152.

[23] S. M. Ross, *Introduction to probability models*. Access Online via Elsevier, 2006.

[24] R. Gail and L. Kleinrock, "An invariant property of computer network power," in *Conference Record, International Conference on Communications*, 1981, pp. 63–1.

[25] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Trans. on Mob. Comp.*, vol. 6, no. 6, pp. 606–620, 2007.

[26] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to probability*. Athena Scientific, Belmont, Massachusetts, 2008.

APPENDIX A
PROOFS AND FURTHER RESULTS

Theorem 1 (Distribution of N): The discrete probability density of N can be approximated by the following:

$$P\{N=1\} = \sum_{n_1=0}^{\infty} \int_{s_0}^{s_1} f_{\hat{S}_0}(t_0) \int_{I_{n_1}-t_0} f_{S_1}(t_1) dt_1 dt_0$$

$$P\{N=2\} = \sum_{n_1=0}^{\infty} \sum_{n_2=n_1+1}^{\infty} \int_{s_0}^{s_1} f_{\hat{S}_0}(t_0) \int_{J_{n_1}-t_0} f_{S_1}(t_1) \cdot \cdot \int_{I_{n_2}-t_0-t_1} f_{S_2}(t_2) dt_0 dt_1 dt_2$$

$$\vdots$$

$$P\{N=k\} = \sum_{n_1=0}^{\infty} \cdots \sum_{n_k=n_{k-1}+1}^{\infty} \int_{s_0}^{s_1} f_{\hat{S}_0}(t_0) \int_{t_1 \in J_{n_1}-t_0} f_{S_1}(t_1) \cdot \cdots \int_{t_{k-1} \in J_{n_{k-1}}-t_0-\cdots-t_{k-2}} f_{S_{k-1}}(t_{k-1}) \cdot \int_{t_k \in I_{n_k}-t_0-\cdots-t_{k-1}} f_{S_k}(t_k) dt_k dt_{k-1} \ldots dt_1 dt_0.$$

Proof: Let us start with event $\{N=1\}$. As discussed above, it occurs when the time of the first contact event falls in an ON interval (Figure 2). Recalling that S_1 denotes the interarrival time between the zero and first contact, we require that $t_0 + S_1 \in [t_0, s_1] \cup \bigcup_{n_1=1}^{\infty} I_{n_1}$, i.e, that $t_0 + S_1$ belongs to any of the ON intervals starting from t_0. Forcing a bit the notation in order to have a compact formula, we define $I_0 = (t_0, s_1]$, which corresponds to the usable portion of the first ON interval after t_0. From this consideration we can obtain the following:

$$P\{N=1|\hat{S}_0=t_0\} = \sum_{n_1=0}^{\infty} P(t_0+S_1 \in I_{n_1}) = \sum_{n_1=0}^{\infty} \int_{I_{n_1}-t_0} f_{S_1}(t_1) dt_1.$$

Then, applying the law of total probability, we get $P\{N=1\} = \int_{s_0}^{s_1} P\{N=1|\hat{S}_0=t_0\} * f_{\hat{S}_0}(t_0) dt_0$, from which Equation (3) immediately follows.

Event $\{N = 2\}$ occurs when the first contact happens during an OFF period and the second contact during an ON period (Figure 2). For the first contact to be in an OFF period, we need that $t_0+S_1 \in \bigcup_{n_1=0}^{\infty} J_{n_1}$. For the second contact to be in an ON period we require that $t_0+S_1+S_2 \in \bigcup_{n_2=n_1+1}^{\infty} I_{n_2}$, i.e., that the second contact happens in any of the ON periods *after* the OFF period in which the first contact has been missed. Conditioning on t_0 we can derive $P\{N = 2|\hat{S}_0 = t_0\}$ as follows:

$$P\{N=2|\hat{S}_0=t_0\} = \sum_{n_1=0}^{\infty} P(t_0+S_1 \in J_{n_1}) \cdot$$
$$\cdot P(t_0+S_1+S_2 \in \bigcup_{n_2=n_1+1}^{\infty} I_{n_2} \mid t_0+S_1 \in J_{n_1}) =$$
$$= \sum_{n_1=0}^{\infty} \int_{J_{n_1}-t_0} f_{S_1}(t_1) \sum_{n_2=n_1+1}^{\infty} \int_{I_{n_2}-t_0-t_1} f_{S_2}(t_2) dt_1 dt_2 =$$
$$= \sum_{n_1=0}^{\infty} \sum_{n_2=n_1+1}^{\infty} \int_{J_{n_1}-t_0} f_{S_1}(t_1) \int_{I_{n_2}-t_0-t_1} f_{S_2}(t_2) dt_1 dt_2.$$

Then, if we apply the law of total probability, we get that $P\{N = 2\} = \int_{s_0}^{s_1} P\{N = 2|\hat{S}_0 = t_0\} f_{\hat{S}_0}(t_0) dt_0$, from which Equation (4) follows.

Let us now consider case $\{N = k\}$ with $k \geq 3$, in which there are $k-1$ consecutive failures (i.e., contact events happening during an OFF period) before the first success. Two consecutive undetected contact events can fall in the same OFF interval (i.e., $n_i = n_{i+1}$), or in two different OFF intervals. If we neglect the case of more than one undetected contacts in the same OFF period[5], the complexity of the problem significantly diminish. For this reason, in the following we assume that this is the case in the scenario under study, and in Section IV-A we derive the conditions under which this assumption is reasonable and we show that these conditions are satisfied by the most popular traces of human contacts. So, assuming that the probability that two consecutive missed contacts fall into the same OFF interval is negligible, the probability of event $\{N = k\}$ can be obtained reasoning in the following way:

1) the first contact event (undetected) happens at time $t_0 + S_1 \in J_{n_1}$;
2) assuming $S_1 = t_1$, the second contact event (undetected) happens at time $t_0 + t_1 + S_2 \in J_{n_2}$, with $n_2 > n_1$;

$$\vdots$$

3) assuming $S_{k-2} = t_{k-2}$, the $(k-1)$-th contact event (undetected) happens at time $t_0+t_1+\cdots+t_{k-2}+S_{k-1} \in J_{n_{k-1}}$, with $n_{k-1} > n_{k-2}$;
4) assuming $S_{k-1} = t_{k-1}$, the k-th contact event (detected) happens at time $t_0 + t_1 + \cdots + t_{k-1} + S_k \in I_{n_k}$, with $n_k > n_{k-1}$.

Translating the above into a mathematical formula, following the same line of reasoning used for cases $N = 1$ and $N = 2$, we obtain Equation (5). ■

Corollary 1 (N with exponential intercontact times): When real intercontact times S_i are exponential with rate λ, the probability density of N is given by:

$$\begin{cases} P\{N=1\} = 1 + \frac{e^{-\lambda\tau}-1}{\lambda\tau} + \frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{\lambda\tau(e^{\lambda T}-1} \\ P\{N=k\} = e^{\lambda\tau}\frac{(1-e^{-\lambda\tau})^2}{\lambda\tau(1-e^{-\lambda T})}\left[\frac{\lambda(T-\tau)}{e^{\lambda T}-1}\right]^{k-1}, \quad k \geq 2 \end{cases}$$

Proof: In this proof we show to how solve Theorem 1 when intercontact times feature an exponential distribution. We use the fact that $f_{S_i}(t) = \lambda e^{-\lambda t}$ for the exponential distribution and the following algebraic relation:

$$\sum_{n_1=0}^{\infty} \sum_{n_2=n_1+1}^{\infty} \cdots \sum_{n_{k-1}=n_{k-2}+1}^{\infty} \sum_{n_k=n_{k-1}+1} x^{n_k} = \frac{x^{k-1}}{(1-x)^k}, \tag{A1}$$

that is true for every $|x| < 1$. We omit the proof for $P\{N = 1\}$ and $P\{N = 2\}$ since it is straightforward to solve them once substituting into Equations (3)-(4) $f_{S_i}(t) = \lambda e^{-\lambda t}$ and $f_{\hat{S}_0}(t_0)$ that we discuss below.

So, let us focus on $P\{N = k\}$. $P\{N = k|\hat{S}_0 = t_0\}$ can be obtained from Equation (5) as described below:

$$\begin{aligned} &P\{N=k|\hat{S}_0=t_0\} = \\ &= \sum_{n_1=0}^{\infty} \cdots \sum_{n_k=n_{k-1}+1}^{\infty} \int_{t_1 \in J_{n_1}} \lambda e^{-\lambda(t_1-t_0)} \\ &\quad \int_{t_2 \in J_{n_2}} \lambda e^{-\lambda(t_2-t_1)} \cdots \\ &\quad \cdots \int_{t_{k-1} \in J_{n_{k-1}}} \lambda e^{-\lambda(t_{k-1}-t_{k-2})} \\ &\quad \int_{t_k \in I_{n_k}} \lambda e^{-\lambda(t_k-t_{k-1})} dt_k dt_{k-1} \ldots dt_1 \\ &= \lambda^{k-1}(s_0+T-s_1)^{k-1} e^{\lambda t_0} \left(e^{-\lambda s_0} - e^{-\lambda s_1}\right) \\ &\quad \sum_{n_1=0}^{\infty} \sum_{n_2=n_1+1}^{\infty} \cdots \sum_{n_{k-1}=n_{k-2}+1}^{\infty} \sum_{n_k=n_{k-1}+1}^{\infty} e^{-\lambda n_k T} \\ &= \lambda^{k-1}(T-\tau)^{k-1} e^{\lambda(t_0-s_0)}(1-e^{-\lambda\tau}) \frac{e^{-(k-1)\lambda T}}{(1-e^{-\lambda T})^k} \\ &= e^{\lambda(t_0-s_0)} \frac{1-e^{-\lambda\tau}}{1-e^{-\lambda T}} \left[\frac{\lambda(T-\tau)}{e^{\lambda T}-1}\right]^{k-1}. \end{aligned}$$

In the above derivation we have used Equation (A1) whith $x = e^{-\lambda T}$, which can be applied as $e^{-\lambda T}$ is positive and smaller than 1. In order to derive $P\{N = k\}$ as $\int_{s_0}^{s_1} f_{\hat{S}_0}(t_0) P\{N = k|\hat{S}_0 = t_0\} \mathrm{d}t_0$, we have to determine the distribution of $\hat{S}_0$. Recall that $\hat{S}_0$ describes the arrival time of the zero event in its ON interval $[s_0, s_1]$. Since we are focusing on exponential intercontact times, the contact process is a Poisson process, hence the probability of an arrival in an interval $[s_0, s_1]$ is uniform over the interval. Thus, we can

[5]Please note that this implies that we also neglect multiple contacts detected in the same ON interval.

easily compute $P\{N=k\}$ as follows:

$$P\{N=k\} = \int_{s_0}^{s_1} \frac{1}{s_1-s_0} P\{N=k|\hat{S}_0=t_0\}dt_0 = \\ = \int_{s_0}^{s_1} \frac{1}{s_1-s_0} e^{\lambda(t_0-s_0)} \frac{1-e^{-\lambda\tau}}{1-e^{-\lambda T}} \left[\frac{\lambda(T-\tau)}{e^{\lambda T}-1}\right]^{k-1} dt_0 = \\ = e^{\lambda\tau} \frac{(1-e^{-\lambda\tau})^2}{\lambda\tau(1-e^{-\lambda T})} \left[\frac{\lambda(T-\tau)}{e^{\lambda T}-1}\right]^{k-1}$$

■

Lemma A1 (First two moments of N): When intercontact times are exponential, the first two moments of N are given by the following:

$$E[N] = 1 - \frac{1-e^{-\lambda\tau}}{\lambda\tau} + \frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{\lambda\tau(e^{\lambda T}-1)} + \\ +(T-\tau)e^{\lambda\tau}\frac{(1-e^{-\lambda\tau})^2}{\tau(1-e^{-\lambda T})}\cdot \\ \cdot\frac{-\lambda(T-\tau)+2e^{\lambda T}-2}{(e^{\lambda T}-1-\lambda(T-\tau))^2} \quad \text{(A2)}$$

$$E[N^2] = 1 - \frac{1-e^{-\lambda\tau}}{\lambda\tau} + \frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{\lambda\tau(e^{\lambda T}-1)} + \\ +(T-\tau)e^{\lambda\tau}\frac{(1-e^{-\lambda\tau})^2}{\tau(1-e^{-\lambda T})}\cdot \\ \cdot\frac{\lambda^2(T-\tau)^2-3\lambda(T-\tau)(e^{\lambda T}-1)+4(e^{\lambda T}-1)^2}{(e^{\lambda T}-1-\lambda(T-\tau))^3} \quad \text{(A3)}$$

Proof: Using the formula in Equation (6) we can calculate the first and the second moments of N, using the properties of the geometrical series. In fact the first moment can be computed, from standard probability theory, as follows:

$$E[N]=\sum_{k=1}^{\infty} kP\{N=k\} = \\ =1-\frac{1-e^{-\lambda\tau}}{\lambda\tau}+\frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{\lambda\tau(e^{\lambda T}-1)} \\ +\sum_{k=2}^{\infty} k\frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{\lambda\tau(1-e^{-\lambda T})}\left[\frac{\lambda(T-\tau)}{e^{\lambda T}-1}\right]^{k-1}.$$

If we are able to show that $\left|\frac{\lambda(T-\tau)}{e^{\lambda T}-1}\right|$ is smaller than 1 we can apply identity $\sum_{k=2}^{\infty} kx^{k-1} = \frac{x(2-x)}{(x-1)^2}$, which holds when $|x|<1$. To this aim, we can rewrite $\frac{\lambda(T-\tau)}{e^{\lambda T}-1}$ as $\frac{\lambda T(1-\Delta)}{e^{\lambda T}-1}$ and observe that this quantity is always greater than or equal to zero. Hence, we can study inequality $\frac{\lambda T(1-\Delta)}{e^{\lambda T}-1} < 1$ and investigate whether it is satisfied or not. The above can be rewritten as $\frac{e^{\lambda T}-1-\lambda T(1-\Delta)}{e^{\lambda T}-1} > 0$. The denominator is by definition greater than or equal to zero, so we can focus on the enumerator. Function $g(y) = e^y - y(1-\Delta) - 1$ is increasing for $y>0$ and such that $g(0)=0$. So $g(y)$ is positive for every positive y, from which follows that identity $\sum_{k=2}^{\infty} kx^{k-1} = \frac{x(2-x)}{(x-1)^2}$ can be applied. After simple mathematical substitutions we then obtain Equation (A2).

The second moment can be obtained as $\sum_{k=1}^{N} k^2 * P\{N=k\}$, from which the equation below follows:

$$E[N^2]=\sum_{k=1}^{\infty} k^2 P\{N=k\} = \\ =1-\frac{1-e^{-\lambda\tau}}{\lambda\tau}+\frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{\lambda\tau(e^{\lambda T}-1)} \\ +\sum_{k=2}^{\infty} k^2\frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{\lambda\tau(1-e^{-\lambda T})}\left[\frac{\lambda(T-\tau)}{e^{\lambda T}-1}\right]^{k-1}$$

We have proved above that $\frac{\lambda(T-\tau)}{e^{\lambda T}-1}$ is positive and less than 1. Exploiting this results, we apply identity $\sum_{k=2}^{\infty} k^2x^{k-1} = -\frac{x(x^2-3x+4)}{(x-1)^3}$, holding for every $|x|<1$, and we obtain Equation (A3). ■

Lemma A2: The first and second moment of $\tilde{S}$ are given by:

$$E[\tilde{S}]=\frac{1}{\lambda}-\frac{1-e^{-\lambda\tau}}{\lambda^2\tau}+\frac{e^{\lambda\tau}(1-e^{-\lambda\tau})}{\lambda^2\tau(e^{\lambda T}-1)}+ \\ +(T-\tau)e^{\lambda\tau}\frac{(1-e^{-\lambda\tau})^2}{\lambda\tau(1-e^{-\lambda T})}\cdot\frac{-\lambda(T-\tau)+2e^{\lambda T}-2}{\left(e^{\lambda T}-1-\lambda(T-\tau)\right)^2} \quad \text{(A4)}$$

$$E[\tilde{S}^2]=\frac{1}{\lambda^2}\left[2-\frac{2(1-e^{-\lambda\tau})}{\lambda\tau}+\frac{2e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{\lambda\tau(e^{\lambda T}-1)}+\right. \\ +2(T-\tau)e^{\lambda\tau}\frac{(1-e^{-\lambda\tau})^2}{\tau(1-e^{-\lambda T})}\cdot \\ \left.\cdot\frac{\lambda(T-\tau)[3+\lambda(T-\tau)-3e^{\lambda T}]+3e^{2\lambda T}-6e^{\lambda T}+3}{\left(e^{\lambda T}-1-\lambda(T-\tau)\right)^3}\right] \quad \text{(A5)}$$

Proof: Exploiting the properties of the random sum of random variables, it is easy to derive ([26] p. 241) that $E[\tilde{S}] = E[N]E[S]$ and $\sigma^2_{\tilde{S}} = E[N]\sigma^2_S + E[S]^2\sigma^2_N$. For the exponential random variable S of parameter λ the first two moments are given by $E[S]=\frac{1}{\lambda}$ and $E[S^2]=\frac{2}{\lambda^2}$. The moments of N are given by (A2)-(A3). Substituting these results into $E[\tilde{S}] = E[N]E[S]$, $\sigma^2_{\tilde{S}} = E[N]\sigma^2_S + E[S]^2\sigma^2_N$, we obtain Equation (A4). ■

Lemma 1: When $\lambda T \ll 1$, the error $\mathcal{E}$ introduced by the approximation of Corollary 1 approaches zero.

Proof: Assuming that condition $\lambda T \ll 1$ holds true, we can use the Taylor series for the exponential function to obtain the following:

$$\mathcal{E}(\tau,T,\lambda) = \frac{1-e^{-\lambda\tau}}{\lambda\tau} - \frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{\lambda\tau(\lambda T+\mathrm{o}(\lambda T)} + \\ - \frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2(T-\tau)}{\left[\lambda T-\frac{\lambda^2T^2}{2}+\mathrm{o}(\lambda^2T^2)\right]\tau\left[\lambda\tau+\frac{\lambda^2T^2}{2}+\mathrm{o}(\lambda^2T^2)\right]} \\ = \frac{1-e^{-\lambda\tau}}{\lambda\tau} - \frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2}{\lambda\tau(\lambda T+\mathrm{o}(\lambda T))} + \\ - \frac{e^{\lambda\tau}(1-e^{-\lambda\tau})^2(T-\tau)}{\tau[\lambda^2T\tau-\frac{\lambda^3t^2\tau}{2}+\mathrm{o}(\lambda^2T^2)]}$$

Observing that $\tau = \Delta T$ and again using the Taylor series for the exponential function, we approximate the errors as below:

$$\begin{aligned}\mathcal{E}(\tau,T,\lambda)=&\frac{\lambda T\Delta+\mathrm{o}(T\Delta)}{\lambda T\Delta}+\\&-\frac{(1+\lambda T\Delta+\mathrm{o}(\lambda T))(\lambda T\Delta+\mathrm{o}(\lambda T))^2}{\lambda^2T^2\Delta+\mathrm{o}(\lambda^2T^2)}+\\&-\frac{(1+\lambda T\Delta+\mathrm{o}(\lambda T))(\lambda T\Delta+\mathrm{o}(\lambda T))^2T(1-\Delta)}{T\Delta[\lambda^2T^2+\mathrm{o}(\lambda^2T^2)]}=\\=&1+\frac{\lambda^2T^2\Delta^2-\mathrm{o}(\lambda^2T^2)}{\lambda^2T^2\Delta+\mathrm{o}(\lambda^2T^2)}+\\&-\frac{[\lambda^2T^2\Delta^2+\mathrm{o}(\lambda^2T^2)](1-\Delta)}{\Delta[\lambda^2T^2\Delta+\mathrm{o}(\lambda^2T^2)]}+\\=&1+\Delta-\frac{\Delta^2(1-\Delta)}{\Delta^2}+\mathrm{o}(1).\end{aligned}$$

From the above, we have that the error goes to 0 when λT is much smaller than 1. ■

Theorem 2: When $\lambda T \ll 1$, the detected intercontact times $\tilde{S}$ follow approximately an exponential distribution with rate $\lambda\Delta$.

Proof: We are going to extend here the proof of Theorem 2 given in the body of the paper, analysing the domains of the MGFs that we previously neglected. In fact, to prove that $\tilde{S}$ can be approximated by an exponential distribution with parameter $\lambda\Delta$, we should demonstrate that for $\lambda\Delta \longrightarrow 0$ both the expressions of the MGFs (i.e. what we proved in the body of the paper) and the domains (i.e what we want to study here) are identical.
Observing the equation of $M_N(s)$ of the proof in the body of the paper, we see that $M_N(s)$ is the result of a power series of functions, and, for this reason, it is defined only in the domain D given by the following:

$$\begin{aligned}D&=\left\{s\in\mathbb{R}:\left|\frac{s\lambda T(1-\Delta)}{e^{\lambda T}-1}\right|<1\right\}=\\&=\left\{s\in\mathbb{R}:|s|<\frac{e^{\lambda T}-1}{\lambda T(1-\Delta)}\right\}.\end{aligned}$$

As $M_{\tilde{S}}(s)=M_N(M_S(s))=M_N\left(\frac{\lambda}{\lambda-s}\right)$, we have that the MGF of $\tilde{S}$ is defined for all $\frac{\lambda}{\lambda-s}\in D$. Observing that $\frac{\lambda}{\lambda-s}$ is a positive number (because $s<\lambda$), $M_{\tilde{S}}$ is defined for all s such that:

$$\frac{\lambda}{\lambda-s}<\frac{e^{\lambda T}-1}{\lambda T(1-\Delta)},\Longleftrightarrow\ s<\lambda-\frac{\lambda^2T(1-\Delta)}{e^{\lambda T}-1}. \quad \text{(A6)}$$

When $\lambda T \ll 1$, the relation (A6) tells us that the domain of $M_{\tilde{S}}$ tends to the domain:

$$\{s\in\mathbb{R}:s<\lambda T\},$$

that is the domain of an exponential random variable with rate $\lambda\Delta$. This completes the proof. ■

Theorem 3: The volume N_Δ of messages delivered by the system under duty cycling Δ is given by:

$$\begin{aligned}N_\Delta\ =\ &\frac{\mu L}{\Delta}-\mu E[D_\Delta]\cdot\\&\cdot\left[1-\frac{1}{2}e^{\frac{L}{E[D_\Delta]\Delta}}\left(e^{-\left(1+\sqrt{\frac{c^2-1}{c^2+1}}\right)}+e^{-\sqrt{\frac{c^2-1}{c^2+1}}}\right)\right],\end{aligned}$$

where c is the coefficient of variation of the delay.

Proof: Every message generated in the network experience a delay that, as we have already seen, has a coefficient of variation that is bigger than 1. For this reason the delay of every message can be approximated with an hyper-exponential random variable with parameters given by the following:

$$\left\{\begin{array}{l}p_1=\frac{1}{2}\left(1+\sqrt{\frac{c^2-1}{c^2+1}}\right)\\ \lambda_1=\frac{2p_1}{E[D_\Delta]}\end{array}\right.\qquad\left\{\begin{array}{l}p_2=1-p_1\\ \lambda_2=\frac{2p_2}{E[D_\Delta]}\end{array}\right. \quad \text{(A7)}$$

Let us denote with T_g the random variable characterising message generation time. If a message is generated at time $T_g=0$, it is delivered if its delay is less than $\frac{L}{\Delta}$. When the message is generated at the time $T_g=t>0$, it arrives at the destination if the delay is less than the remaining time $\frac{L}{\Delta}-t$. For this reason its probability to be delivered is given by

$$\begin{aligned}P\left\{T_g+D_\Delta\leq\frac{L}{\Delta}\middle|T_g=t\right\}=&\\=1-\left(p_1e^{-\lambda_1(\frac{L}{\Delta}-t)}+p_2e^{-\lambda_2(\frac{L}{\Delta}-t)}\right),&\end{aligned}$$

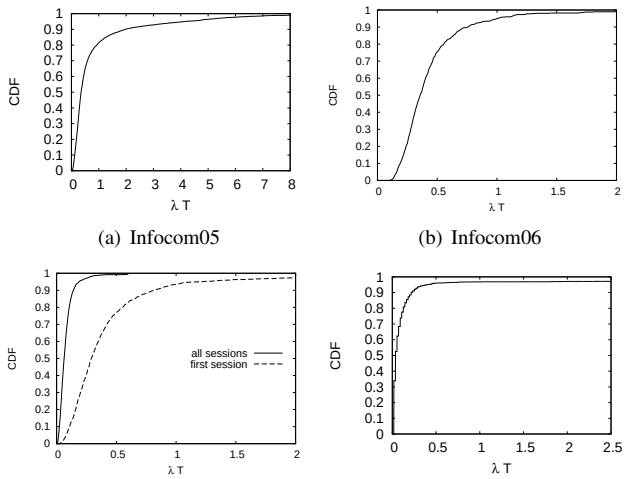
exploiting the CCDF of the hyper-exponential distribution. As in a Poisson process of rate μ, in a time interval dt are generated μdt messages, the total number of messages delivered in the network lifetime $\frac{L}{\Delta}$ is given by:

$$\begin{aligned}N_\Delta=&\int_0^{\frac{L}{\Delta}}P\left\{T_g+D_\Delta\leq\frac{L}{\Delta}\middle|T_g=t\right\}\mu\,dt\\=&\int_0^{\frac{L}{\Delta}}\mu\left[1-\left(p_1e^{-\lambda_1(\frac{L}{\Delta}-t)}+p_2e^{-\lambda_2(\frac{L}{\Delta}-t)}\right)\right]dt\\=&\frac{\mu L}{\Delta}-\mu\left[\frac{p_1}{\lambda_1}\left(1-e^{-\lambda_1\frac{L}{\Delta}}\right)+\frac{p_2}{\lambda_2}\left(1-e^{-\lambda_2\frac{L}{\Delta}}\right)\right]\\=&\frac{\mu L}{\Delta}-\mu\left[E[D_\Delta]-\frac{E[D_\Delta]}{2}\left(e^{-\lambda_1\frac{L}{\Delta}}+e^{-\lambda_2\frac{L}{\Delta}}\right)\right],\end{aligned} \quad \text{(A8)}$$

where we used that $\frac{p_1}{\lambda_1}=\frac{p_2}{\lambda_2}=\frac{E[D_\Delta]}{2}$ from Equation (A7). This concludes the proof. ■

A. Validation: pairwise analysis

In this section we discuss whether the condition $\lambda T \ll 1$ at the basis of our approximation holds true for the pairs of nodes in the datasets that we have studied in Section V-A, where, due to lack of space, we have only performed an average study. In Figure 11 we have plotted the CDF of quantity λT across all pairs in the datasets. Please note that we have only considered contacts between internal devices (i.e., only those that were explicitly taking part to the experiments). For the Reality Mining datasets, we have neglected contacts during the summer break, as done in previous studies [21]. From Figure 11, it is clear that the results that we have obtained for the average case are able to accurately represent what happens at the level of individual pairs. For the RollerNet and Reality Mining experiments, the vast majority ($> 90\%$) of pairs satisfy the condition for the approximation. In Infocom06 there is still a large fraction of pairs satisfying the approximation, while for Infocom05 the approximation seems to hold for $\sim 60\% - 70\%$ of pairs. These pairwise results confirm what we have derived in Section V-A.



(a) Infocom05 (b) Infocom06

(c) RollerNet (d) Reality Mining

Fig. 11. CDF of λT in the reference datasets

The Strength of Vicinity Annexation in Opportunistic Networking

Tiphaine Phe-Neau†, Marcelo Dias de Amorim†, and Vania Conan‡
†CNRS and UPMC Sorbonne Universités {tiphaine.phe-neau, marcelo.amorim}@lip6.fr
‡Thales Communications vania.conan@fr.thalesgroup.com

***Abstract*—Most disruption-tolerant networking protocols available have focused on mere contact and intercontact characteristics to make forwarding decisions. We propose to relax such a simplistic approach and include multi-hop opportunities by annexing a node's vicinity to its network vision. We investigate how the vicinity of a node evolves through time and whether such information is useful when routing data. By analyzing a modified version of the pure WAIT forwarding strategy, we observe a clear tradeoff between routing performance and cost for monitoring the neighborhood. By observing a vicinity-aware WAIT strategy, we emphasize how the pure WAIT misses interesting end-to-end transmission opportunities through nearby nodes. Our analyses also suggest that limiting a node's neighborhood view to four hops is enough to improve forwarding efficiency while keeping control overhead low.**

***Index Terms*—Opportunistic networks, disruption-tolerant networks, contact, intercontact, vicinity.**

I. Introduction

As our urban society lives on, the more technologically nomadic its citizens get. During their daily commuting, people carry electronic devices like smartphones, portable game stations, or laptops. Previsions show how there will likely be more than 1 billion smartphones in 2016 [1]. Common devices embed wireless interfaces and important storage abilities traveling with their owners. The penetration of such technology on our daily life leads to new and increased mobile usages as well as new potential networking paradigm like disruption-tolerant networking (DTN) also known as opportunistic networking [2]. Devices like smartphones are ideal actors of DTN, they leverage people's mobility to carry information toward new places or new persons without requiring any help from an infrastructure, therefore avoiding the payment of a subscription to any provider.

Opportunistic networks rely on user mobility to store and forward information. Unlike usual wired or MANET networks, nodes in opportunistic networks inherently lack global network knowledge, as they are only aware of what they learned via encounters (aka *contacts*) [2]. Routing in DTN is thus challenging by nature. There are two solutions that bound the performance of routing strategies in such networks. On the one hand, the most efficient solution in terms of communication costs consists in waiting until the source meets the destination to transfer the message (at the cost of longer delays) [3]. On the other hand, full epidemic forwarding (flooding) yields the shortest delay but generates the costliest traffic overhead [4]. In between, other solutions such as PRoPHET or Spray-and-Wait employ a wiser strategy [5], [6]. Nodes choose their next hops based on probabilistic likeliness of meeting the destination or through distributed flooding.

All the solutions listed above share a common characteristic – they consider that whenever nodes are not in contact they are in *intercontact*. As we will see later on, this leads to suboptimal results as nodes are likely to miss transfer opportunities when the destination of a message is nearby but not in direct contact. In this paper, *we provide elements toward the adoption of a node's close vicinity as a more appropriate mean to help deliver messages in DTN*. The idea is to leverage short-length, multi-hop paths whenever possible to achieve immediate message delivery while keeping signaling overhead low. The motivation behind our work is that nodes that show interest to communicate are likely to occupy similar geographic areas, even if not within direct communication range [7]. We provide the following contributions:

- We propose and evaluate the interest of extending vicinity knowledge beyond one hop. To this end, we analyze both real-world and synthetic mobility traces.
- We show the perks of using short-length multi-hop paths in the WAIT protocol and validate the strategy of annexing a node's vicinity as a routing asset.
- We investigate if there is any empirical tradeoff in terms of vicinity knowledge that allows better waiting times while constraining monitoring costs.

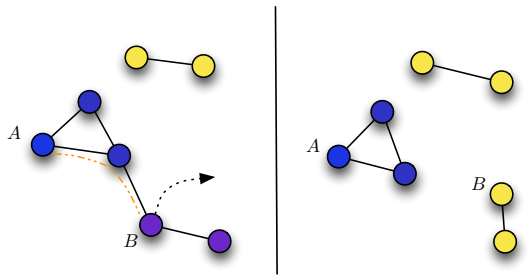
In the remainder of this paper, we first clarify the problem in Section II. Then, we define the notion of vicinity in DTN and enunciate the metrics used for our evaluation in Section III. We analyze the gains of vicinity annexation and its optimal setting in Section IV. In Section V, we list how our work relates to previous analyses and finally, we conclude the paper in Section VI.

II. Contact-based vs. Deeper Vicinity Vision

Currently, DTN protocols deduce their transmission opportunities using contact observations. We consider the traditional definition that a contact happens between two nodes whenever they are within each other's wireless range.

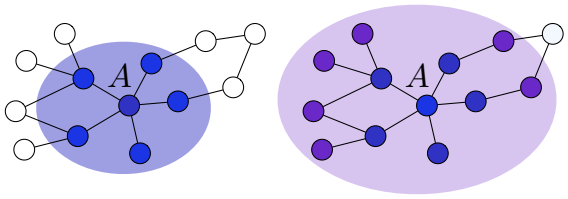
A. The WAIT protocol

In the WAIT protocol, the source stores the message until it meets the destination. The main criticism on this approach, although its minimal communication cost, is that the source may wait for a quite long time before being able to deliver the



(a) t_0 (b) t_1

Fig. 1. A motivating example where A wants to send a message to B. Fig. 1(a): at t_0, nodes A and B are at a 2-hop distance but not in contact. Fig. 1(b): at t_1, B moves away without ever coming in contact with A. A missed the opportunity to send its message while B was nearby. In this situation, vicinity annexation could help A. If A had known B was so close, it could have used the existing end-to-end path.



(a) 1-vicinity (b) 2-vicinity

Fig. 2. Node A's κ-vicinity illustration, here, $\kappa = \{1, 2\}$. In Fig. 2(a), we represented node A's 1-vicinity. When $\kappa = 1$, A only knows nodes in contact with him. In Fig. 2(b), we have node A's 2-vicinity ($\kappa = 2$). A knows all nodes within a 2-hop distance and so on. A has an end-to-end path, of at most length κ, to any member of its κ-vicinity.

message or, worst, to completely fail delivering it. As we will see in Section IV, we observe reduction of delivery delays of up to 80% in average by extending of only one hop the vicinity knowledge; in some cases, delays can be reduced by several hours. This means that a simple variation of the WAIT protocol can be now applied in contexts that could not be considered previously. The WAIT protocol is also known as Direct Transmission.

B. The pros and cons of monitoring contacts only

In a node-centric scheme, gathering contact information comes naturally. Any device can sense its surrounding and feel whenever other nodes are around through appropriate probing tactics. The application of contact knowledge in the WAIT protocol is straightforward: the source delivers a content to the destination when they get in contact with each other. In practice, a node may be collocated with many other nodes but it is not always or ever in direct contact with some of them. An example is illustrated in Fig. 1. At a given time t_0, A is at a 2-hop distance from B. Then at time t_1, B decides to leave and breaks any existing contemporaneous path from A to B. Let us consider that A knew the network topology at t_0. If A had known the path to destination B at t_0, it could have considered sending the message to B using multiple hops (in this example, only two), instead of waiting until meeting B (or trying some other non-deterministic strategy), which might take forever.

PRoPHET uses delivery predictabilities (preds) to evaluate the utility of using one node as a relay toward the destination. The protocol updates these preds only when the node comes in contact with other nodes and ages them as long as they move away (even if nearby). PRoPHET also uses a transitivity property to update its preds but mellows its impact via a β factor that can be quite low. Moreover, it does not give any freshness insights on this "transitive proximity". Concerning Spray-and-Wait, the distributed flooding approach sends L copies of a message to L different nodes, in somehow a "first in contact first served" fashion. As in the case of PRoPHET, Spray-and-Wait only uses knowledge about nodes that get in direct contact, without considering vicinity information beyond one hop that could be helpful.

C. Why not taking a look around?

DTN protocols rely on contacts as they are easy to gather, while extended neighborhood knowledge is more costly. Due to the DTN nature, offering nodes a consistent and full knowledge of the network topology is unrealistic. An alternative would be to make nodes have information on the connected component they are in. This would allow each node knowing with whom it has a contemporaneous path. However, as nodes do not know a priori the size of the connected component, it becomes difficult to limit the control overhead. Still, by limiting the scope of a node vision (to nodes up to a few hops, say two or three), we also limit the signaling overhead. Localized vicinity knowledge can be an important asset for DTN nodes. However, the main challenge in providing extended vicinity knowledge to nodes is how to find a good balance between efficiency and probing costs. In fact, we can wonder how far a node should probe its vicinity for surrounding knowledge. The more information about the network, the better decisions we can make but the higher the costs induced.

III. Waiting less with vicinity knowledge

We evaluate how the vicinity knowledge can help improve the responsiveness of a routing scheme for DTN. In this paper, we consider the particular case of the WAIT protocol because of its specific properties: upper bound in terms of delay and lower bound in terms of signaling overhead. We leave the evaluation with other protocols for future work.

A. The κ-vicinity

We define how nodes see their vicinity with the κ-vicinity. We assume that nodes monitor the presence of other nodes up to a distance of κ hops [7]:

Definition 1. ***κ-vicinity.*** *The κ-vicinity of a node A, noted κ_A, is the set of nodes with a contemporaneous end-to-end path of at most κ hops to A. All other nodes are considered outside κ_A.*

In Fig. 2, we show an example of node A's κ-vicinity for $\kappa = 1$ and $\kappa = 2$. The traditional definition of a contact

corresponds to the case of $\kappa = 1$. Now, whenever node A wants to send a message to node B, A scans its neighborhood up to κ hops. If B arrives within A's κ-vicinity, A can send the message to B via a multi-hop contemporaneous path, avoiding unnecessary waiting periods.

B. The WAIT protocol with extended knowledge

To observe the impact of vicinity awareness, we investigate the effects of neighborhood knowledge in the WAIT protocol by focusing on the waiting parameter – the time a node waits before being able to send a message straight to the destination. The secret, which is application-dependent, is how to balance delay and signaling cost. Recall that in the case of the WAIT protocol, the waiting time only stops when the source meets the destination (no intermediate relays), which is the worst case in terms of delay. The waiting time has a particular meaning as it is related to the user experience – any user notices the duration before its message gets delivered (if it gets delivered at all) and judges a service accordingly. Instead of analyzing delivery percentage, we chose to focus on waiting delays which is a direct representation of neighborhood observation raw advantages.

C. Costs

To take into account the costs of multi-hop messaging and neighborhood monitoring, we identified two main sources of overhead. We use the message as the unit of comparison.

Data Overhead (D_o): represents the total cost to deliver a message. Clearly, any protocol with extended neighborhood knowledge is costlier than its simple version. Whenever the source switches to multi-hop transmission mode, the message follows a contemporaneous end-to-end path to the destination and has to sustain several store and forward processes. The "extra" cost of such a communication, in terms of additional messages sent, is the number n of hops between A and B minus one:

$$D_o = n - 1. \tag{1}$$

Neighborhood Knowledge Overhead (N_o): represents the signaling overhead to gather information about the neighborhood. Node A broadcasts a discovery message (`DM`) to its contacts with a TTL set to κ. All nodes receiving the `DM` rebroadcast this message with a TTL set to $\kappa - 1$, and so on. We assume that each transmission is acknowledged (see Fig. 3 for a detailed example). This leads to a cost of:

$$N_o = ||\kappa_A|| + ||(\kappa - 1)_A|| + 1, \tag{2}$$

where $||\cdot||$ stands for cardinality. N_o does not depend on the path length that `DM`s have to cross. With little aggregation, N_o only depends on the number of neighbors in a node's connected component.

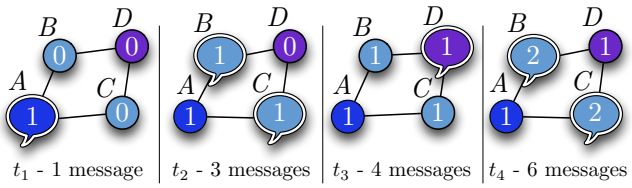


Fig. 3. Neighborhood knowledge discovery technique. At t_1, A ignites the discovery by broadcasting a message with a TTL set to 2. Its contacts, B and C receive the message. At t_2, they broadcast a message with a TTL set to $2 - 1 = 1$. At t_3, D received discovery messages with a TTL of 1, then broadcasts its reply. At t_4, B and C aggregate all replies they received and send their knowledge to A. In the end, we obtain 6 sent messages.

IV. Performance analysis

To evaluate the performance gains enabled by neighborhood awareness, we simulate the WAIT protocol for different values of κ (recall that $\kappa = 1$ corresponds to the basic WAIT protocol) using various mobility data.

A. Datasets

We performed our study with a wide range of datasets both synthetic and real-life based. Real-life experiments all involved devices (iMotes, T-Motes) carried by participants. These Motes logged the presence of other devices within a 10-meter range unless specified. Researchers derive contact intervals from these presence logs. We chose the following scenari captured during various experiments. As all these experiments are available to the community, we also provide the downloading links.

- **Infocom05** measurement was held during the Infocom 2005 conference [8], [9]. Researchers gave iMotes to 41 attendees. We study a 12-hour interval with the highest activity. Each iMote probes its environment every 120 seconds. *Infocom05* represents conference environment.
- **Rollernet** had 62 participants during a 3-hour dominical rollerblading tour in Paris [10], [11]. Motes probed their surrounding every 30 seconds. This dataset illustrates a dynamic sport event.
- **Unimi** involved 48 people among the students, faculty members, and staff from the University of Milano for two weeks in 2008 [12], [13]. Devices probed their surroundings every second. *Unimi* provides a longer academic and working situation.

We also generated scenari from two well-known mobility models:

- **Random Trip** is an improved Random-Waypoint mobility model [14], [15]. We generated the patterns of 20 nodes on a surface of 50 x 60 m^2 with speed between 0 and 7 m/s.
- **Community** is a mobility model reflecting human tendencies to aggregate in specific locations [16], [17]. We created 50 nodes with a 10m wireless range on a 1,500 x 2,500 m^2 plane during 9 hours.

TABLE I
AVERAGE NUMBER OF NEIGHBORS IN A NODE'S κ-VICINITY (WHOLE DATASET DURATION).

	κ							
Dataset	1	2	3	4	5	6	7	8+
Community	2.0	4.0	4.6	**4.7**	4.7	4.7	4.7	4.7
RandomTrip	2.0	3.2	4.7	**5.7**	6.3	6.7	6.9	7.1
Infocom05	1.5	3.8	5.3	**6.0**	6.4	6.4	6.4	6.4
Rollernet	1.4	3.2	4.7	**5.7**	6.3	6.7	6.9	7.0
Unimi	0.3	0.7	0.9	**1.1**	1.1	1.2	1.2	1.2

TABLE II
NEIGHBORS κ-DISTRIBUTION IN A NODE'S κ-VICINITY.

	κ							
Dataset	1	2	3	4	5	6	7	8+
Community	2.4	**2.3**	0.7	0.1	0.0	0.0	0.0	0.0
RandomTrip	2.3	2.3	2.0	**1.4**	0.8	0.4	0.2	0.1
Infocom05	3.0	4.4	3.0	**1.4**	0.7	0.2	0.1	0.0
Rollernet	2.0	2.5	2.1	**1.5**	0.9	0.6	0.3	0.2
Unimi	1.5	**1.0**	0.7	0.4	0.2	0.1	0.0	0.0

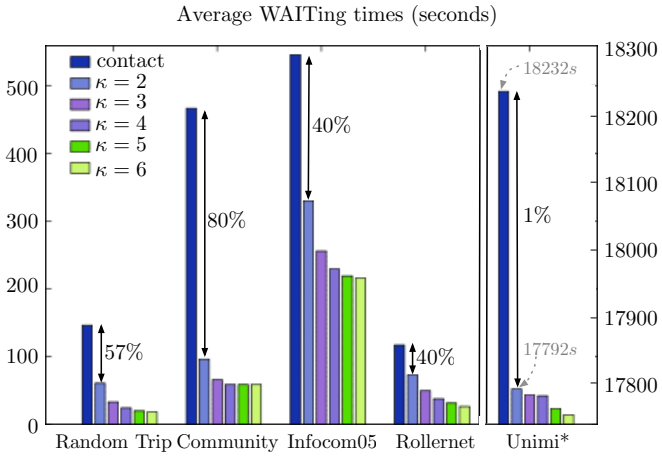


Fig. 4. Averaged waiting times according to the threshold κ. For all traces, there is a clear improvement between the first and second bar (contact only vs. 2-vicinity). Being aware of a node's κ-vicinity can lead to divide waiting times by 4 like in *Community*. The higher the κ, the better the waiting delays, yet, above $\kappa > 4$, gains become negligible. Note that, for the *Unimi** dataset we focused on its top values. The average waiting time in contact is 18,232 seconds while in the 2-vicinity, it is 17,792 seconds. These high values come from the dataset length (two weeks).

B. Threshold Optimization

Neighborhood monitoring is an expensive process in opportunistic networks. To lower its costs, we investigate the optimal κ threshold. First, we consider the amount of people needed in the vicinity then we observe the difference between a static and a dynamic κ setting.

For each node, we analyze the average number of neighbors in their κ-vicinity. Table I shows this value for the whole dataset duration. We understand that above a certain threshold κ_t, a node's κ-vicinity does not expend much (except for the *RandomTrip* dataset, which has a random movement pattern and a high density). In *Community* or *Infocom05*, a node's κ-vicinity does not grow significantly anymore above $\kappa_t = 4$. The same phenomenon appears with *Unimi* but with lower figures. The *Unimi* dataset is longer (two weeks) than other datasets. As we chose to analyze the average number of neighbors for the whole experiment duration, *Unimi*'s length lowered the expected average number of nodes.

In the next table, we focused on instants where nodes had at least one close neighbor. For each dataset, we analyzed all nodes' inner κ-vicinity distribution. Whenever a node had at least one neighbor, we observed the average number of neighbors located at a κ-hop distance (see Table II). In *Infocom05*, we see that on average within a non void κ-vicinity, a node could find 3.0 nodes in contact, 4.4 at 2 hops, 3.0 at 3 hop, 1.4 at 4 hops etc. For *Community* and *Unimi*, the number of nodes at $\kappa > 2$ falls below 1. For *RandomTrip*, *Rollernet* and *Infocom05*, the fall occurs after $\kappa = 4$. Above the threshold $\kappa_t = 4$, there will rarely be nodes at higher distances. Moreover, these distributions are linked to each datasets average diameter. All datasets average distributions are concentrated on shorter distance with $\kappa \leq 4$. For instance, *Community* does not have components larger than 4-hop distance. *Unimi* has components of at most 6-hop length. But within *Unimi*'s components, most neighbors appear in contact or at a 2-hop distance. The κ_t represents a high enough threshold so as to capture most of a node's surroundings. As a result, one would conclude that setting up a 4-vicinity monitoring for each node is optimal.

C. Routing Effects

1) Loss & Delays: For each mobility trace and each pair of nodes, we randomly generated 10 messages at different time instants. We chose to generate sparse messages for waiting times to better reflect the impact of neighborhood monitoring. The most symptomatic situation arises when a pair of nodes never come into contact, but once and a while they belong to the same connected component. In this situation, the WAIT protocol drops the message whereas the neighborhood-aware variant can manage to forward it correctly.

As scarce as this situation may sound, it happens for 10% of pair of nodes in *Infocom05*, 12% of *Unimi* nodes, 53% in *Community*, and around 55% of *Rollernet* nodes. If these nodes try to send a message using the WAIT protocol, they will simply fail. These fractions of nodes have *infinite waiting delays* when WAIT is in use. Otherwise, with the neighborhood-friendly version, they manage to deliver messages with bounded waiting times.

For these nodes with *bounded waiting delays*, we analyze to which extent neighborhood knowledge helps lower their waiting times. In Fig. 4, we show the averaged pairwise waiting times for each dataset. Each bar represents the average waiting delay we obtain with κ-vicinity probing. For every dataset, between the first and second bars, we notice significant reduction in the waiting times: 40% in *Infocom05* and *Rollernet*, 57% in *RandomTrip*, and around 80% in *Community*. The *Unimi* dataset stands out because of its time scale. The

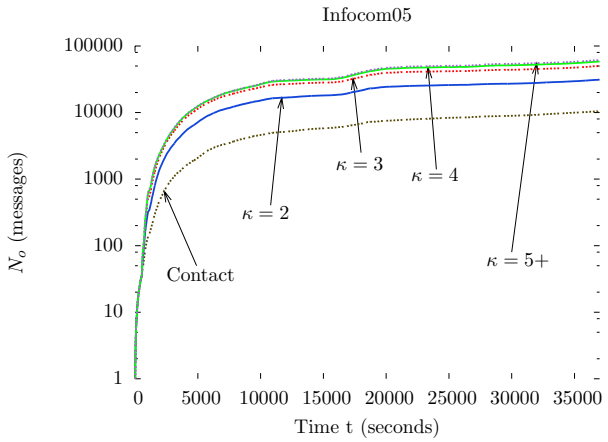


Fig. 5. Neighborhood Knowledge Overhead (N_o) in terms of message sent by the discovery technique `Reg` for a node in the *Infocom05* dataset. On average, probing κ-vicinity with $\kappa > 4$ costs as much as probing the 4-vicinity. This version of neighborhood probing is very expensive. Note the logscale on the y-axis.

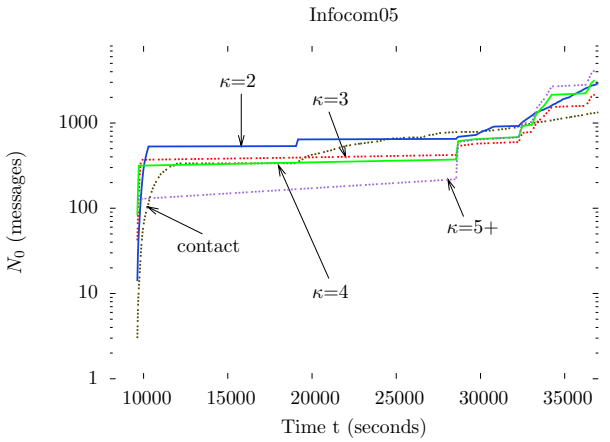


Fig. 6. Neighborhood Knowledge Overhead (N_o) using `OnD` for a pair of node in *Infocom05*. Contact monitoring drops $9/10^{th}$ of messages and keeps monitoring its contacts without being able to deliver any messages. For 7 delivered messages, sensing 3-vicinity (or beyond) ends up cheaper than observing 2-vicinity. $\kappa \geq 3$ leads to shorter waiting delays and shorter probing periods than with 2-vicinity. Note the logscale on the y-axis.

experiment lasting two weeks, the random message generation process may choose values during weekends or nights. Even though, the relative difference between the first and second bar is more limited, the time reduction is still present between the first two bars.

For all datasets we observe that, although we keep reducing the waiting delays, the gains for $\kappa > 4$ are much smaller. This corroborates our first feeling that localized knowledge should be enough and suggests that we can, in practice, keep κ small.

2) Overheads: Supporting vicinity knowledge monitoring does not come for free. Any node needs to probe its vicinity and create a flow of messages around.

Impact of neighborhood knowledge overhead. There are many strategies for connected component gathering, from link state-like solutions to flooding techniques. For our study, we chose to compare two naive behaviors:

- Nodes keep monitoring their κ-vicinity at regular time intervals (called `Reg` hereafter).
- Nodes monitor their κ-vicinity when they have a message to send and stop when it expires (called `OnD` for "On Demand").

With `Reg` probing every 30 seconds, we quantified the volume of generated messages for different values of κ. Monitoring only contacts induces fewer overheads than any deeper neighborhood monitoring. For $\kappa = \{2, 3\}$, we have larger volumes of N_o. Beyond $\kappa = 4$, there are no noticeable differences for N_o. Overall behaviors are quite alike and depend on the surrounding density.

In Fig. 6, we plot N_o of the same source node as before. This time, we use the `OnD` method for neighborhood analysis. The reason we have noticeable jumps in all curves is, when the destination comes into the source's κ-vicinity, this latter stops monitoring its surroundings. Contact monitoring drops all but one message and is only plotted for the reader's information. As a result, the `OnD` technique appears more efficient than the naive `Reg`. In Fig. 6, we see how N_o evolves with time. With a simple probing technique `OnD`, we manage to constrain message overheads and deliver more messages than with the WAIT protocol. Also, an interesting result is how, for the same number of delivered messages (7), probing the 3-vicinity and beyond gives better results than probing only the 2-vicinity in terms of N_o. The faster the source finds the destination, the shorter the waiting delay and the lower the N_o.

Impact of data overhead. N_o seems to be the most expensive in terms of messages sent, yet, we also have to consider D_o (i.e., the number of messages over an end-to-end path). D_o adds an insignificant number of messages to N_o. It is important to underline that having a large D_o (i.e., a long path between the sender and the destination) can lead to undelivered messages. This is why one would prefer smaller κ.

V. Related and Future Work

Concerning DTNs, researchers have found various ways to leverage a node's neighborhood. Some techniques choose to use the social behavior of the participants. As in a city people tend to cluster into communities around different points of interests, Ött *et al.* presented a protocol leveraging end-to-end and multi-hop DTN paths [18]. End-to-end paths occur among connected components whereas DTN ones happen between these temporary components. Sarafijanovic-Djukic *et al.* made a similar observation in the VANET environment [19]. Later, Heimlicher and Salamatian laid their study over the groundwork that mobile wireless networks tend to have connected crowds [20]. The main punch line for all these studies is: for each node, there are immediate neighborhood structures to use.

Other analyses preferred another point of view towards the vicinity in DTN. Instead of considering a node's instantaneous

vicinity, they considered a node's "reachable" vicinity in terms of nodes we can reach during a given time window. Chaintreau *et al.* analyzed spatio-temporal clusters diameter in a network [21]. Tang *et al.* focused on the nature of these spatio-temporal paths to better understand how to use them [22]. Whitbeck *et al.* proposed an interesting way to capture a node's reachable vicinity through a new graph type [23].

In our work, the κ-vicinity reflects the existing topology and uses it to lower transmission delays. The κ-vicinity leverages immediate neighborhood structure further than techniques mentioned in the first paragraph and considers reachability in a node's connected component. We do no wait for potential contacts appearing later, we use existing links beyond mere contacts. As an additional example of use, in their latest paper Diana *et al.* applied a similar vicinity notion to satellite communications [24]. By leveraging neighboring stations, their proposal allowed valuable routing performance gains.

In Section IV-B, the recommended threshold of $\kappa = 4$ is static. Deploying a strategy with a dynamic κ threshold may be another alternative. As seen in Section IV-C, adapting your vicinity vision according to your needs allows considerable overhead gains. Determining the accurate policy for vicinity sensing may be linked to applications needs as well as performance requirements. As a future work, we would like to investigate the relationship between vicinity probing policies and local densities to provide

VI. Conclusion

In this paper, we examine the impact of vicinity awareness on the waiting time in a variation of the pure WAIT protocol. Most DTN techniques only focus on sensing direct contacts and do not inquire about their nearby neighbors. This strategy is the most straightforward and rational. However, we cannot deny the sociological nature of DTNs. People do not wander randomly in a city. They gather around specific persons or locations. To our opinion, ignoring a node's immediate neighborhood results in a loss of useful information.

Our findings show that neighborhood probing significantly improves performances of the WAIT protocol in terms of waiting delays. When delays used to be infinite, they are now bounded. When delays were high, they are now lowered by a factor up to four in our scenarios. Yet, neighborhood monitoring ignites messaging overhead. But, by limiting a node's vicinity knowledge to a threshold κ of four is an optimal setting. We constrain costs and still enhance performance. According to our observations, we have now enough motivation to consider potential performance gains that vicinity knowledge could bring to more elaborated DTN schemes.

Acknowledgment

This work is partially supported by the European Commission in the framework of the FP7 Mobile Opportunistic Traffic Offloading (MOTO) under grant agreement number 317959.

References

[1] Forrester Research, "Mobile is The New Face of engagement," 2012, http://www.forrester.com/go?objectid=RES60544.

[2] K. Fall, "A Delay-Tolerant Network Architecture for Challenged Internets," in *ACM Sigcomm*, Karlsruhe, Germany, Aug. 2003.

[3] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: the single-copy case," *IEEE/ACM Transactions on Networking*, vol. 16, no. 1, pp. 63–76, Feb. 2008.

[4] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Duke University, Tech. Rep., 2000.

[5] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic routing in intermittently connected networks," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 7, July 2003.

[6] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Spray and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks," in *ACM SIGCOMM workshop on Delay-Tolerant Networking*, Philadelphia, Pennsylvania, USA, Aug. 2005.

[7] T. Phe-Neau, M. Dias de Amorim, and V. Conan, "Fine-Grained Intercontact Characterization in Disruption-Tolerant Networks," in *IEEE Symposium on Computers and Communication*, Kerkyra, Greece, Jun. 2011.

[8] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 606–620, 2007.

[9] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, "CRAWDAD data set cambridge/haggle (v. 2009-05-29)," Downloaded from http://crawdad.cs.dartmouth.edu/cambridge/haggle, May 2009.

[10] P.-U. Tournoux, J. Leguay, F. Benbadis, J. Whitbeck, V. Conan, and M. D. de Amorim, "Density-aware routing in highly dynamic DTNs: The rollernet case," *IEEE Transactions on Mobile Computing*, vol. 10, pp. 1755–1768, 2011.

[11] J. Leguay and F. Benbadis, "CRAWDAD data set upmc/rollernet (v. 2009-02-02)," Downloaded from http://crawdad.cs.dartmouth.edu/upmc/rollernet, Feb. 2009.

[12] S. Gaito, E. Pagani, and G. P. Rossi, "Fine-Grained Tracking of Human Mobility in Dense Scenarios," in *IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, Rome, Italy, Jun. 2009.

[13] P. Meroni, S. Gaito, E. Pagani, and G. P. Rossi, "CRAWDAD data set unimi/pmtr (v. 2008-12-01)," Downloaded from http://crawdad.cs.dartmouth.edu/unimi/pmtr, Dec. 2008.

[14] S. PalChaudhuri, J.-Y. Le Boudec, and M. Vojnovic, "Perfect Simulations for Random Trip Mobility Models," in *IEEE Infocom*, Miami, Florida, USA, Aug. 2005.

[15] S. Pal Chaudhuri, J.-Y. Le Boudec, and M. Vojnovic, "The random trip mobility model," http://icawww1.epfl.ch/RandomTrip/, Apr. 2005.

[16] M. Musolesi and C. Mascolo, "Designing mobility models based on social network theory," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 11, pp. 59–70, July 2007.

[17] ——, "Community based mobility model," http://www.cl.cam.ac.uk/research/srg/netos/mobilitymodels/, May 2006.

[18] J. Ott, D. Kutscher, and C. Dwertmann, "Integrating DTN and MANET routing," in *ACM MobiCom workshop on Challenged Networks*, Pisa, Italy, Sep. 2006.

[19] N. Sarafijanovic-Djukic, M. Pidrkowski, and M. Grossglauser, "Island Hopping: Efficient Mobility-Assisted Forwarding in Partitioned Networks," in *IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, Reston, VA, USA, Sep. 2006.

[20] S. Heimlicher and K. Salamatian, "Globs in the Primordial Soup – The Emergence of Connected Crowds in Mobile Wireless Networks," in *ACM Mobihoc*, Chicago, Illinois, USA, Sep. 2010.

[21] A. Chaintreau, A. Mtibaa, L. Massoulie, and C. Diot, "The Diameter of Opportunistic Mobile Networks," in *ACM Conext*, New York, NY, USA, Dec. 2007.

[22] J. Tang, M. Musolesi, C. Mascolo, and V. Latora, "Characterising Temporal Distance and Reachability in Mobile and Online Social Networks," *SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 1, pp. 118–124, January 2010.

[23] J. Whitbeck, M. Dias de Amorim, V. Conan, and J.-L. Guillaume, "Temporal Reachability Graphs," in *ACM Mobicom*, Istanbul, Turkey, Aug. 2012.

[24] R. Diana, E. Lochin, L. Franck, C. Baudoin, E. Dubois, and P. Gelard, "A DTN routing scheme for LEO satellites topology," in *IEEE Vehicular Technology Conference*, Québec, Canada, Sep. 2012.

Beyond Contact Predictions in Mobile Opportunistic Networks

Alexandru Tatar*, Tiphaine Phe-Neau*, Marcelo Dias de Amorim*◇, Vania Conan†, and Serge Fdida*

*Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France ◇ CNRS †Thales Communications & Security

***Abstract*—When studying and designing protocols for mobile opportunistic networks, most works consider only direct contact patterns between mobile nodes. Tracking these contacts is important for end-to-end communications but relying only on this kind of information provides a limited view about transmission possibilities. Mobile users are often in intercontact, but still separated by only a few hops, which translate into effective communication opportunities between nodes. In this paper, we focus on such a type of communication opportunities and investigate to what extent they can be predicted. Using real-world datasets, we provide evidences about the predictable nature of nodes' proximity and evaluate the benefits of these results compared to direct contact predictions.**

***Index Terms*—Opportunistic networks, intermittent connections, support vector machine, link prediction.**

I. Introduction

The delay-tolerant network (DTN) paradigm has developed over the last few years [1]. Its growth comes from the democratization of recent mobile technologies like smartphones, laptops, tablets, or game stations. These devices allow each person to carry wireless technologies and enables them to collect and disseminate data on the go. During the day, people encounter each other at random locations in the street, the public transportation system or at work. By coming close to one another, these persons are able to transmit information to each other via short range wireless technologies like Wi-Fi Direct or Bluetooth.

The design of efficient communication protocols in DTNs depends in great part on the capacity to understand and predict human mobility patterns. Thus, over the last years, several studies have revealed important insights about contact durations between mobile users [2], the periodicity of human encounters [3], or the network structures created by human meeting patterns [4]. Uncovering these mobility patterns can then be used to design measures that facilitate the prediction of contacts between nodes. This includes the use of frequency of contacts to identify similarities in mobility characteristics [5], or in finding strongly-connected mobile users that could serve as message carriers [6]. While these metrics are used as good heuristics to the contact prediction problem, they provide only a limited view on the future contact opportunities. A more advantageous but laborious approach to this problem is to predict the contact patterns.

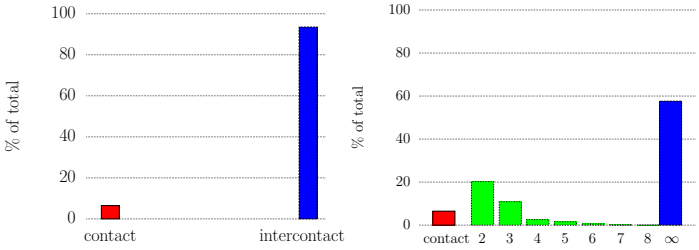
978-1-4799-4937-3/14/$31.00

Recent studies have addressed the problem of contact prediction – predict if two nodes are going to be in direct transmission range – and have revealed that, under the right prediction method and predictive features, contacts between mobile users are to a certain extent predictable [7]. This result is valuable as it allows one to predict the evolution of the network of human interactions that can be used to design more effective DTN communication protocols.

But contacts between nodes are not the only type of relationship between mobile users. Often, users may find themselves not in direct transmission range but still in the nearby vicinity. Thus, to have a broader view on the available communication opportunities, the extended notion of contact, namely κ-contact, has recently been proposed [8]. Previous analyses showed that considering only contacts between nodes ends up in a biased suboptimal network understanding while studying κ-contacts provides a more complete understanding on the available end-to-end communication opportunities.

In this work, we study the predictability of extended contact opportunities in DTNs. Using data from three human-based contact traces, we show that κ-contact opportunities are more predictable than direct contact relationships. To measure the possible impact of this finding in a real-world application, we propose an experimental setting that supports the idea that κ-contact prediction has an interesting potential usage. Our contributions in this paper can be summarized as follows:

- We provide insights on the κ-contacts relationships between mobile nodes and show that considering only direct contacts covers a limited part of the end-to-end transmission possibilities. We reveal the instability of periods of time nodes stay at the same distance, and that κ-contact intervals display better predictability characteristics: intervals frequency and length.
- Using a supervised prediction framework, we study the predictive nature of κ-contacts and compare it with the traditional case of predicting contacts between nodes. Our results indicate that, in highly dynamic mobile settings (e.g. rollerblading scenario), predicting that nodes will remain at a distance of at most two hops from one another, can attain twice the performance of a direct contact prediction.
- Through simulations, we evaluate the impact of κ-contact prediction in a service that would benefit from predicting contacts between mobile users. The experimental results



(a) Current vision. (b) Vicinity-aware vision.

Fig. 1. *Sig09* example: current vision versus vicinity awareness.

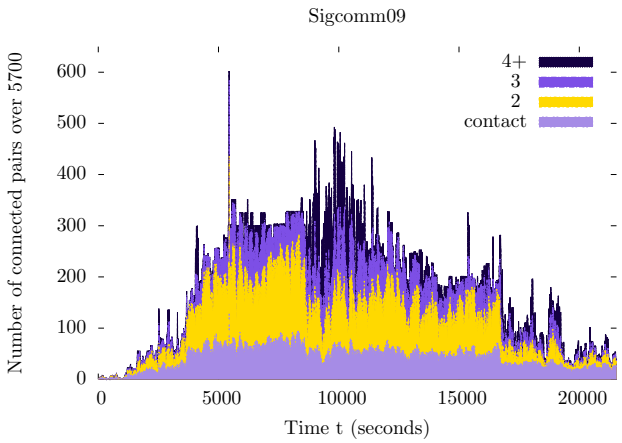


Fig. 2. *Sig09* end-to-end transmission opportunities.

show that there is a higher potential on relying on κ-contact prediction compared to the traditional contact case.

The remainder of this paper is structured as follows: Section II details the κ-vicinity notion as well as κ-contact and the datasets used in our study. Sections III exposes interesting facts about κ-contacts relationships while Section IV describes the prediction framework and reports the prediction performance. Finally, we link our work to the existing literature in Section VI and conclude our study in Section VII.

II. Vicinity and Datasets

A. Is contact enough?

Current DTN approaches only consider network knowledge coming from nodes in contact. While this approach may be simple, it has proven to be efficient in making decisions to forward data in DTNs. However, we realize that there is more at hand than simple contact information. In Fig. 1, we represent some interesting facts about two nodes from *Sig09* dataset (see Section II-C for more details on this dataset).

In Fig. 1(a), we plot the proportion of time nodes spend in contact and in intercontact (not in contact) according to the traditional approach. They stay around 6% of the time in contact and the remaining 94% in intercontact. However, when we observe the same situation under a vicinity-aware point of view (see Fig. 1(b)), we see that these nodes stay 6%

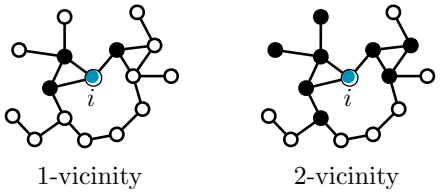


Fig. 3. Example of κ-vicinity. The 1-vicinity consists in only node i's contacts (1-hop distance). The 2-vicinity consists in all i neighbor's whose shortest distance is inferior to 2 hops.

of their time in contact but they also remain at a 2-hop distance around 20% of the time and at a 3-hop distance around 10% of the time. The time that the nodes spend without any end-to-end path linking them (∞) is only 57% of the experiment duration; far below the 94% intercontact time illustrated in Fig. 1(a). Considering only the time spent at a 2-hop distance improves our understanding of the duration that the two nodes are linked by an end-to-end path by a factor of 3.

At a network-wide level, we observe that contacts also represent a minor part of end-to-end transmission opportunities. In Fig. 2, we represent the number of connected pairs by their shortest distances for the beginning of the *Sig09* dataset. The bottom layer indicates the number of nodes that are in contact, the yellow layer shows nodes connected by 2-hop paths and so on. It becomes clear that most end-to-end opportunities come from 2-hop paths and not from contacts. Therefore, ignoring such opportunities results in a waste of connectivity assets.

B. κ-vicinity, κ-contact, and κ-intercontact

To characterize a vicinity in DTN, we use the concept of κ-vicinity [9]. We discriminate a node i's vicinity according to the number of hops between i and its surrounding neighbors. Note that in our definition, we assume that connectivity is bidirectional, as a result κ-vicinity relationships are symmetric.

Definition 1. ***κ-vicinity.*** *The κ-vicinity $\mathcal{V}^i_\kappa$ of node i is the set of nodes with shortest paths of length at most κ hops from i.*

Clearly, $\mathcal{V}^i_{\kappa-1} \subset \mathcal{V}^i_\kappa$. In Fig. 3, we illustrate the 1-vicinity and 2-vicinity for node i. The κ-vicinity's relevancy comes from the proximity of nodes. The shortest distance is one of the easiest characteristics to gather in opportunistic networks. We need to extend the notion of contact and intercontact to the κ-vicinity as from now on, they represent our point of view in the network.

Definition 2. ***κ-contact.*** *Two nodes are in κ-contact when they dwell within each other's κ-vicinity, with $\kappa \in \mathbb{N}^*$. More formally, two nodes i and j are in κ-contact when $\{i \in \mathcal{V}^j_\kappa\} \equiv \{j \in \mathcal{V}^i_\kappa\}$. In other words, a contemporaneous path of length at most κ hops links i and j. Note that, 1-contact represents mere contact.*

Definition 3. ***κ-intercontact.*** *Two nodes are in κ-intercontact when they do not belong to each other's κ-vicinity (there is*

TABLE I
DATASETS CHARACTERISTICS.

Dataset	*#*	*Duration*	*Probing*	*Type*
Infocom05	41	12h	120s	Conference
Sig09	76	1 day	120s	Conference
Rollernet	61	1h30	15s	Sport

TABLE II
DISTANCE AVERAGE DURATION (IN SECONDS).

	κ						
Dataset	1	2	3	4	5	6	7
Infocom05	399	296	224	175	131	154	212
Sig09	149	83	41	25	18	13	11
Rollernet	48	65	76	89	105	114	129

no path of length κ or less linking the two nodes). Note that, 1-intercontact represents simple intercontact.

C. Datasets

We consider several real-world contact traces throughout our experiments.

Infocom05 measurement was held during a 5 days conference in 2005 [2]. 41 attendees carried iMotes collecting information about other iMotes nearby within a 10m wireless range. We study a 12-hour interval bearing the highest networking activity. Each iMote probes its environment every 120 seconds. *Infocom05* represents a professional meeting framework.

Sig09 was taken during the first day of a conference in Barcelona [10]. The experiment recorded 76 users relationships using Bluetooth-based smartphones. Each phone logged contacts every 120 seconds.

Rollernet had 62 participants measuring their mutual connectivity with iMotes during a 1 hour and a half rollerblading tour in Paris [11]. These iMotes sent beacons every 15 seconds. This experiment shows a specific sport gathering scenario.

In Table I, we recapitulate all datasets characteristics. *#* is the number of participating nodes. *Duration* indicates the dataset duration. *Probing* shows the probing intervals of the measuring devices.

III. PAIRWISE INTERACTIONS UNDER THE κ-CONTACT CASE

Given the new definitions of contact and intercontact we analyze different characteristics of the pairwise interactions. For more detailed information concerning κ-contact and κ-intercontact, please refer to [9].

A. Pairwise minimum distance

We begin by studying the pairwise minimum distance, i.e., how close nodes come to each other during the duration of a trace. For instance, if two nodes meet at least once, we mark this distance as 1. If they come as close as 3 hops, we consider the minimum distance to be 3. For nodes that never come in κ-contact, we consider this distance as ∞.

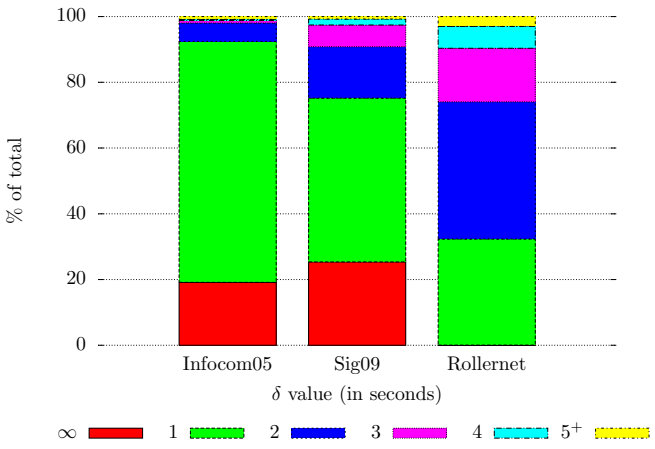


Fig. 4. Pairwise minimum distance for *Infocom05*, *Sig09*, and *Rollernet*.

TABLE III
κ-CONTACT AVERAGE DURATION (IN SECONDS).

	κ						
Dataset	1	2	3	4	5	6	7
Infocom05	399	322	274	247	230	224	224
Sig09	149	101	72	60	54	51	50
Rollernet	48	61	68	75	81	86	90

We represent the results in Fig. 4. In terms of pairs of nodes that come in direct contact, we observe that in conference settings, characterized by a high number of nodes in restricted physical spaces, the number of connected pairs is reasonably high: 49% for *Sig09* and 73% for *Infocom05*. *Rollernet* on the other hand shows a lower network connectivity, with only 33% of nodes coming in a direct contact. But the analysis of contact alone yields an incomplete picture as there is a considerable amount of nodes who come close to each other but never in direct contact. For example, the percentage of pairs that come at a distance of 2 is 5% for *Infocom05*, 16% for *Sig09*, and 41% for *Rollernet*. For *Rollernet* the percent of nodes that come at a 2-hops distance is even higher than the nodes that come in direct contact and one can observe that a non negligible amount of nodes advance up to a distance 3 (6%) and 4 (16%).

B. Average distance duration

In Table II, we present the average duration of an interval during which nodes remain at a distance of κ-hops from one another. For *Infocom05* and *Sig09*, we observe that close connections are more stable, with smaller average durations as the distance between nodes increases. This shows how for conference settings, network stability comes from the core of the κ-vicinity. However, we observe the opposite phenomenon for *Rollernet* dataset. With larger κ we have an increase of the average duration that nodes spend at a certain distance from one another. Thus, due to nodes' movement in a highly dynamic scenario, meeting between users lasts for very short

TABLE IV
κ-CONTACT NUMBER OF INTERVALS ($\times 1,000$).

	κ						
Datasets	1	2	3	4	5	6	7
Infocom05	3.7	14.7	28.9	40.0	46.7	50.3	51.9
Sig09	13.3	49.7	96.9	131.6	152.2	163.4	168.8
Rollernet	2.6	9.4	18.4	27.5	35.2	41.3	45.7

periods of time but nodes spend a significant amount of time in the nearby vicinity.

C. Average κ-contact duration

We also study the average κ-contact durations (see Table III), i.e., we observe the average duration of any κ-contact interval. Following our logic, since we cover a wider spatial range with our κ-vicinity, nodes coming closer are likely to be in κ-contact earlier and leave the κ-contact later, therefore we should obtain longer κ-contact intervals. With *Rollernet*, we observe that the greater the value for κ, the longer the durations. Surprisingly for *Infocom05* and *Sig09*, this is not the case, we actually notice the opposite phenomenon. With larger κ, we seem to have smaller κ-contact intervals. So does that mean that increasing our network vision with the κ-vicinity reduces the duration of end-to-end transmission possibilities?

Table IV shows how wrong this conclusion may be. In this table, we show the actual number of κ-contact intervals for each κ and each dataset. For all of them, the greater the value of κ, the greater the number of κ-contact intervals. So, with higher κ values, we multiply the possibility of observing a κ-contact interval. They may be on average of shorter length (for *Infocom05* and *Sig09*) yet we multiply the possibility of having pairwise end-to-end paths. In addition, the cumulated κ-contact duration grows with larger κ. A similar observation as well as an explanation has been made in a companion paper [9].

IV. PREDICTING κ-CONTACT ENCOUNTERS

A. Dynamic graph representation

The mobile traces analyzed in this paper represent dynamic networks composed of a set of mobile users that sporadically come in contact. We represent this network using a dynamic graph structure, $G_{0,T} = (V, E_{0,T})$, with V the set of mobile users observed during a finite period of time [0, T) and $E_{0,T}$ the set of temporal edges between them. We consider an edge $e_{uv} \in E_{0,T}$ if any two users $u, v \in V$ have been at least once into contact during the period [0, T). To analyze the evolution of this network over time, we split time into fixed time-windows of duration w and represent the dynamic network as a time series of network snapshots G_{t_1}, G_{t_2}, ..., G_{t_n}, with $n = \lceil \frac{T}{w} \rceil$. G_{t_i} represents the aggregate graph G_{t_{i-1},t_i} that records the contacts between mobile users during the period [t_{i-1}, t_i). In a dynamic network, the future changes of the network may depend not only on the most recent state of the network but also on older ones. To model the dynamic evolution and catch possible periodicities in human encounters, the data used as input in the prediction process is represented as a successive series of static snapshots $G_{t_{i-m}}, ..., G_{t_{i-2}}, G_{t_{i-1}}$. Thus, given data from the previous m time-windows our objective is to predict the κ-contacts during the next target period G_{t_i}. We will later discuss how the choice of w and m affect the prediction performance.

B. κ-contact prediction problem

We formulate the prediction task as a binary classification problem where, given past data recorded until a moment in time t_{i-1}, the goal is to predict if any two mobile nodes will be in κ-contact during the subsequent period [t_{i-1}, t_i).

We rely on two types of information in the prediction model: the frequency of κ-contact occurrences and the structural properties of the connectivity network. The first type of information measures the strength of κ-contact relationships, quantified by the *duration* and the *number of times* any pair of nodes has been in κ-contact in the past. A longer duration and a greater number of κ-contacts can provide stronger evidence that two nodes will be in κ-contact in the future. For the second type of information, to quantify the structural properties of the network, we extract various features that capture the proximity between nodes in the graph of past interactions. These features showed predictive power in various applications such as collaborative filtering and link prediction problems [12], [13], [14]. In this work we use four common proximity measures:

- *Common neighbors (CN)*. For each pair of nodes $u, v \in V$, CN represents the number of common neighbors:

$$CN_{(u,v)} = \mid \mathcal{V}_1^u \cap \mathcal{V}_1^v \mid. \tag{1}$$

- *Adamic Adar* [15]. This measure extends the notion of common neighbors by weighting each neighbor by the inverse logarithm of its degree centrality:

$$AdamicAdar_{(u,v)} = \sum_{x \in \{\mathcal{V}_1^u \cap \mathcal{V}_1^v\}} \frac{1}{\mid \mathcal{V}_1^x \mid}. \tag{2}$$

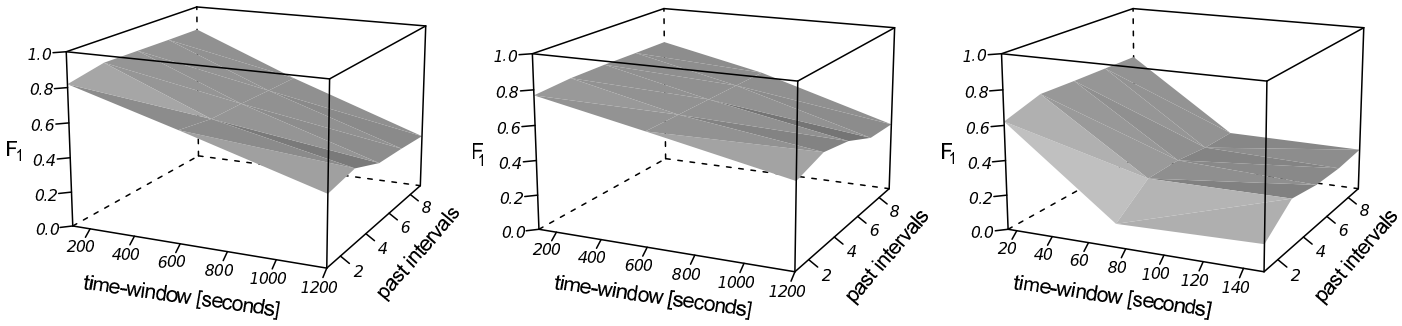
- *Katz* [16]. This feature counts all the paths between any pair of nodes, giving a higher weight to shorter paths. If $path_{u,v}^l$ represents the set of paths of length l between two nodes u and v, and β is a damping factor (set to 0.05 in our evaluation), the Katz score is calculated using the following formula:

$$Katz_{(u,v)} = \sum_{l=1}^{\infty} \beta^l \times \mid path_{u,v}^l \mid. \tag{3}$$

- *Preferential attachment* [17]. This feature is built on the premise that the probability of a new contact is correlated with the product of nodes' degree.

$$PA_{(u,v)} = \mid \mathcal{V}_1^u \mid \times \mid \mathcal{V}_1^v \mid \tag{4}$$

The two types of features provide complementary information about nodes' contact patterns. The frequency of interactions catches the persistence of κ-contact relationships but its predictive power is conditioned by the past contact occurrences



(a) *Infocom05* (b) *Sig09* (c) *Rollernet*

Fig. 5. Prediction performance for different time-window durations and by varying the number of training time-windows (past intervals).

TABLE V
NOTATION FOR THE BINARY CLASSIFICATION CONFUSION MATRIX

		Predicted value	
		predicted = 1	predicted = 0
Actual value	actual = 1	TP	FN
	actual = 0	FP	TN

(using these features one can only predict the reoccurrence of a κ-contact). Topological features, on the other hand, allow us to capture complex data patterns about the structure of the network of interactions. We build the prediction model and report the results using the entire set of features as we observed that taking these features together achieves a higher performance than using them separately.

We adhere to a supervised learning procedure in our evaluation. Each mobile trace is split in two equal-sized temporal parts: the first period is used as the training set and the remaining part serves to report the prediction performance. We examined two classification algorithms: SVM (using LIBSVM library [18]) and logistic regression, under different parameter settings and used a validation set to avoid overfitting. We report the quality of the prediction using the F_1 score (also called F-measure), expressed as the harmonic mean between precision ($\frac{TP}{TP+FP}$) and recall ($\frac{TP}{TP+FN}$) as defined by the confusion matrix (Table V).

C. The effect of time-window duration and past data

The prediction performance is influenced by the duration of the time-window. Aggregating data over longer durations may lose useful temporal information about the structure of the dynamic network. Another important aspect is given by the number of training time-windows. Including more information from the past may capture important temporal patterns but also increase the computational cost.

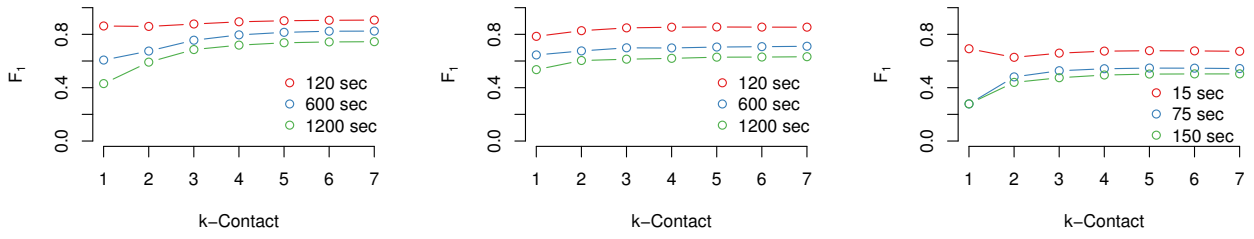
To evaluate the impact of past information in the prediction performance we vary the amount of data used in the prediction model and include information from the previous $\{1, 3, 5, 7, 9\}$ time-windows. We illustrate the results for the 1-contact case as we observed that the remarks made on this value are consistent with other κ values as well. For the size of the time-window we select the most granular duration (the scanning rate used in the mobile trace) and two other values that represent $5\times$ and $10\times$ this duration. Thus, we consider time-windows of duration $\{120, 600, 1200\}$ seconds for *Sig09* and *Infocom05* and use $\{15, 75, 150\}$ seconds for *Rollernet* (which has a more granular frequency).

The results are presented in Fig. 5 by means of 3D plots that represent the F_1 score as a function of the time-window duration and the number past time-windows used in the prediction model. On the x-axis we examine different time-window durations and the y-axis (labeled *past intervals* in Fig. 5) denotes the number of time-windows used in the prediction model. For example, a past interval of length 9 for a time-window of 1200 seconds means that, based on the contacts recorded during the previous 9 intervals of 1200 seconds, we predict contacts during the next time-window.

The figure illustrates that the most recent information plays the most important role in the prediction performance. For all three datasets, using data from the latest three time-periods achieves the highest performance and older information has little predictive power. This indicates that the most recent interactions are the most important in predicting the immediate future. We can also observe that the longer the duration of the time-window, the less accurate the prediction performance. This suggests that aggregating data over longer durations is prone to larger errors. Taking the example of *Infocom05* (Fig. 5(a)), the results show that predicting the contact opportunities during the next 2 minutes shows an F_1 score of 0.8 and the performance drops with 50% when trying to predict what will happen during the next 20 minutes. For *Rollernet*, which represents a more dynamic scenario, the drop of performance is even higher with a 70% decrease when trying to predict the contacts during the next 150 seconds compared to a 15-seconds time-window.

D. κ-contact prediction results

Based on the previous observations of the optimal number of past intervals we assess the performance of predicting κ-contact relationships. We vary the value of κ from 1 to 7 and consider three durations for the time-window: $\{120, 600, 1200\}$ seconds for *Infocom05* and *Sig09* and $\{15, 75, 150\}$ seconds for *Rollernet*. The results are illustrated in Fig. 6. First, we observe that predicting that two nodes will be in direct communication range shows particularly poor results in very dynamic mobile settings (e.g. *Rollernet* that describes a



(a) *Infocom05* (b) *Sig09* (c) *Rollernet*

Fig. 6. The efficiency of predicting κ-contact relationships for different durations of the time-window. On the y-axis we represent the prediction performance and on the x-axis we vary the value of κ-contact from 1 to 7.

rollerblading activity) and for longer durations of the time-window. Thus, in situations that involve important changes in the network topology, predicting that nodes will be in direct contact is prone to large errors.

Relaxing the prediction objective beyond direct contact relationships reveals more accurate predictive power. Overall, the greater the value for κ the more effective the prediction performance. On average (for all mobility traces and different time-window durations) predicting that nodes will be at most at a distance 2, 3, and 4 shows an improvement of 7%, 10%, and 11% compared to the case where we want to predict direct meetings between mobile users. While the improvement is important for small values of κ we notice that there is little benefit in extending the prediction for a κ greater than 3. The most significant increase, compared to the direct contact case, can be observed for κ = 2 with an average increase of 10% for *Rollernet*, 7% for *Infocom05*, and 6% for *Sig09*. The benefit is negligible when trying to predict the network change in the immediate horizon but it becomes significant when trying to make predictions over longer periods of time. Taking the case of *Infocom05* for a time-window of 1200 seconds and *Rollernet* for 150 seconds, predicting that nodes will be separated by at most two nodes (κ-contact = 3) reveals an improvement of 60% for *Infocom05* and 74% for *Rollernet* compared to the direct contact prediction case.

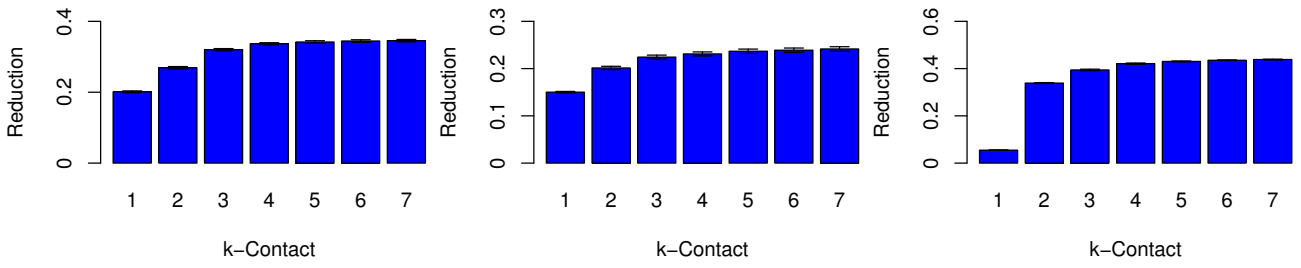
While we leave a more detailed analysis of these findings for future work, we give two plausible explanations for these results. First, as we showed in Fig. 4, a non-negligible number of nodes, although never in direct contact, they come at a 2-hop distance. By extending the prediction objective to 2-hop contacts, we include these potential events into consideration, which appear to have a more predictable nature. Then, as showed in Section III direct contacts between mobile users are scarce and short-lived, which makes them more difficult to predict in very dynamic scenarios and for longer time horizons. This explains the low prediction effectiveness observed with *Rollernet* and for longer time-windows for *Sig09* and *Infocom05*. Thus, extending the notion of contact to κ-contact gives us access to more stable connections (nodes leave direct connectivity but remains in κ-contact for longer durations) that reveal a more predictable nature.

V. Practical Implications

To capture the possible benefit that κ-contact prediction would bring in practical scenario we propose and evaluate the following use-case example.

We consider a content producer, located on the Internet, that regularly publishes content for a known group of collocated mobile users that communicate with the server using the cellular infrastructure. Content is categorized in topics. Users subscribe to these topics and content is pushed to users upon creation. We also consider that, in order to reduce the amount of cellular traffic caused by content delivery, the content producer collects data about the mobile contact traces and relies on a κ-contact prediction functionality when transmitting information to users. More specifically, at the publication of a content, instead of individually transmitting the content to each subscriber, the content producer optimizes the delivery process based on the predicted κ-contact opportunities. For example, if the server predicts that two users, interested in the same content, will be in κ-contact, a message is sent to only one of these nodes which will opportunistically forward the message to the other node when they will be κ-contact. We also assume that nodes are capable of sensing their κ-vicinity and can detect when a targeted user is in κ-contact. To collect nearby topological knowledge, we assume the existence of a link-state protocol gathering nearby knowledge under the form of a connectivity graph. The implementation itself is beyond the scope of our study, yet a previous analysis studied the impact of monitoring overhead [8].

We design the experimental setting using ONE simulation environment [19]. In our experiments we set the number of topics to 100. Each mobile node randomly subscribes to 20 up to 100 topics. For the prediction module, we use a time-window of 75 seconds for *Rollernet*, and 600 seconds for *Infocom05* and *Sig09*. Content is uniformly created throughout the duration of the experiments (that covers the duration of a mobility trace) and the results are averaged over 10 simulation runs. We also consider an infinite cache size at the user side and assume that the content is small enough to fit into one message in the communication between content producer and the users and between the mobile users. To measure the impact of κ-contact prediction we report the reduction in the number of messages in the communication between the



(a) *Infocom05* (b) *Sig09* (c) *Rollernet*

Fig. 7. The percent of traffic with the infrastructure that can be reduced through κ-contact prediction and mobile opportunistic communications. On the y-axis we represent the traffic reduction compared to the case where content is sent to mobile users using only the infrastructure. On the x-axis we present different values for κ-contact.

content producer and the mobile users when using κ-contact prediction module compared to a case where the content is individually sent to each user using the cellular infrastructure.

The results are presented in Fig. 7. First, we observe that the greater the value of κ-contact, the greater the potential of traffic reduction. The biggest improvement of predicting beyond direct neighbors is noticed for κ = 2, that shows an improvement of 6% for *Sig09*, 7% in *Infocom05*, and 30% for *Rollernet*. The potential traffic reduction is directly affected by the characteristics of the traces: κ-vicinity properties (presented in Fig. 4) and prediction performance (presented in Fig. 6). Taking the example of *Sig09*, even if the effectiveness of the prediction showed little improvement for κ = 2 compared to κ = 1 the potential reduction is nevertheless important (6%). This is explained by the significant number of nodes located at a 2-hop distance detected with the κ-contact prediction. The benefit is even more substantial in the case of *Rollernet*. By counting on the pairs of nodes connected at a 2-hop distance (that exceed the number of direct contact opportunities), the traffic reduction attains a performance of 33% compared to 5% when using only direct contact prediction.

VI. Related work

The DTN community used different ways to benefit from a node's neighborhood. Some relied on the social behavior of users. In urban areas, where people tend to form communities around points of interests, Ott et al. presented a protocol leveraging end-to-end and multi-hop DTN paths [20]. Sarafijanovic-Djukic et al. made a similar observation for VANET [21]. Heimlicher and Salamatian demonstrated that mobile wireless networks tend to have connected crowds [4]. The transient notion between all these studies is that there are immediate neighborhood structures to use but none of them actually defined a notion of vicinity for DTN. Similar principles have also been considered even in other contexts, such as wireless mesh networks [22].

In the latest years, analyzing human mobility to detect useful patterns and to derive accurate prediction models has been well studied. Song et al. observed that, despite the many decisions influencing our daily routines, there is a 93% potential in predicting user mobility [23]. Clauset and Eagle revealed strong periodicities in contact periods between mobile users that may depend on the environment under study (the physical place and the type of user activity) [3]. Zayani et al. studied the problem of predicting contact opportunities between mobile users [7]. Using a tensor-based link prediction technique, they provide evidence about the predictive power of various features that capture both the topological distance and the physical proximity between users. In this work we focus on one specific aspect of human mobility, i.e. predicting if two nodes will be in each other κ-vicinity. Our analysis is close to the work of Zayani et al. but differs in the prediction goal (we extend the contact prediction to the κ-contact case), the mobility traces under study, and the prediction framework (we use a supervised learning framework compared to the unsupervised setting used in their paper).

From a general point of view, our prediction objective is related to the link prediction problem in complex networks. This topic is an important research direction in several domains that define relationships between different entities. This includes predicting the co-authorship of research publications, hyperlinks between web pages, or human communication patterns [13], [14], [24], [25]. Liben-Nowell and Kleinberg have studied the predictive power of various topological features and observed that the Katz measure performs consistently well. While analyzing the predictive power of non-topological attributes Al Hasan et al. observed that the frequency of interactions (e.g. co-authorship of scientific papers) is an efficient predictive variable [26]. We build on this knowledge, and we analyze the predictive powers of different features (topological measures and the frequency of users encounters) in predicting the κ-contact relationships between mobile users.

VII. Conclusion and future work

In this work, we addressed the problem of predicting κ-contact opportunities between mobile users – predict if users will find themselves at a distance of at most κ-hops from one another. By analyzing three real-world contact traces, we observed that one can obtain better performances when predicting 2^+-contacts compared to the direct contact case. To assess the impact of these findings in a real-world application, we proposed a simulation experiment in which, by combining

mobile opportunistic communications with κ-contact prediction one can reduce the amount of traffic used in the communication of mobile nodes with the infrastructure. Our results suggest that services benefiting from contact predictions can efficiently exploit the predictable nature of κ-contacts.

Research in the area of κ-contact prediction and its applications is an open subject with many possible extensions. First, as the observations made in this paper are based on specific mobility settings (conferences and rollerblading) more work is needed in order to understand to what extend these observations can be generalized to other mobility scenarios. The quality of the prediction shows promising performances, yet not optimal, and suggests there is still room for improvement. One way to increase the prediction performance is to consider additional features in the prediction model. This includes information about the geographical co-location patterns of nodes (not available in the traces used in this paper) that showed strong predictive power in the link prediction problem [7], [14]. Then, in this study, we only consider a one-step ahead prediction problem (we use information received in the previous m time periods to predict what will happen during $m+1$ period). To cover a larger range of situations the prediction objective can be extended to further time periods, i.e. predict contacts during subsequent time periods $m+1$, $m+2$, ..., $m+n$. This can be particularly favorable in situations where collecting and processing data adds a significant delay, for which immediate prediction can be considered outdated.

Finally, the current evaluation of the applicability of κ-contact prediction in real-world scenarios considers some simplifying assumptions and more work would be needed to assess the benefit of κ-contact prediction in practice. This includes an evaluation of the additional cost of collecting data about nodes' mobility and the feasibility of implementing a mechanism capable of detecting the κ-contact communication opportunities.

ACKNOWLEDGMENT

Alexandru-Florin Tatar, Tiphaine Phe-Neau, Marcelo Dias de Amorim, and Serge Fdida carried out part of the work at LINCS (http://www.lincs.fr). This work was partially funded by the European Community's Seventh Framework Programme under grant agreement no. FP7-317959 MOTO.

REFERENCES

[1] K. Fall, "A Delay-Tolerant Network Architecture for Challenged Internets," in *ACM Sigcomm*, Karlsruhe, Germany, Aug. 2003.

[2] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 606–620, 2007.

[3] A. Clauset and N. Eagle, "Persistence and periodicity in a dynamic proximity network," *arXiv preprint arXiv:1211.7343*, 2012.

[4] S. Heimlicher and K. Salamatian, "Globs in the Primordial Soup – The Emergence of Connected Crowds in Mobile Wireless Networks," in *ACM Mobihoc*, Chicago, Illinois, USA, Sep. 2010.

[5] V. Erramilli, A. Chaintreau, M. Crovella, and C. Diot, "Diversity of forwarding paths in pocket switched networks," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 2007, pp. 161–174.

[6] V. Erramilli, M. Crovella, A. Chaintreau, and C. Diot, "Delegation forwarding," in *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2008, pp. 251–260.

[7] M.-H. Zayani, V. Gauthier, and D. Zeghlache, "Improving link prediction in intermittently connected wireless networks by considering link and proximity stabilities," in *WOWMOM*, 2012, pp. 1–10.

[8] T. Phe-Neau, M. Dias de Amorim, and V. Conan, "The Strength of Vicinity Annexation in Opportunistic Networking," in *IEEE NetSciCom*, Torino, Italy, Apr. 2013.

[9] ——, "Vicinity-based DTN Characterization," in *ACM MobiOpp*, Zurich, Switzerland, Mar. 2012.

[10] A.-K. Pietilänen and C. Diot, "Dissemination in opportunistic social networks: the role of temporal communities," in *ACM Mobihoc*, Hilton Head, South Carolina, USA, Jun. 2012.

[11] P.-U. Tournoux, J. Leguay, F. Benbadis, J. Whitbeck, V. Conan, and M. D. de Amorim, "Density-aware routing in highly dynamic DTNs: The rollernet case," *IEEE Transactions on Mobile Computing*, vol. 10, pp. 1755–1768, 2011.

[12] Z. Huang, X. Li, and H. Chen, "Link prediction approach to collaborative filtering," in *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*. ACM, 2005, pp. 141–142.

[13] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *J. Am. Soc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, May 2007.

[14] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi, "Human mobility, social ties, and link prediction," in *ACM SIGKDD*, 2011, pp. 1100–1108.

[15] L. A. Adamic and E. Adar, "Friends and Neighbors on the Web," *Social Networks*, vol. 25, no. 3, pp. 211–230, 2003.

[16] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39– 43, 1953.

[17] A.-L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations," *Physica A: Statistical Mechanics and its Applications*, vol. 311, no. 3, pp. 590–614, 2002.

[18] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[19] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE Simulator for DTN Protocol Evaluation," in *International Conference on Simulation Tools and Techniques*, Rome, Italy, Mar. 2009.

[20] J. Ott, D. Kutscher, and C. Dwertmann, "Integrating DTN and MANET routing," in *ACM Chants*, Pisa, Italy, Sep. 2006.

[21] N. Sarafijanovic-Djukic, M. Pidrkowski, and M. Grossglauser, "Island Hopping: Efficient Mobility-Assisted Forwarding in Partitioned Networks," in *IEEE SECON*, Reston, VA, USA, Sep. 2006.

[22] M. E. M. Campista, L. H. M. K. Costa, and O. C. M. B. Duarte, "A routing protocol suitable for backhaul access in wireless mesh networks," *Computer Networks*, vol. 56, no. 2, pp. 703–718, Feb. 2012.

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

[24] B. Taskar, M.-F. Wong, P. Abbeel, and D. Koller, "Link prediction in relational data," in *Neural Information Processing Systems*, vol. 15, 2003.

[25] Z. Huang and D. K. Lin, "The time-series link prediction problem with applications in communication surveillance," *INFORMS Journal on Computing*, vol. 21, no. 2, pp. 286–303, 2009.

[26] M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," in *SDM06: Workshop on Link Analysis, Counter-terrorism and Security*, 2006.