An Arrival-based Framework for Human Mobility Modeling

Dmytro Karamshuk
IMT Lucca
Lucca, Italy
dmytro.karamshuk@imtlucca.it

Chiara Boldrini, Marco Conti, Andrea Passarella
IIT-CNR
Pisa, Italy
firstname.lastname@iit.cnr.it

Abstract—**Modeling human mobility is crucial in the performance analysis and simulation of mobile ad hoc 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 newest features of mobility properties (e.g., visiting patterns to locations or inter-contact times) as they came up from trace analysis. As a consequence, typically these models are neither flexible (i.e., features of mobility cannot be changed without changing the model) nor controllable (i.e., the exact shape of mobility properties cannot be controlled directly). In order to take into account the above requirements, in this paper we propose a mobility framework whose goal is, starting from the stochastic process describing the arrival patterns of users to locations, to generate pairwise inter-contact times and aggregate inter-contact times featuring a predictable probability distribution. We validate the proposed framework by means of simulations. In addition, assuming that the arrival process of users to locations can be described by a Bernoulli process, we mathematically derive a closed form for the pairwise and aggregate inter-contact times, proving the controllability of the proposed approach in this case.**

I. INTRODUCTION

Modeling human mobility is crucial in performance analysis of networking protocols for mobile ad hoc networks. In mobile ad hoc networks messages are routed by the users of the network (which exchange them upon encounter with other users) and eventually delivered to their destinations. The delay experienced by messages is thus a function of the pairwise inter-contact time, which is the time between two consecutive contacts for a pair of nodes. Characterizing the inter-contact time is therefore essential for modeling the performance of networking protocols for mobile ad hoc networks. 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.

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 [8] [20]) and social (i.e., how long users stay together and how long they have to wait before meeting again [4]) 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 characterised. Many hypotheses have been made (about them featuring an exponential distribution [7], a Pareto distribution [4], a Pareto with exponential cut-off distribution [10], etc.), but the problem has yet to be solved.

Building upon the above findings, the current approach to human mobility modeling has been so far based on trying to reflect in the model the newest features of mobility properties as they came up 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 travelled by the users [14]. 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 [6] [21]. 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 [1] [2].

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 test their robustness to different mobility behaviors. Second, on more practical terms, it allows for changing the model upon new discoveries from trace analysis without the need to start

over from 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 grain with the majority of available mobility models. For example, in social-based mobility, where the number of social relationships determines the shape of inter-contact times, an appropriate configuration can lead to heavy tailed inter-contact times [1]. However, there is no direct way for 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 twofold. First, we propose a mobility framework aiming to be flexible and controllable (Section III). Our framework 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. This information can be extracted from available mobility traces or collected from location-based social networking applications such as Gowalla and Foursquare. Based on the input social graph, communities are identified and are assigned 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 *arrival network*), where a link between a generic user i and a location l characterizes the way user i visits location l. As emerges from the above description, the framework combines the social dimension of user movements together with spatial and temporal preferences, while at the same time allowing for flexibility in visiting patterns and in the resulting inter-contact times (as shown in Section IV).

The second contribution of this work lies in considering a specific instance of the framework, and analytically deriving the relation between arrival patterns of users to locations and the resulting pairwise and aggregate inter-contact time distribution. In our analysis we assume that the time is slotted, and we represent the way users arrive to locations as Bernoulli process. We prove that when the arrival process describing how users visit their assigned locations is Bernoulli, then also the contact process is Bernoulli, and the pairwise inter-contact time features a geometric distribution whose parameter can be derived in closed form (Section V-A). This shows that there can be a direct control on the output of the mobility model using the proposed framework, even if we do not have a general enough analytical model to cover all cases.

As for aggregate inter-contact times, recently Passarella and Conti [19] have investigated how aggregate inter-contact times depend on the pairwise statistics from which they originate. Understanding this dependence is extremely important, because aggregate statistics are much easier to collect in practice than pairwise ones, and, in the past, conclusions were often drawn from the former considering them to be representative of the latter. In this paper we take a step further into the investigation of aggregate and pairwise statistics by studying how individual arrival patterns to locations affect the aggregate inter-contact time. More specifically, we prove (Section V-B) that heavy-tailed aggregate inter-contact times, which have emerged from the analysis of real mobility traces [4], can be obtained from simple heterogenous Bernoulli arrivals. This confirms the main result in [19], i.e., that heterogeneity in pairwise statistics can lead to aggregate statistics that are very distant in distribution.

II. RELATED WORK

A comprehensive overview of the state-of-the-art in mobility modeling was presented in [11]. The main findings in human mobility research can be classified along the three axes of *spatial*, *temporal*, and *connectivity* 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), connectivity properties to the interactions between users.

Similarly, three dominating techniques have emerged in the 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 regular reappearance of users at the sets of preferred locations. Their general approach is to store the maps (i.e., the sets) of preferred places for each of the users and to explore them while deciding on the next destination for their walks. The main representatives of this group are SLAW [20] and the model proposed by Song et al. [14]. Being able to satisfy the main spatial properties of human mobility trajectories, these models do not pay enough attention to the other - social and temporal - aspects of the movements.

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 [21]. The model incorporates detailed geographic topology, personal schedules and motion generator defined for more than 30 different types of activities. Although the model gives an extremely thorough representation of human movements in very particular scenarios, it does not explain the main driving forces of human mobility and it is too complex for mathematical analysis.

The most recent and most rapidly evolving trend in modeling human mobility is based on incorporating complex network theory and 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.
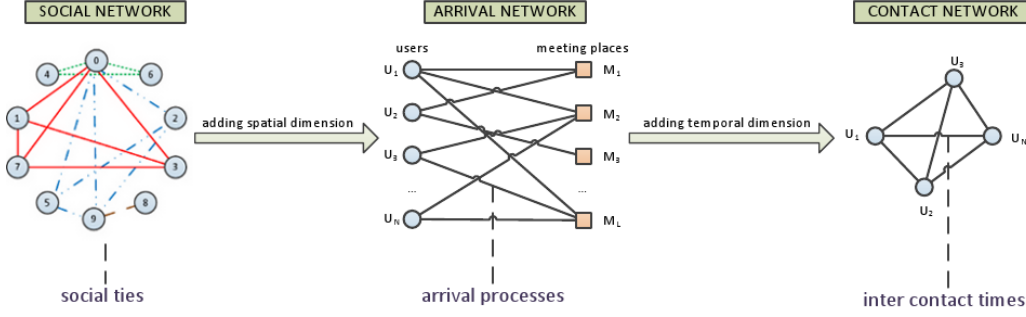
Figure 1. Framework overview

The first models of this class were CMM [15] and HCMM [1], although many others have recently been developed.

Although existing models are able to reproduce realistic mobility trajectories in a wide range of scenarios, there is still the need for a better understanding of the correlations between the main characteristics of human movements, i.e., patterns of user co-appearance at locations and emerging connectivity properties, i.e., inter-contact times. To the best of our knowledge, here we present the first approach where the time sequence of user re-appearance in shared meeting places are explicitly modeled. Therefore, with respect to existing models, our approach is easily customizable to any temporal pattern of user co-appearance in locations and gives a natural framework for the mathematical analysis of the contact sequences between them. Additionally, we reflect the social dimension of human mobility in the distribution of the shared meeting places between communities of tightly coupled users.

III. THE PROPOSED MOBILITY FRAMEWORK

In this section we introduce our mobility framework, designed around the three main dimensions – social, spatial and temporal – of human mobility (see Figure 1). The social dimension is explicitly captured in the framework by taking a graph of human social relationships as input parameter. This graph can be any well known graph, such as a random graph [16] or a scale-free graph [16], 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 its movements. We explore the fact that the structure of communities in the social graph has a significant impact on human mobility, thus we assign users to meeting places such that communities of tightly connected users (*cliques*, in complex network terminology) share similar 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 [22]. A stochastic point process is a stochastic process that characterizes how events (usually called arrivals) 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.

A. The social dimension 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]. As anticipated, in our analysis we consider proximity-based communities, i.e., communities whose members share a common meeting place (e.g., office, bar, apartment). The fact that all members of the community visit a shared meeting place 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 real social networks exhibit a *overlapping* and *hierarchical structure* [17] [18]. 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 users of the parent clique. For example, a company shares a set of offices visited by all its employees, while each division has its own working place.

B. Adding the spatial dimension to social graphs

The goal of the proposed framework is to reflect in the spatial behavior of users the structure of their social communities. 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 I) for generating the arrival network starting from the input social graph we mainly need two components: a clique finding algorithm (that also detects overlapping cliques) and a way for reproducing hierarchical cliques.

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 [3] 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) Repeat the procedure starting from the first step, until there are no links left in the input graph.
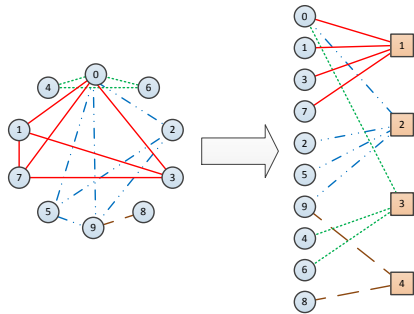

Figure 2. A round of assigning social cliques to meeting place; cliques are marked with different line styles

The first component corresponds to steps 1 and 2 in Table I. In each round, the social graph is divided into a set (called *cover*) of overlapping cliques, such that each link of the graph is assigned to exactly one clique. To this purpose, we use the BronKerbosch algorithm [3]. 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 the 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 are reflected into corresponding meeting places.

The second component (step 3 in Table I) 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 (*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 include 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 I) when there are no more links to be removed in the social graph.

1) From meeting places to geographical locations: The analysis of the algorithm in Table I reveals that the number of generated meeting places 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 want to 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 this aim, we exploit the observation that a general urban environment is characterized by the existence of locations with an extremely large number of meeting places (e.g., business centers), while for the majority of locations this number is significantly smaller (e.g., private houses). This is conceptually similar to a power law distribution of meeting places per location. Therefore, the grouping of the meeting places to a given number of physical locations can be done with the de-facto standard for reproducing such distribution, i.e., the *preferential attachment* scheme [16]. In this scheme, meeting places are one-by-one randomly attached to one of the available locations, with the probability of selecting a specific location being proportional to the number of places already attached to it.

C. 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 want to further characterize the way users visit locations by considering the temporal properties of such visits. To this aim, we assign to each link in the arrival network a stochastic point process A_i^l that describes the arrivals of user i to 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. As an example, in light of the results from trace analysis, a single time slot can be taken to be equal to one day.

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. The contact graph can be fully mathematically characterized (we provide an example of this characterization in Section V for the case of arrival processes being heterogenous Bernouilli processes) or it can be

obtained from simulations. From the simulation standpoint, the output of the proposed framework is a contact trace in the form `<user1, user2, startTime, endTime>`, where the first two elements are the node identifiers and the last two denote the start and end time slots of the contact. This trace can be then fed into a networking simulator like the ONE simulator [13].

IV. CASE STUDIES

We discussed in Section I that a desired property of a mobility framework is its ability to reproduce different distributions for the main mobility properties. From this standpoint, in this section we analyze, by means of simulations, the behavior of the framework described above, focusing on the distribution of inter-contact times that it generates. Complementary to this, in [12] we provide a preliminary validation of the model based on real world mobility traces.

As we already discussed, inter-contact times play a major role in the delay experienced by messages in mobile ad hoc networks. The evaluation provided below, despite preliminary, clearly indicates that the proposed framework is able to generate different distributions for the inter-contact time. In some cases, as in the first scenario discussed below, we have a complete understanding of how a given inter-contact time distribution can be obtained from the input parameters of the framework. However, the derivation of a general analytical model relating the arrivals of users and the resulting inter-contact times is an ongoing work.

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 (Table I). As input social graph we consider two random graphs G_{n_1,χ_1} and G_{n_2,χ_2} of $n_1 = 500$ and $n_2 = 1000$ users, in which each possible edge exists with probability $\chi_1 = 0.2$ and $\chi_2 = 0.1$, respectively. Recall that a random graph G_{n_i,χ_i} is obtained by starting with a set of n_i vertices and randomly adding edges between each pair of vertices with probability χ_i [16]. We evaluate both these two graphs 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 the 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 changing the arrival processes A_i^l of users to meeting places. Simulations are run for 10000 time units of simulated time, and results are shown with a confidence level of 99.9%.

In the first experiment we model users' arrivals to places with Bernoulli arrival processes. In a Bernoulli process, the probability of an arrival at a given time slot is constant

and equal to $\rho_{A_i^l}$, which also corresponds to the rate of the process. Here 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 characterize in Section V. Figure 3 depicts the result of simulations for each arrival network. For instance, Figure 3.a depicts simulation results for the network with parameters $n = 500$, $\chi = 0.2$ 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: while the structural properties of the network influence the parameter of the aggregate inter-contact time CCDF, the shape of the distribution remains the same for all the networks, and it can be approximated with a power law function of exponent $\gamma = -2$.
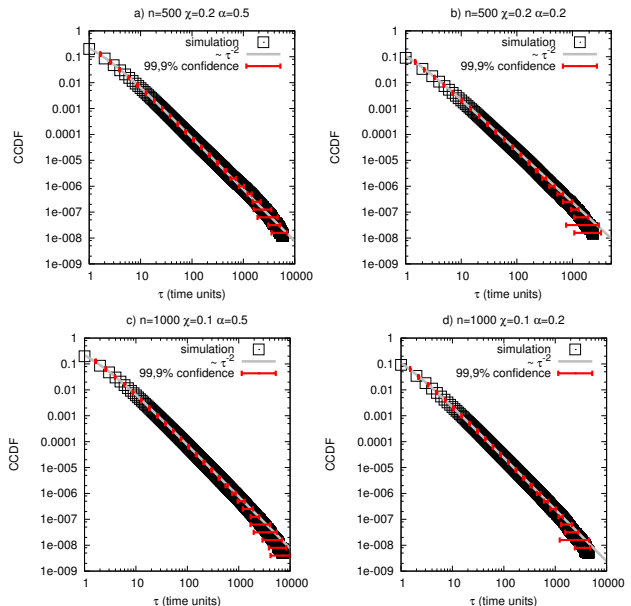

Figure 3. The aggregate inter-contact time for different arrival networks

In the second experiment we change the type of arrival processes in the network and study the corresponding change in the distribution of aggregate inter-contact times. More specifically, we consider the case when the arrival processes are point processes with independent uniformly distributed intervals, i.e., the intervals between arrivals of process A_i^l are drawn from a discrete uniform distribution of range $[1, b]$, where $b = 2 \times \lfloor 1/\rho_{A_i^l} \rfloor - 1$, $\lfloor \ \rfloor$ is the floor (greatest integer) function and $\rho_{A_i^l}$ is drawn from the same distribution as in the first experiment (note that the average inter-arrival of a user to a location is approximately the same as in the previous experiment). We model two networks with parameters $\{n_1 = 500, \chi_1 = 0.2, a_1 = 0.5\}$ and $\{n_2 = 500, \chi_2 = 0.2, a_2 = 0.2\}$. Therefore, the only difference with respect to the first experiment is the type of arrival processes. The results presented in Figure 4 clearly show that the

aggregate inter-contact time distribution also in this case decays as a power law with exponent $\gamma = -2$. In other words, this result reflects the corresponding result for the Bernoulli arrival processes with similar distribution of the arrival rates. Thus, it may indicate that the distribution of the arrival rates (which is approximately the same in both experiments and equal to $\rho_{A_i^l}$) plays a major role in the emergence of the power law aggregate inter-contact times distribution.
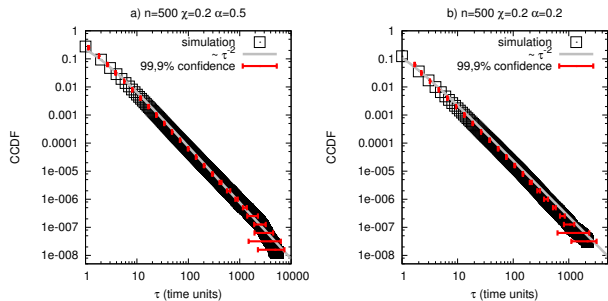
Figure 4. The aggregate inter-contact time for the case of arrival processes with independent uniformly distributed intervals

To better understand the influence of the arrival rates on the resulting inter-contact times characteristic, in the third experiment we simulate arrival networks where arrival processes are Bernoulli processes, like in the first experiment, but this time with identical rates. More specifically, we model two networks with the same parameters $\{n = 500, \chi = 0.2, a = 0.5\}$, in which all the rates of arrival processes are identical and equal to $\rho_{A_i^l}^{(1)} = 1/2$ for the first network, and $\rho_{A_i^l}^{(2)} = 1/3$ for the second one. 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 the network where 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 the network with average inter-arrival time for all processes equal to $1/\rho_{A_i^l}^{(2)} = 3$ time units. From Figure 5 (plotted in lin-log scale) we can see that the aggregate inter-contact times characteristic in such networks differs from the results we obtained before. More specifically, the resulting distribution of the aggregate inter-contact times decays as an exponential function rather than as a power law function. This result additionally supports the hypothesis that the resulting aggregate inter-contact times characteristic significantly depends on the distribution of arrival rates.

The preliminary results discussed in this section provide an indication that the distribution of the rates of the arrival processes plays a major role in the resulting aggregate inter-contact times between users. Moreover, these results allowed us to show how very different distributions for the aggregate inter-contact times can be obtained starting from simple Bernoulli arrival processes. This finding is very interesting from the standpoint of a mathematical analysis of the proposed framework, as Bernoulli processes possess
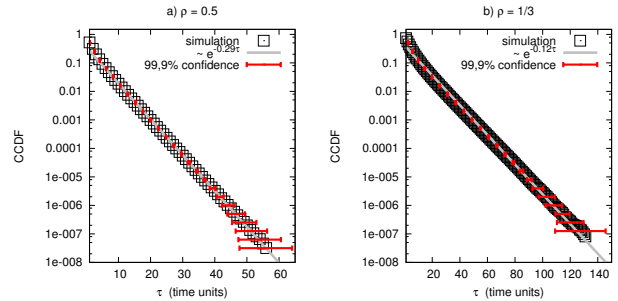
Figure 5. The aggregate inter-contact time for arrival network with identical arrival rates

a number of properties (e.g., single parameter, memoryless property) that significantly simplify the analysis, as we show in the following section.

V. ANALYTICAL MODEL FOR THE CONTACT PROCESS

In this section we show how, in a concrete instance of the framework, we are able to analytically derive the distributions of pairwise and aggregate inter-contact times starting from a given discrete stochastic process describing how users visit locations, thus highlighting the controllability of the framework. For the sake of completeness, the findings from this section are validated in [12] by comparing analytical results with simulations.

In the following we study the dependence between individual arrival processes of users and the corresponding contact processes between them. A *contact process* describes 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. 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 inter-arrival times for a Bernoulli process feature a geometric distribution, we obtain that from geometric inter-arrival times to specific meeting places (corresponding to Bernoulli arrivals) a geometric distribution of pairwise inter-contact times follows. 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 to derive analytically also the aggregate inter-contact times as a function of the arrival rates of users to meeting places. Specifically, we describe the conditions for which the aggregate inter-contact time has a power law shape.

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 II. Due to space limitations, the complete proofs for the results shown in this section can be found in the associated technical report [12].

Table II
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 to 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

A. Contact process for a pair of users

In this section, assuming Bernoulli arrivals to 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 in 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, 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 of them 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 Figure 6. In the following lemma we prove that the single-place contact process C_{ij}^l is also a Bernoulli point process.
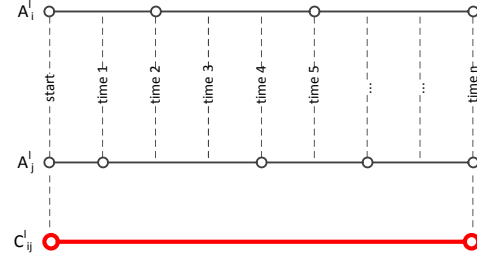

Figure 6. The single-place contact process as an intersection of arrival processes

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_i^l, of rates $\rho_{A_i^l}$ and $\rho_{A_j^l}$ respectively, is a Bernoulli process of rate $\rho_{C_{ij}^l} = \rho_{A_i^l} \times \rho_{A_j^l}$.

The intuitive proof for the above lemma is that 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 $\rho_{A_i^l} \times \rho_{A_j^l}$ is again a Bernoulli process.

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 at least in one of the L_{ij} meeting places that they share. Thus, the contact process between users U_i and U_j can be obtained merging (as shown in [22]) their single-place contact processes (Figure 7). In the following theorem we show that if single-place contact processes are Bernoulli, then also the contact process is Bernoulli.
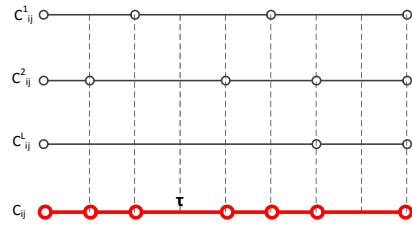

Figure 7. The compound contact process as a merging of single-place contact processes

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 point process with the rate $\rho_{C_{ij}} = 1 - \prod_{l=1}^{L_{ij}} (1 - \rho_{A_i^l} \times \rho_{A_j^l})$.

The intuitive proof for the above result is that 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. Then, Theorem 1 follows.

The contact process described in the 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 times distribution between a pair of users U_i and U_j, meeting at a number L_{ij} of meeting places, and whose arrivals to these meeting places are described as Bernoulli arrival processes A_i^l and A_j^l of rates $\rho_{A_i^l}$ and $\rho_{A_i^l}$, is geometric with rate $\rho = 1 - \prod_{l=1}^{L_{ij}} (1 - \rho_{A_i^l} \times \rho_{A_j^l})$.

B. 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 a concrete case by providing the conditions on the Bernoulli arrival processes of users to locations such that the resulting aggregate inter-contact time distribution is heavy-tailed. Heavy-tailed distributions for aggregate inter-contact times are important as they have often emerged from the analysis of real mobility traces [4]. Our derivation shows how such heavy tailed behavior can result from simple heterogenous Bernoulli arrival processes, which are very convenient to deal with for mathematical analysis. This result also confirms the main finding of [19]: 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 of the work by Passarella and Conti [19], which describes the dependence between the aggregate inter-contact time distribution and the inter-contact time distributions of individual pairs of users. Specifically, they consider a heterogeneous scenario, where pairwise inter-contact times 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 [19] 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 $F(\tau) = \frac{1}{E[P]} \int_0^\infty \rho f_P(\rho) F_\rho(\tau) \mathrm{d}\rho$, where $F_\rho(\tau)$ denotes the CCDF

of the inter-contact times between a pair of nodes whose rate is equal to ρ.

We extend this finding 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 times 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 of the geometric distribution of the pairwise inter-contact times, i.e., $F_\rho(\tau) = (1 - \rho)^\tau$, $\tau \in \{1, 2, 3, ...\}$, can be re-written in a discrete exponential form, i.e., $F_\lambda(\tau) = e^{-\lambda \tau}$, $\tau \in \{1, 2, 3, ...\}$, 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 and the result in Theorem 2, in Lemma 2 we derive under which condition for parameter λ the aggregate inter-contact time is heavy-tailed.

Lemma 2: In a network where pairwise inter-contact times have a discrete exponential distribution of the form $F_\lambda(\tau) = e^{-\lambda \tau}$, $\tau \in \{1, 2, 3, ...\}$, and parameters λ are drawn from an exponential distribution with rate α, the aggregate inter-contact time distribution is $F(\tau) = \frac{\alpha + \alpha^2}{(\tau + \alpha)(\tau + \alpha + 1)}$.
The complete proof for the above Lemma and for all results introduced below can be found in the associated technical report [12].

More generally, the result in Lemma 2 says that the aggregate inter-contact times 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 individual inter-contact times is exponential. In the rest of the section we develop this case and show how the exponential distribution of the parameter of individual inter-contact times emerges in the 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 Corollary 1, 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 parameters λ of the individual inter-contact

times is exponential. To this aim, we prove the following lemma.

Lemma 3: 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, then the resulting pairwise inter-contact times parameters λ are exponentially distributed with parameter α.

A condition for Lemma 2 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 III. Therefore, the result of Lemma 2 can be applied to the networks generated by the mobility framework. Finally, we combine the results of Lemma 2 and Lemma 3 in the following theorem.

Theorem 3: 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 inter-contact times is given by Lemma 2.

VI. Conclusions

In this paper we have proposed a mobility framework that, starting from an input social graph, characterizes the way users visit locations. The spatial dimension of mobility is added imposing that people belonging to the same social community are assigned the same location, which is where the people of that community meet. Then, the way users visit their assigned locations over time is described by means of a stochastic process. We have shown that this framework is flexible, i.e., is able to generate inter-contact times featuring different distributions. Finally, assuming that users arrive to locations according to heterogenous Bernoulli processes, we have analytically derived the pairwise and aggregate inter-contact times, thus showing the predictability of the model in this case.

Acknowledgment

References

[1] C. Boldrini and A. Passarella. HCMM: Modelling spatial and temporal properties of human mobility driven by users' social relationships. *Computer Communications*, 33(9):1056–1074, 2010.

[2] V. Borrel, F. Legendre, M.D. De Amorim, and S. Fdida. Simps: Using sociology for personal mobility. *IEEE/ACM Trans. Netw.*, 17(3):831–842, 2009.

[3] C. Bron and J. Kerbosch. Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575–577, 1973.

[4] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Trans. Mobile Comp.*, pages 606–620, 2007.

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

[6] F. Ekman, A. Keränen, J. Karvo, and J. Ott. Working day movement model. In *ACM/SIGMOBILE workshop on Mobility models*, pages 33–40. ACM, 2008.

[7] Wei Gao, Qinghua Li, Bo Zhao, and Guohong Cao. Multicasting in delay tolerant networks: a social network perspective. In *MobiHoc '09*, pages 299–308. ACM, 2009.

[8] M.C. Gonzalez, C.A. Hidalgo, and A.L. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[9] W. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy. Modeling time-variant user mobility in wireless mobile networks. In *IEEE INFOCOM 2007*, pages 758–766, 2007.

[10] T. Karagiannis, J.Y. Le Boudec, and M. Vojnovic. Power law and exponential decay of intercontact times between mobile devices. *IEEE Trans. Mobile Comp.*, 9(10):1377–1390, 2010.

[11] D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella. Human mobility models for opportunistic networks. *IEEE Commun. Mag*, (12):157–165, December 2011.

[12] D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella. An arrival-based framework for human mobility modeling. Technical report, http://cnd.iit.cnr.it/chiara/techrep/wowmom12_tr.pdf, IIT-CNR 2011.

[13] Ari Keränen, Jörg Ott, and Teemu Kärkkäinen. The ONE Simulator for DTN Protocol Evaluation. In *SIMUTools '09*, 2009.

[14] K. Lee, S. Hong, S.J. Kim, I. Rhee, and S. Chong. Slaw: A new mobility model for human walks. In *IEEE INFOCOM 2009*, pages 855–863, 2009.

[15] M. Musolesi and C. Mascolo. Designing mobility models based on social network theory. *ACM SIGMOBILE CCR*, 11(3):59–70, 2007.

[16] M.E.J. Newman. The structure and function of complex networks. *SIAM review*, pages 167–256, 2003.

[17] M.E.J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

[18] G. Palla, I. Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.

[19] A. Passarella and M. Conti. Characterising aggregate inter-contact times in heterogeneous opportunistic networks. *NETWORKING 2011*, pages 301–313, 2011.

[20] C. Song, T. Koren, P. Wang, and A.L. Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, 2010.

[21] Q. Zheng, X. Hong, J. Liu, and D. Cordes. Agenda driven mobility modelling. *International Journal of Ad Hoc and Ubiquitous Computing*, 5(1):22–36, 2010.

[22] M. Zukerman. An introduction to queueing theory and stochastic teletraffic models, 2007.