Modelling Inter-contact Times in Social Pervasive Networks*

Andrea Passarella, Marco Conti, Chiara Boldrini
IIT-CNR
Pisa, Italy
{a.passarella,m.conti,c.boldrini}@iit.cnr.it

Robin I.M. Dunbar
Institute of Cognitive & Evolutionary Anthropology
University of Oxford, Oxford, UK
robin.dunbar@anthro.ox.ac.uk

ABSTRACT

Thanks to the diffusion of mobile user devices (e.g. smartphones) with rich computing and networking capabilities, we are witnessing an increasing integration between the cyber world of devices and the physical world of users. In this perspective, a possible evolution of pervasive networking (hereafter referred to as *social pervasive networks*, SPNs) consists in closely mapping human social structures in the network of the devices. Links between devices would correspond to social relationships between users, and communication events between devices would correspond to communications between users. It can be shown that fundamental convergence properties of SPN forwarding protocols are determined by the distributions of inter-contact times between the individual nodes (i.e. the time elapsed between two successive communication events between the nodes). Individual pairs inter-contact times are hard to completely charaterise, while the distribution of the *aggregate* inter-contact times is often a much more convenient figure. However, the aggregate distribution is not always representative of the individual pairs distributions. Therefore using it to characterise the properties of SPN forwarding protocols might not be correct. In this paper we provide an analytical model based on fundamental models of human social networks from the anthropology literature, which shows the exact dependence between the two in heterogeneous SPNs. Moreover, we use the model to i) study cases in which analysing the aggregate distribution is not enough, and ii) find sufficient conditions that guarantee that studying the aggregate distribution is enough to characterise the properties of SPN forwarding protocols.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]; C.2.1 [**Network Architecture and Design**]

*This work was funded by the European Commission under the SCAMPI (258414) FIRE project. The data were collected as part of the TESS project, funded by EPSRC. R.I.M. Dunbar's research is also funded by the British Academy Centenary Research project.

MSWiM'11, October 31–November 4, 2011, Miami, Florida, USA.

General Terms

Performance, Human Factors

Keywords

pervasive networks, social networks, modelling

1. INTRODUCTION

Last generation smartphones, tablets and similar pervasive devices feature extremely rich networking, computing, and sensing capabilities. It is nowadays argued in the research community that the penetration of this class of devices in the mass market is - for the first time - providing concrete grounds and a real opportunity for a massive deployment of pervasive networking applications [21]. Moreover, the fact that pervasive devices are almost constantly carried by users pushes towards the convergence of the "cyber" world, formed by the users' networked pervasive devices, and the "physical" world, formed by the users interacting with each other. In particular, an emerging design paradigm for pervasive networks consists in using off-line models and online information about the users' social behaviour to design, for example, routing [3, 19], data dissemination [4, 2], and mobile social networking [24] solutions. According to this paradigm, the human social plane (i.e., the structure and properties of social relationships between users) is translated into the cyber world to optimise the behaviour of pervasive networking systems.

The networking environment we consider in this paper, referred to as *social pervasive networks* (SPN), is a possible evolution of the pervasive networking paradigm enabled by this tight integration of the cyber and physical worlds. Assuming that the diffusion of pervasive technologies will enable, in principle, communication between any two users anytime and anywhere, the resulting network might in fact be formed by edges that correspond to communication channels activated because of a social relationship between two users, and only when those users communicate due to their social relationship. In other words, the network and the communication events between the devices might closely map the corresponding human social network and the interaction patterns of the users. Multi-hop communications will still occur in this type of networks, for example to enable diffusion of information among groups of people that not necessarily have mutual social relationships. Besides being a natural design approach, another advantage of such a design paradigm would be that activated communication channels will naturally inherit the trust level existing between their users, which is typically hard to assess in pervasive networks. Note that there is significant evidence suggesting that human social networks are almost invariant with respect to the specific technology that mediates social interactions [28]. Therefore, results in the anthropology domain that describe the properties of human social networks are already a solid starting point to investigate the properties of SPNs.

Within this scenario, the specific focus of this paper is to study some fundamental properties of inter-contact times between users. In SPNs, *contacts* are communications between two users due to a social interaction, and *inter-contact times* are time intervals between two consecutive contacts. Inter-contact times play a fundamental role for SPNs, as they have shown to do for a related networking environment, opportunistic networks [27]. In opportunistic networks, face-to-face contacts between users are exploited to forward messages. The foundational results presented in [9] highlight the impact of the distribution of inter-contact times on the convergence of opportunistic network routing protocols. Unlike in SPN, contacts in opportunistic networks require physical co-location of users. However, results in [9] hold for any network where messages can be exchanged only upon contacts between nodes, and therefore they apply also to SPNs. Specifically, [9] shows that when all inter-contact times between individual users follow a power law distribution with shape less than 2, then a large family of forwarding protocols (termed "naïve" protocols) diverge, i.e., yield infinite delay. In naïve protocols, nodes do not exploit any information describing the status of the network when taking forwarding decisions, and are only aware of the id of the destination. These protocols are attractive because they are very lightweight and simple to implement and analyse, and have been widely used in the literature [33, 16, 15, 31]. Note that the algorithm used in [15] to derive fundamental results on the capacity of opportunistic networks also falls in this category of forwarding protocols. Naïve forwarding protocols are attractive also for SPNs due to their simplicity and light cost. For example, they could be used for information dissemination and gossiping applications, where information should be spread to a large subset of users who will not necessarily share direct social relationships, and who will thus require multi-hop communication.

Although results in [9] apply to the distributions of *individual pairs* inter-contact times, it has been common in the literature [20, 6, 7, 29, 8, 5] to characterise opportunistic networks through the *aggregate* distribution of inter-contact times, i.e., the distribution of *all* inter-contact times between any two pairs considered altogether. Actually, using the aggregate distribution instead of all the distributions of individual pairs would be very convenient also in SPNs for a number of reasons:

- From a scalability standpoint, it is much less costly to compute, distribute, and store the parameters of a unique distribution than the parameters of all individual pairs distributions.

- From a statistical accuracy standpoint, much fewer samples are required to estimate with sufficient accuracy a unique aggregate distribution than all individual pairs distributions.

- From a privacy standpoint, it is much less sensitive to collect and distribute information about the aggregate distribution than about each individual pair, as from the former it is much harder to track individual users' behaviours.

Unfortunately, the aggregate distribution is in general not representative of the individual pairs distributions. Theoretically, the only case when it is representative is a completely homogeneous network, where all pairs inter-contact times are identically distributed, and thus the aggregate distribution is exactly the same as the distributions of individual pairs. However, for the reasons highlighted above, it is sensible to ask whether there are other cases in *heterogeneous* SPNs where studying the aggregate distribution is sufficient to characterise the convergence properties of forwarding protocols. To this end, it is necessary to have a clear understanding of the dependence between the individual pairs distributions and the aggregate distribution. Recently, [25] has analytically characterised this dependency for the case of opportunistic networks. In this paper we focus - instead - on the totally different scenario of SPN, where contacts do not require users mobility and physical co-location, but are driven by the structure of human social networks.

This paper provides the following contributions. We provide an analytical model showing the dependence between the inter-contact times distributions of individual pairs and the aggregate inter-contact time distribution in heterogeneous SPNs. Moreover, we highlight several cases of heterogeneous networks where considering the aggregate distribution is *not* sufficient to draw correct conclusions on the convergence properties of naïve forwarding protocols in SPNs. Specifically, we show cases where the aggregate distribution presents a power law, while all individual pairs distributions present a light tail. We highlight that, under certain conditions, this is also the case of one of the key datasets used in the anthropology literature to derive structural properties of human social networks [30]. Finally, we derive sufficient conditions for concluding that studying the aggregate distribution *is* sufficient to characterise the convergence properties of SPN naïve forwarding protocols.

The rest of the paper is organised as follows. In Section 2 we review the state-of-the-art relevant for this paper. Section 3 describes the models of human social networks available in the anthropology literature as the basis of our work. Section 4 presents the model showing the dependence between the inter-contact time distributions of individual pairs and the aggregate inter-contact time distribution. In Section 5 we use the model to analyse relevant cases of heterogeneous social pervasive networks. Finally, Section 6 concludes the paper.

2. RELATED WORK

This paper is mainly related to two bodies of work. The first one consists of the anthropology literature about models of human social networks. This body of work is described in detail in Section 3. The second body of work consists of the literature about the study of inter-contact times in opportunistic networks.

Results in [9] have demonstrated the fundamental impact of inter-conctat times on the convergence properties of opportunistic network routing protocols. As mentioned already, authors show that when the inter-contact times of individual pairs present a power law with shape less than 2, a large family of routing protocols yield infinite expected delay. [9] also analyses real traces of face-to-face inter-contact times, both originally presented in the paper and collected by others [23, 17, 32, 13]. Assuming that the network is homogeneous, authors focus on the distribution of aggregate inter-contact times, finding a good fit with a Pareto distribution with shape less than 2. These results challenge the actual applicability of popular routing protocols.

This view has been softned, to a certain extent, in [20], where authors have analysed the same traces of [9] (and, in addition, a proprietary GPS trace), noticing that the aggregate inter-contact times distribution actually presents an exponential cut-off in the tail. For what concerns the dependence between aggregate inter-contact times and the inter-contact times of individual pairs, [20] provides an initial result deriving analytically the dependence between them when the contact rates between individual pairs are known. In addition, [20] does not spend too much effort on the issue of heterogeneity, after noticing that, for a subset of the pairs in their traces, invidual inter-contact times are power law.

Results in [9, 20] had a very important impact on the subsequent literature, although not much attention has been put on the critical issue of heterogeneity. The fact that aggregate inter-contact times in popular traces present a power law has typically resulted in assuming that all distributions of individual pairs are power law. One of the most important examples is the area of mobility models. Most of the recent proposals (e.g., [6, 5, 22, 29]) aim at generating inter-contact times of individual pairs and/or aggregate inter-contact times following a power law. Similarly, other papers try to highlight which characteristics of reference mobility models generate a power law in inter-contact times [7, 8].

Authors of [10] analyse mathematically the dependence between inter-contact times of individual nodes and aggregate inter-contact times in a more general setting with respect to the model in [20]. They re-analyse the same traces used in [9, 20] showing that the distributions of inter-contact times of individual pairs are definitely heterogeneous. They propose a model to describe how heterogeneity impacts on the distribution of aggregate inter-contact times. However, as highlighted in [25], they miss to consider an important aspect, thus deriving an imprecise model. [25] presents the most precise model, as far as we know, to describe the dependence between the inter-contact time distributions of individual pairs and the aggregate inter-contact time distribution in opportunistic networking environments.

To the best of our knowledge, this is the first paper in the literature that analyses this dependence in social pervasive networks, considering models of interactions between users derived in the anthropology literature. With respect to [25], this results in a totally different model for describing the heterogeneity of inter-contact times of individual pairs. Moreover, in this paper, in addition to studying cases in which the aggregate inter-contact times distributions cannot be used to analyse the convergence of forwarding protocols, we also provide sufficient conditions under which focusing only on the aggregate distribution is enough.

3. HUMAN SOCIAL NETWORKS

Before presenting our analysis, it is worth describing our reference model for the structure of human social networks, which is based on the concept of ego network. An ego network is the network seen from the standpoint of a single individual (ego). It includes only other people (alters) the ego has social relationships with (represented by an edge in the ego network).

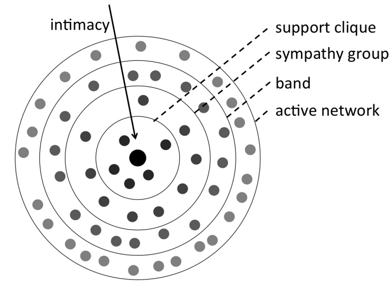

Figure 1: Ego-network's hierarchical structure.

Ego networks have been extensively studied in the anthropology literature [11, 12, 18, 30, 35], resulting in a detailed model of their structure (Figure 1). [35] has shown that ego networks can be represented as a series of concentric layers centred around the ego. Starting from the inner-most layer, layers are characterised by a decreasing level of *intimacy* with the ego. On the other hand, the *size* of the layers (the number of alters within the layer) increases with a factor approximately equal to 3. Extensive studies have identified four layers, i.e. the support clique, the sympathy group, the band and the active network, with size approximately equal to 5, 15, 45 and 150 [18, 12, 11]. The size of the active network (150) is usually referred to as the *Dunbar's number*, and represents the maximum number of alters an ego can - on average - maintain social relationships with [18]. This is a limit related to cognitive capabilities of the human brain [12]. Many more alters can be outside the active network, corresponding to people known to the ego, but with whom the ego does not establish any significant social relation. These alters are usually not represented in the model. Note that this hierarchical structure depends very little on the communication means supporting social relationships [28].

Authors of [18] have also shown that the *emotional closeness* of the ego with a given alter is the key parameter determining the position of the alter in the layers. Moreover [18, 30] show that there is a strong correlation between the emotional closeness and the frequency of communication between the ego and the alter. Therefore, it follows that the structure of the ego network depicted in Figure 1 naturally determines the contact rates between the ego and alters in its social network. Specifically, contacts are more frequent with alters in the inner-most layer (usually referred to as *strong ties*), while the frequency progressively declines for external layers, resulting in *weaker* ties. This property is one of the starting points of the analysis presented in Section 4.

Finally, it is worth pointing out that, for our purposes, focusing on ego networks is sufficient. In general a social network contains more information than the set of ego networks of its members, as the latter does not capture correlations. However, it is straightforward to note that inter-contact times between any pair of users can be fully described by looking at ego networks only, because they only depend on the relationship between these two users, which is captured by the ego-network model.

4. INTER-CONTACT TIMES MODEL

In this section we study, through an analytical model, the dependence between the distributions of the individual pairs and aggregate inter-contact times, in a network where contacts can be described with the ego-network model presented in Section 3.

An important requirement of our model is to represent heterogeneous networks in which the distributions of inter-contact times between individual pairs are not iid. We take heterogeneity into account in the definition of the model for contact *rates* (the reciprocal of the average inter-contact times). We assume that the contact rates are random variables (r.v.) following a known distribution (hereafter Λ_p denotes the contact rate of the generic pair p). In addition, we assume that individual pairs inter-contact times are distributed according to a known *type* of distribution (e.g., Pareto, exponential, ...). For each pair p, the parameters of the inter-contact times distribution are a function of Λ_p, i.e., the parameters are set such that the average inter-contact time is equal to $1/\Lambda_p$. This allows us to model heterogeneous environments in which not all individual inter-contact times are identically distributed, and to control the type of heterogeneity through the r.v. describing the contact rates.

Therefore, three distributions play a key role in our analysis, i.e. i) the distributions of individual pairs inter-contact times (whose CCDF is hereafter denoted as $F_\lambda(x)$), ii) the distribution of individual pairs contact rates (whose density is hereafter denoted as $f(\lambda)$), and iii) the distribution of the aggregate inter-contact times (whose CCDF is hereafter denoted as $\mathcal{F}(x)$).

4.1 Modelling human networks contact patterns

As a first step in the model, we describe how we account for the human social network structures described in Section 3. This is taken into consideration in the definition of the distribution of the contact rates. Figure 2 provides a schematic representation of a generic distribution. As, in any given ego network, contacts with alters in inner shells occur more frequently than contacts with alters in outer shells, contact rates with peers in the inner-most shell should be drawn from the tail of the distribution, while contact rates with peers in the outer-most shell should be drawn from the head. Based on this observation, we divide the possible range of rates in L sectors, where L is the number of layers of an ego network, and layer 1 denotes the inner-most layer. The challenge is to meaningfully identify in the contact rate distribution the boundaries of the sectors corresponding to each layer or, in other words, to define the sectors of the contact rate distribution from where to draw contact rate samples for alters in any given social layer. The average number of relationships in each layer $n_l, l = 1, \dots, L$, and the total number of relationships N can be derived from the results presented in Section 3 [18, 12, 11]. We can thus compute the fraction of relationships in each layer as n_l/N (note that $n_L = N$). If we denote with $\lambda_0, \dots, \lambda_L$ the values of λ that identify the sectors of the contact rates distribution corresponding to the layers, the values of $\lambda_i, i = 1, \dots, L$ can be computed as the $(1 - \frac{n_l}{N})$-th percentiles of the rates distribution (note that λ_L and λ_0 are the minimum and maximum possible values of λ, respectively). Therefore, contact rates with a peer in layer $l = 1, \dots, L$ are drawn from the sector identified by λ_l, λ_{l-1}. It thus follows that the density of contact rates for relationships in layer l is as follows

$$f_l(\lambda) = \begin{cases} 0 & \lambda < \lambda_l \vee \lambda > \lambda_{l-1} \\ C_l f(\lambda) & \lambda_l \le \lambda \le \lambda_{l-1} \end{cases} \tag{1}$$

where C_l is a constant such that $\int_0^\infty f_l(\lambda) d\lambda = 1$, i.e. $C_l = [G(\lambda_{l-1}) - G(\lambda_l)]^{-1}$, $G(\lambda)$ being the CDF of Λ.

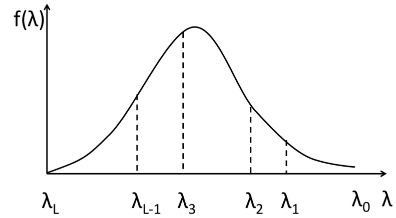

Figure 2: A representative contact rates distribution in human social networks

Note that we only consider the distribution of contact rates for alters with a contact rate greater than 0. In principle, the distribution of contact rates presents a significant mass probability in 0, corresponding to the fact that an ego "knows" alters also outside the active network layer, but relationships are so weak that the contact rate is zero.

4.2 General inter-contact times model

The starting point of our model is a result originally presented in [20] (and recalled in Lemma 1), which describes the dependence between the distributions of the individual pairs and aggregate inter-contact times, *when the contact rates are known a priori.* Let us assume that P pairs are present in the network, that $n_p(T)$ contact events between pair p occur during an observation time T. Let us denote with $N(T)$ the total number of contact events over T, with θ_p the contact rate of pair p, with θ the total contact rate ($\theta = \sum_p \theta_p$), and with $F_p(x)$ the CCDF of inter-contact times of pair p. Then, the following lemma holds.

LEMMA 1. *In a network where P pairs of nodes exist for which inter-contact times can be observed, the CCDF of the aggregate inter-contact times is:*

$$\mathcal{F}(x) = \lim_{T\to\infty} \sum_{p=1}^{P} \frac{n_p(T)}{N(T)} F_p(x) = \sum_{p=1}^{P} \frac{\theta_p}{\theta} F_p(x) \tag{2}$$

Lemma 1 is rather intuitive. The distribution of aggregate inter-contact times is a mixture of the individual pairs distributions. Each individual pair "weights" in the mixture proportionally to the number of inter-contact times that can be observed in any given interval (or, in other words, proportionally to the contact rate).

In this paper we significantly extend this result, by i) assuming that contact rates are random variables, thus unknown a priori, and ii) exploiting an anthropology model for describing contacts between humans. Specifically, we can derive the following Theorem[1].

THEOREM 1. *In a social pervasive network where contact rates are determined by the hierarchical structure of ego networks, the CCDF of the aggregate inter-contact times is:*

$$\mathcal{F}(x) = \sum_{l=1}^{L} \frac{p_l C_l}{\sum_{l=1}^{L} p_l E[\Lambda_l]} \int_{\lambda_l}^{\lambda_{l-1}} \lambda f(\lambda) F_\lambda(x) d\lambda \tag{3}$$

where p_l is the probability that a social relationship of any given user is in layer l of its ego network, and Λ_l denotes the contact rates between an ego and its alters in layer l (i.e., its density is as in Equation 1).

[1]Unless otherwise stated, proofs are provided in the [26].

While in [26] we provide the complete proof of Equation 3, here we briefly discuss its physical meaning. First of all, Equation 3 can be seen as the weighted sum of components related to the individual layers of ego networks. Specifically, by defining $\mathcal{F}_l(x)$ as follows:

$$\mathcal{F}_l(x) = \frac{C_l}{E[\Lambda_l]} \int_{\lambda_l}^{\lambda_{l-1}} \lambda f(\lambda) F_\lambda(x) d\lambda \quad (4)$$

we can write $\mathcal{F}(x)$ as

$$\mathcal{F}(x) = \sum_{l=1}^{L} \frac{p_l E[\Lambda_l]}{\sum_{l=1}^{L} p_l E[\Lambda_l]} \mathcal{F}_l(x) \quad (5)$$

In [26] we show that $\mathcal{F}_l(x)$ is actually the CCDF of the aggregate inter-contact times over layer l only. Equation 5 highlights an intuitive result. Each such component ($\mathcal{F}_l(x)$) "weights" in the aggregate proportionally to the fraction of pairs falling in the layer (p_l), and to the average contact rates of the layer (i.e., to the average number of inter-contact events that is generated by a pair in that layer).

Besides a more formal derivation shown in [26], the form of the individual layer CCDF in Equation 4 has a more intuitive derivation, starting from the result in Lemma 1. Specifically, it can be obtained by considering a modified network in which we assume that all rates $\lambda \in [\lambda_l, \lambda_{l-1}]$ are possibly available (for pairs in layer l), each with a probability $f_l(\lambda)d\lambda$. $\mathcal{F}_l(x)$ is thus the aggregate over all the resulting individual pairs inter-contact times distributions. As the number of such distributions becomes infinite and is indexed by Λ_l (a continuous random variable), the summation in Equation 2 becomes an integral over λ. Moreover, the weight of each distribution (θ_p in Equation 2) becomes $\lambda \cdot p(\lambda) = \lambda f_l(\lambda) d\lambda$, while the total rate ($\theta$ in Equation 2) becomes $\int_0^\infty \lambda f_l(\lambda) d\lambda = E[\Lambda_l]$. The expression in Equation 4 follows immediately.

5. STUDY OF REPRESENTATIVE SOCIAL PERVASIVE NETWORKS

The following analysis starts from the result in Theorem 1, and is divided in two parts. In the first part (Section 5.1) we show relevant cases where the aggregate inter-contact times distribution is not representative of the distributions of the individual pairs, and thus studying all individual pairs distributions is necessary to characterise the convergence properties of SPN forwarding protocols. Specifically, we identify by analysis cases where the individual distributions present a light tail, while the aggregate turns out to follow a power law. Interestingly, under the conditions discussed in Section 5.1, we show that the human social network used to collect the data at the basis of [30] falls in this category. In the second part (Section 5.2), we derive analytically sufficient conditions for concluding that the aggregate inter-contact times distribution (instead of all the individual pairs distributions) can be used to characterise the convergence properties of SPN forwarding protocols. Specifically we prove that it is sufficient that the distribution of even a single pair follows a power law for the aggregate distribution to also be power law. We thus conclude that when the aggregate is *not* power law, then all individual pairs distributions must present a light tail, and therefore the network does not satisfies the conditions for forwarding protocols divergence found in [9].

It is important to note that, to carry on our analysis, it is sufficient to study the aggregate inter-contact times distribution over individual layers only, provided by Equation 4. It is, in fact, sufficient that one such aggregate presents a heavy tail for the whole aggregate to be heavy tailed. Thus, Equation 4 is the key starting point for the following analysis.

5.1 Networks where analysis of individual pairs distributions is needed

The first network we consider is one where individual pairs distributions are exponential, and the contact rates follow a gamma distribution. Being all individual pairs distributions exponential, this network does not satisfy the conditions for divergence of forwarding protocols found in [9]. Note that assuming exponential individual pairs distributions is relevant. At least in the case of face-to-face contacts, real traces analysis has shown that the distributions of a high fraction of pairs are exponential [14, 10].

Considering contact rates following a gamma distribution is motivated by the analysis of the dataset used to derive the properties of ego networks described in [30], at the basis of the model described in Section 3. The dataset collects information about 251 ego networks. Each relationship in each ego-network provides a sample of contact rate, for a total of over 20000 samples. We fit the resulting empirical distribution to reference distributions (i.e., gamma, exponential and Pareto) using the Maximum Likelihood (ML) method [34], and compare the fitted distributions against the data using the Akaike Information Criterion (AIC, [1]). Figure 3 shows a visual comparison of the samples obtained from [30] and the ML fittings of the considered contact rates distributions (ML estimators of the parameters are provided in Table 1). As for the gamma and exponential distributions we consider the standard definitions with shape α and rate b, and with rate b, respectively. As for the Pareto distribution we consider the CCDF form $F(\lambda) = \left(\frac{b}{b+\lambda}\right)^\alpha, \alpha > 0, \ \lambda > 0$.

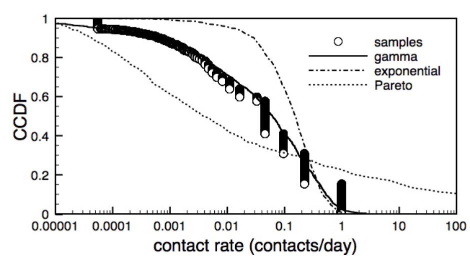

Figure 3: Fitting distributions

The intuition from Figure 3 is that the gamma distribution is the best fit for our dataset. This is confirmed by the AIC test, whose values are shown in Table 1. Remember that in AIC tests the best alternative is the one with the lowest AIC value [1].

Distribution	Best fit parameters	AIC value
Gamma	$\alpha = 0.34$, $b = 1.63$	-50280.62
Exponential	$b = 4.86$	-23505.08
Pareto	$\alpha = 0.16$, $b = 5.5\text{x}10^{-5}$	-31289.34

Table 1: AIC values for the tested distributions.

Lemma 2 and Theorem 2 characterise the distribution of the aggregate inter-contact times in a network where individual pairs follow an exponential distribution, and contact rates follow a gamma distribution.

LEMMA 2. *When contact rates follow a gamma distribution and individual inter-contact times an exponential distribution, the CCDFs of inter-contact times aggregated over individual layers ($\mathcal{F}_l(x)$) all decay, for large x, faster than a power law with exponential cutoff, but the CCDF corresponding to the outer-most layer, which decays as a power law. Specifically, if the contact rates follow a gamma distribution with shape α and rate b, the following relations hold true, for large x:*

$$\begin{cases} \mathcal{F}_l(x) \leq \frac{Re^{-\lambda_l(b+x)}}{x} & l = 1, \ldots, L-1 \\ \mathcal{F}_L(x) \simeq \frac{K}{x^{\alpha+1}} \end{cases} \tag{6}$$

where R and K do not depend on x.

THEOREM 2. *In a social pervasive network where individual pairs inter-contact times are exponentially distributed and contact rates follow a gamma distribution, the distribution of the aggregate inter-contact times features a heavy tail. Specifically, the following relation holds true:*

$$f(\lambda) = \frac{\lambda^{\alpha-1} b^{\alpha} e^{-b\lambda}}{\Gamma(\alpha)}, F_\lambda(x) = e^{-\lambda x}$$
$$\Rightarrow \mathcal{F}(x) \simeq \frac{K}{x^{\alpha+1}} \text{ for large } x$$

where K does not depend on x.

PROOF. This follows immediately from Lemma 2, by recalling the relationships between $\mathcal{F}_l(x)$ and $\mathcal{F}(x)$ in Equation 5, and noting that $\mathcal{F}_L(x)$ dominates over all the other components for large x. □

To validate the analytical results, we compare the result of Theorem 2 with simulations. Specifically, we simulate an ego-network with 150 alters divided in layers according to the model presented in Section 3. Ego and each alter meet with exponential inter-contact times, with rates drawn from a gamma distribution. Sectors of the distribution corresponding to the layers are defined as described in Section 4.1. Each simulation run reproduces an observation of the network for a time interval T, defined according to the following algorithm. For each alter a, we first generate 100 inter-contact times, and then compute the total observation time after 100 inter-contact times, T_a, as the sum of the pair inter-contact times. T is defined as the maximum of $T_a, a = 1, ..., 150$. To guarantee that all alters are observed for the same amount of time, we generate additional inter-contact times for each alter until T_a reaches T. Simulations have been replicated 20 times with independent seeds, and confidence intervals (with 99% confidence level) have been computed.

Figure 4 shows a very good agreement between the analytical and the simulation models. Recall that the analysis predicts that the tail of the aggregate inter-contact times distribution decays as $\frac{1}{x^{\alpha+1}}$ where α is the shape parameter of the contact rates distribution. Figure 4 shows that - as also found in the analysis - the lower the shape of the contact rates distribution, the heavier the tail of the aggregate inter-contact times. This results from the fact that lower shape parameters result in a higher mass of probability of contact rates around 0, i.e., in an increasing probability of very long inter-contact times.

Theorem 2 and Lemma 2 provide two interesting insights. First, the presence of aggregate inter-contact times with a

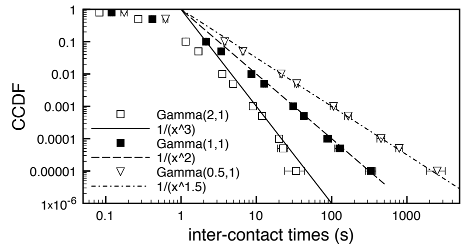

Figure 4: Aggregate inter-contact times with gamma contact rates

heavy tail distribution does not necessarily mean that routing protocols risk divergence in SPNs, as such a heavy tail can emerge starting from exponentially distributed individual pairs. Therefore, when the contact rates follow a gamma distribution, looking at the distribution of aggregate inter-contact times is not sufficient to check whether routing protocols may diverge or not. Instead, the distributions of individual pairs inter-contact times must be analysed. Second, the power law of $\mathcal{F}(x)$ appears because of the power law of the inter-contact times aggregated over the outer-most layers, $\mathcal{F}_L(x)$. Due to the shape of the gamma distribution, in the outer-most layers contact rates can be arbitrarily close to 0, thus resulting in arbitrarily large inter-contact times. Intuitively, this suggests a more general behaviour: Whenever the distribution of the contact rates is such that rates arbitrarily close to 0 can be drawn, the distribution of the aggregate inter-contact times presents a heavy tail.

The case of contact rates following a gamma distribution is not the only one in which these two properties hold true. Actually, the case where rates follow a Pareto distribution also present similar properties (still assuming that individual inter-contact times follow an exponential distribution). Lemma 3 and Theorem 3 analyse this case.

LEMMA 3. *When contact rates follow a Pareto distribution whose CCDF is in the form $F(\lambda) = \left(\frac{b}{b+\lambda}\right)^{\alpha}$, $\lambda > 0$ and individual inter-contact times are exponential, the CCDFs of inter-contact times aggregated over individual layers ($\mathcal{F}_l(x)$) all decay, for large x, at least as fast as a power law with exponential cutoff, but the CCDF corresponding to the outer-most layer, which decays as a power law. Specifically, the following relations hold true for large x:*

$$\begin{cases} \mathcal{F}_l(x) \leq \frac{Re^{-\lambda_l x}}{x} + \frac{Qe^{-\lambda_{l-1} x}}{x} & l = 1, \ldots, L-1 \\ \mathcal{F}_L(x) \simeq \frac{K}{x^2} \end{cases} \tag{7}$$

where R, Q and K do not depend on x.

THEOREM 3. *When contact rates follow a Pareto distribution whose CCDF is in the form $F(\lambda) = \left(\frac{b}{b+\lambda}\right)^{\alpha}$, $\lambda > 0$ and individual inter-contact times are exponential, the CCDF of the aggregate inter-contact times decays, for large x, as a power law with shape equal to 2. Specifically, the following relation holds true*

$$F(\lambda) = \left(\frac{b}{b+\lambda}\right)^{\alpha}, F_\lambda(x) = e^{-\lambda x}$$
$$\Rightarrow \mathcal{F}(x) \simeq \frac{K}{x^2} \text{ for large } x$$

where K does not depend on x.

PROOF. This comes immediately from Lemma 3 by noticing that the slowest decaying component of $\mathcal{F}(x)$ is the one corresponding to the outer-most layer, and using the corresponding expression from Equation 7. □

Also in this case, we validate the results by comparing analysis and simulation. Figure 5 shows that also in this case the analytical results capture very well the shape of the tail of the aggregate inter-contact times distribution obtained in simulation. Note that, as predicted by the model, the aggregate distribution decays as a power law with shape equal to 2, irrespective of the parameters of the contact rates distribution.

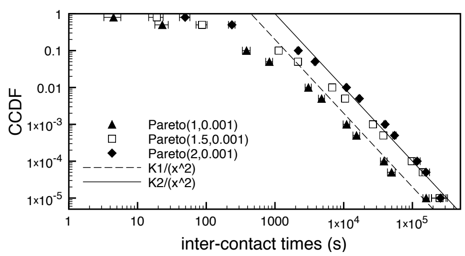

Figure 5: Aggregate inter-contact times with Pareto contact rates.

As anticipated, also in this case considering the aggregate inter-contact times distribution is not sufficient to characterise the convergence of forwarding protocols. Note that, also in this scenario, contact rates can be arbitrarily close to 0. This confirms the intuition that when contact rates present this property, the aggregate distribution of inter-contact times presents a heavy tail.

5.2 The effect of individual power-law inter-contact times distributions

In this section we study analytically the effect of even a single heavy tail individual inter-contact times distribution on the distribution of the aggregate inter-contact times. Specifically, we assume that the inter-contact times distribution of a given pair p presents a heavy tail, i.e. we assume that $F_p(x)$ is as follows:

$$F_p(x) \simeq x^{-\eta} \text{ for large x}$$

For the rest of the individual pairs distributions we consider the same assumptions used in Section 4, and, in addition, we assume that they are not power law. In other words, the distribution of pair p is the only one in the network presenting a heavy tail. Finally, we assume that the number of pairs is finite. Then, the following lemma holds true.

LEMMA 4. *In a network with a finite number of pairs, where there exists one pair whose individual inter-contact times distribution follows, for large x, a power law with shape η, the distribution of the aggregate inter-contact times, for large x, follows a power law at least as heavy as $x^{-\eta}$, i.e.*

$$\exists\ p \text{ s.t. } F_p(x) \simeq x^{-\eta} \text{ for large } x \Rightarrow$$
$$\mathcal{F}(x) \geq Cx^{-\eta} \text{ for large } x \text{ and for some constant } C > 0$$

Figure 6 provides a concrete example of the result in Lemma 4. Specifically, we first consider an ego-network such that the distribution of the aggregate inter-contact times does not present a heavy tail. As shown in [26], this can be obtained, for example, by using exponentially distributed inter-contact times, and sampling the contact rates from a slightly different Pareto distribution with respect to the one considered in Section 5.1. Specifically, it is necessary to consider a CCDF in the form $F(\lambda) = \left(\frac{b}{\lambda}\right)^{\alpha}, \alpha > 0,\ \lambda > b$. In this case, the tail of the resulting aggregate distribution presents a power law with an exponential cut-off. In the Figure, the percentiles obtained by simulation are marked with white squares, and the corresponding analytical curve (derived in [26]) is plotted with a solid line. Then, we added to the same ego network one more alter, whose inter-contact times with the ego follow a Pareto distribution with shape $\eta = 1.1$ (while the scale parameter, defining the minimum inter-contact time, is set to 1 day, as this was also the minimum inter-contact time found in the traces used in Section 5.1). In the figure, percentiles obtained by simulation are marked with dark diamonds, while the corresponding analytical curve predicted by Lemma 4 is plotted with a dashed curve. According to Lemma 4, i) the existence of even a single pair whose inter-contact times are power law implies that the tail of the aggregate distribution is also heavy, and ii) the tail of the resulting aggregate distribution can be lower bounded by a power law with shape equal to $\eta = 1.1$. Figure 6 confirms both results obtained in Lemma 4.

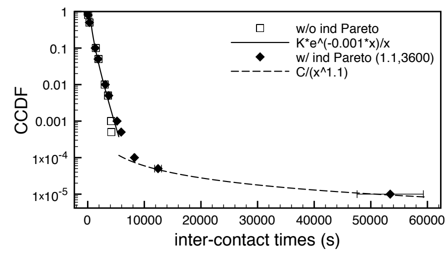

Figure 6: Aggregate inter-contact times with and without a single Pareto ICT pair.

Lemma 4 allows us to immediately identify the cases where considering the aggregate inter-contact times distribution is sufficient to characterise the convergence properties of forwarding protocols in SPNs. Specifically, the following theorem holds true.

THEOREM 4. *In a network with a finite number of pairs, if the distribution of the aggregate inter-contact times does not present a heavy tail, then no individual inter-contact times distribution can present a heavy tail.*

PROOF. This comes straightforwardly from Lemma 4. If even a single individual inter-contact times would present a heavy tail, then the aggregate distribution would also present a heavy tail. □

In practical terms, Theorem 4 tells that when the aggregate inter-contact times distribution does not present a heavy tail, it is not necessary to study all the distributions of individual inter-contact times to check the conditions on the convergence of forwarding protocols found in [9], because no individual inter-contact times distribution can follow a power law. This result is dual to those found in Section 5.1, which tell that, instead, when the aggregate distribution presents a heavy tail, a detailed analysis of the individual pairs distributions is necessary.

6. CONCLUSION

In this paper we studied fundamental properties of inter-contact times in social pervasive networks, basing our analysis on reference models of human social networks available in the anthropology literature. Social pervasive networks are a possible evolution of current pervasive networks, where the communication network maps directly the human social network of users, and communication between devices occurs when users communicate as an effect of their social tie.

In social pervasive networks, forwarding actions occur upon such contact events between users. In this scenario, it is fundamental to characterise the properties of inter-contact times between individual users (i.e., the time between two consecutive contact events), as this has been shown to play a key role in determining convergence properties of forwarding algorithms when forwarding occurs upon contacts only. A complete characterisation of all individual inter-contact times distribution might be impractical to achieve. From this standpoint, using the aggregate distribution of inter-contact times would be much more convenient. Unfortunately, in general the aggregate distribution is not representative of the distributions of individual pairs. A clear understanding of the dependence between the two is thus needed.

In this paper we provided a mathematical model based on models of human social networks available in the anthropology literature to formally characterise the dependence between the individual pairs distributions and the distribution of aggregate inter-contact times. The model highlights the importance of the network heterogeneity (captured through the distribution of contact rate) in determining the shape of the aggregate distribution. We have used the model to study relevant networks in which, unfortunately, focusing on the aggregate distribution is not enough, and all individual pairs distributions must be analysed to characterise the convergence properties of forwarding algorithms. Interestingly, the model allowed us to also find sufficient conditions to be sure that considering the aggregate distribution is enough.

Beyond the specific applications presented, the contribution of the paper consists in providing a clear understanding of the dependence between the different distributions of inter-contact times in SPNs, and a practical tool to understand which statistics must be used to correctly study the convergence properties of SPN forwarding protocols.

7. REFERENCES

[1] H. Akaike. A new look at the statistical model identification. *IEEE Tran. on Aut. Control*, 19(6):716 – 723, 1974.

[2] S. Allen, M. Chorley, G. Colombo, and R. Whitaker. Self adaptation of cooperation in multi-agent content sharing systems. In *IEEE SASO*, 2010.

[3] C. Boldrini, M. Conti, and A. Passarella. Exploiting users' social relations to forward data in opportunistic networks: The HiBOp solution. *Pervasive and Mobile Computing*, 4(5):633 – 657, 2008.

[4] C. Boldrini, M. Conti, and A. Passarella. Design and performance evaluation of ContentPlace, a social-aware data dissemination system for opportunistic networks. *Comput. Netw.*, 54(4):589–604, March 2010.

[5] 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.

[6] 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.

[7] H. Cai and D. Y. Eun. Crossing over the bounded domain: from exponential to power-law inter-meeting time in manet. In *ACM MOBICOM*, pages 159–170, 2007.

[8] H. Cai and D. Y. Eun. Toward stochastic anatomy of inter-meeting time distribution under general mobility models. In *ACM MobiHoc*, pages 273–282, 2008.

[9] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Trans. Mob. Comput.*, 6(6):606–620, 2007.

[10] V. Conan, J. Leguay, and T. Friedman. Characterizing pairwise inter-contact patterns in delay tolerant networks. In *Autonomics*, 2007.

[11] R. Dunbar and M. Spoors. Social networks, support cliques, and kinship. *Human Nature*, 6(3):273–290, 1995.

[12] R. I. M. Dunbar. The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, 6(5):178–190, 1998.

[13] N. Eagle and A. (Sandy) Pentland. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10:255–268, March 2006.

[14] W. Gao, Q. Li, B. Zhao, and G. Cao. Multicasting in delay tolerant networks: a social network perspective. In *ACM MobiHoc*, 2009.

[15] M. Grossglauser and D. N. C. Tse. Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Trans. Netw.*, 10:477–486, August 2002.

[16] Z. J. Haas and T. Small. A new networking model for biological applications of ad hoc sensor networks. *IEEE/ACM Trans. Netw.*, 14:27–40, February 2006.

[17] T. Henderson, D. Kotz, and I. Abyzov. The changing usage of a mature campus-wide wireless network. In *ACM MOBICOM*, 2004.

[18] R. Hill and R. Dunbar. Social network size in humans. *Human Nature*, 14(1):53–72, 2003.

[19] P. Hui, J. Crowcroft, and E. Yoneki. Bubble rap: Social-based forwarding in delay tolerant networks. *IEEE Trans. Mob. Comput.*, 2010.

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

[21] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell. A survey of mobile phone sensing. *IEEE Comm. Mag.*, 48(9):140 –150, 2010.

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

[23] M. McNett and G. M. Voelker. Access and mobility of wireless pda users. *ACM MC2R*, 9, 2005.

[24] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell. Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application. In *ACM SenSys*, 2008.

[25] A. Passarella and M. Conti. Characterising aggregate inter-contact times in heterogeneous opportunistic networks. In *IFIP Networking*, 2011.

[26] A. Passarella, M. Conti, C. Boldrini, and R. Dunbar. Modelling inter-contact times in human Social Pervasive Networks. Technical report, IIT-CNR-08, 2011, http://cnd.iit.cnr.it/andrea/docs/mswim11-tr.pdf.

[27] L. Pelusi, A. Passarella, and M. Conti. Opportunistic networking: data forwarding in disconnected mobile ad hoc networks. *IEEE Comm. Mag.*, 44(11):134 –141, 2006.

[28] T. Pollet, S. Roberts, and R. Dunbar. Use of social network sites and instant messaging does not lead to increased social network size, or to emotionally closer relationships with offline network members. *Cyberpsychology, Behavior, and Social Networking*, 2010.

[29] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong. On the levy-walk nature of human mobility. In *INFOCOM*, pages 924–932, 2008.

[30] S. Roberts and R. Dunbar. Communication in social networks: Effects of kinship, network size, and emotional closeness. *Personal Relationships*, 2010.

[31] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Efficient routing in intermittently connected mobile networks: the multiple-copy case. *IEEE/ACM Trans. Netw.*, 16:77–90, February 2008.

[32] J. Su, A. Chin, A. Popivanova, A. Goel, and E. de Lara. User mobility for opportunistic ad-hoc networking. In *IEEE MSCA*, 2004.

[33] A. Vahdat and D. Becker. Epidemic routing for partially connected ad hoc networks. Technical report, Duke University, 2000.

[34] L. Wasserman. *All of statistics: a concise course in statistical inference*. Springer Verlag, 2004.

[35] W. X. Zhou, D. Sornette, R. A. Hill, and R. I. M. Dunbar. Discrete hierarchical organization of social group sizes. *Proceedings of the Royal Society B: Biological Sciences*, 272(1561):439–444, 2005.