Consiglio Nazionale delle Ricerche

Modelling inter-contact times in human social pervasive networks

A. Passarella, M. Conti, C. Boldrini, R. Dunbar

IIT TR-08/2011

Technical report

giugno 2011

Istituto di Informatica e Telematica

Modelling inter-contact times in human social pervasive networks

Technical Report

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, UK
robin.dunbar@anthro.ox.ac.uk

ABSTRACT

Thanks to the diffusion of mobile user devices (e.g. smartphones) with rich computation 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 (throughout referred to as *social pervasive networks*, SPNs) might consist 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 PSN 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 characterise, 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, such that using it to characterise the properties of PSN forwarding protocols might not be correct. In this paper we provide an analytical model showing the exact dependence between the two in heterogeneous SPNs. Moreover, we use the model to i) study cases in which studying 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 PSN forwarding protocols.

Categories and Subject Descriptors

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

General Terms

Performance

Keywords

social pervasive networks, human social networks, modelling, inter-contact times

1. INTRODUCTION

Last generation smartphones, tablets and similar pervasive devices feature extremely rich networking, computation 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 [4, 19], data dissemination [5, 3], 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* (PSN), 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. 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 [27]. Therefore, current 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 studying 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 [26]. In opportunistic networks, face-to-face contacts between users are exploited to forward messages. The foundational results presented in [10] highlight the impact of the distribution of inter-contact times on the convergence of opportunistic network routing protocols. Unlike in PSN, contacts in opportunistic networks require physical co-location of users. However, results in [10] hold for any network where messages can be exchanged only upon contacts between nodes, and therefore they apply also to SPNs. Specifically, [10] shows that when the all inter-contact times between individual users follow a power law distribution with shape less than 2, then a large family of forwarding protocols diverge, i.e. yield infinite delay. Note that the algorithm used in [16] to derive fundamental results on the capacity of opportunistic networks also fall in this category of forwarding protocols. Although results in [10] apply to the distributions of *individual pairs* inter-contact times, it has been common in the literature [20, 7, 8, 28, 9, 6] 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 present an analysis similar to that presented in [25], focusing - however - on the totally different scenario of PSN where contact do not require users mobility and physical co-location, but are completely 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 time 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 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 [29]. Finally, we derive sufficient conditions for concluding that studying the aggregate distribution *is* sufficient to characterise the convergence properties of PSN 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 at 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 in 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 in the literature about the study of inter-contact times in opportunistic networks.

Results in [10] 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 parameter lower than 2, a large family of routing protocols yield infinite delay. [10] also analyses real traces of face-to-face inter-contact times, both originally presented in the paper and collected by others [23, 17, 30]. Assuming that the network is homogeneous, it focuses on the distribution of aggregate inter-contact times, finding a good fit with a Pareto distribution with shape parameter lower than 2. These results posed an important warning in the opportunistic networking community, casting a rather pessimistic view on 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 [10] (and, in addition, the well known Reality Mining trace [14]), noticing that the aggregate inter-contact time 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 the two distributions when the contact rates between individual pairs are known. In addition, [20] does not spend too much effort on the issue of heterogeneity in the inter-contact times of individual pairs, after noticing that, for a subset of the pairs in their traces, invidual inter-contact times are power law.

Results in [10, 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 for opportunistic networks. Most of the recent proposals (e.g., [7, 6, 22, 28]) 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 of individual pairs [8, 9].

Authors of [11] have analysed 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-analysed the same traces used in [10, 20] focusing much more than previous papers on the analysis of inter-contact times of individual pairs. They show that the distributions of inter-contact times of individual pairs are definitely heterogeneous. They propose a model to describe how heterogeneity impact 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. To the best of our knowledge [25] presents the most precise model to describe the dependence between the inter-contact time distribution 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, based on quantitative surveys with users. With respect to [25], this results in a totally different model for describing the heterogeneity of inter-contact times of individual pairs. While the line of reasoning for deriving the model is similar, the model itself turns out to be significantly different because of the different networking environment.

3. HUMAN SOCIAL NETWORKS

Before presenting our analysis, it is worth describing our reference model for the structure of human social networks. We consider a particular model of human social networks, 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).

Ego networks have been extensively studied in the anthropology literature [12, 13, 18, 29, 31], resulting in a detailed model of their structure (Figure 1). [31] 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, 13, 12]. 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 [13]. Many more alters can be outside the active network, corresponding to people known to the ego, but with whom the ego do 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 [27].

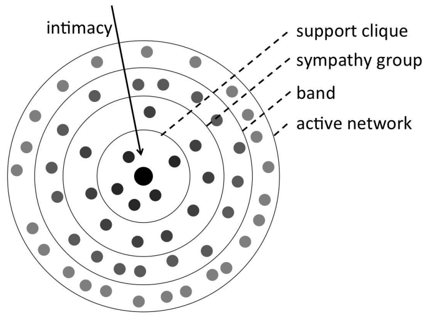

Figure 1: Ego-network's hierarchical structure.

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, 29] 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

Before deriving 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 contact rates distribution. 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, 13, 12]. 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

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

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.

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 "knowns" alters also outside the active network layer, but relationships are so weak that the contact rate is zero.

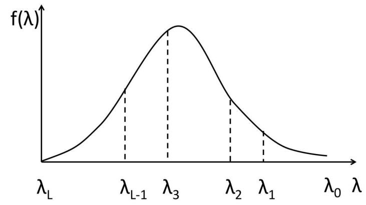

Figure 2: A representative contact rates distribution in human social networks

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 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:*

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

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 rate of inter-contact times).

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

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

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

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.

Proof. See Appendix A.

While in Appendix A we provide the complete proof of Equation 1, 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 human social networks. Specifically, by defining $\mathcal{F}_l(x)$ as follows:

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

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

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

In appendix A 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 Appendix A, 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.

Theorem 1 shows the dependence between the three distributions that characterise the properties of inter-contact times. The key property we study in the following is under which conditions, and starting from which distributions of individual inter-contact times and contact rates, the distribution of aggregate inter-contact times presents a heavy tail. This provides a solid mathematical foundation to understand whether focusing on the aggregate inter-contact times is sufficient or not for assessing the convergence properties of routing protocols in SPNs.

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. STUDY OF REPRESENTATIVE SOCIAL PERVASIVE NETWORKS

In this section we exploit the model derived in Section 4 to study two different aspects. First, in Sections 5.1 and 5.2, we study how the inter-contact *rates* impact on the distribution of the aggregate inter-contact times when individual pairs inter-contact times are exponentially distributed. This analysis follows the footsteps of [25], and highlights important cases in which the existence of aggregate inter-contact times following a power-law is *not* a sufficient condition for forwarding protocols divergence. Second, in Section 5.3, we study analytically the impact on the aggregate inter-contact times distribution of even a single pair of users with power-law inter-contact times. This analysis allows us to derive sufficient conditions on the properties of the aggregate distribution to conclude that forwarding protocols will *not* diverge. All in all, these results provides concrete guidelines on how to interpret the properties of the aggregate inter-contact times distribution, showing when it is sufficient to consider the aggregate distribution to characterise the convergence properties of forwarding in SPNs.

As far as the first part of the analysis, note that [25] studied the relationship between individual pairs inter-contact times and aggregate inter-contact times for face-to-face contacts in mobile opportunistic networks. Authors considered a network where individual pairs inter-contact times are exponential (thus not falling in the divergence condition found in [10]), and show that the distribution of inter-contact rates can be responsible for a heavy tail in the aggregate inter-contact times distribution. In particular, the considered inter-contact rates following, respectively, gamma, exponential and Pareto distributions. Therefore, in the following we consider the same representative distributions for contact rates. First of all, we analyse the dataset presented in [29], which has been one of the basis for the results summarised in Section 3. The dataset collects information about 251 ego networks. Each relationship in each network provides a sample of contact rate, for a total of over 20000 contact rates samples. We fit the resulting empirical distribution to the reference distributions of this paper using the Maximum Likelihood (ML) method, and compare the fitted distributions against the data using the Akaike Information Criterion (AIC, [2]). As we find that a gamma distribution provides the best fit, we carry on a detailed analysis of this case (Section 5.1). For completeness, the study with the other reference contact rates distributions is presented in Section 5.2.

As far as the second part of the analysis, we assume that there exists a single pair of nodes whose inter-contact times distribution presents a heavy tail. We prove that, no matter what the rest of the individual pairs distributions are, the aggregate inter-contact times distribution presents the same heavy tail. Seen from a complementary standpoint, this result allows us to conclude that if the aggregate inter-contact time distribution is *not* heavy tailed, then no individual pair can have a heavy tail inter-contact times distribution.

5.1 The role of inter-contact rates: study of a measured case

Figure 3 shows a visual comparison of the samples obtained from [29] and the ML fittings of the considered contact rates distributions (ML estimators of the parameters are provided in Table 1). For the gamma distribution we consider the following definition (for the density)

$$f(\lambda) = \frac{\lambda^{\alpha-1} b^\alpha e^{-b\lambda}}{\Gamma(\alpha)} \tag{6}$$

where α and b are the shape and rate parameters, respectively. For the exponential distribution we consider the standard definition (resulting in the density in Equation 7)

$$f(\lambda) = be^{-b\lambda} \tag{7}$$

where b is the rate parameter. For the Pareto distribution, we consider the two possible definitions resulting in the CCDFs below:

$$\begin{array}{l} F(\lambda) = \left(\frac{b}{\lambda}\right)^{\alpha}, \alpha > 0,\ \lambda > b \\ F(\lambda) = \left(\frac{b}{b+\lambda}\right)^{\alpha}, \alpha > 0,\ \lambda > 0 \end{array} \tag{8}$$

where α and b are the shape and scale parameters. The difference between the two forms is that in the first case λ cannot take values arbitrarily close to 0, while in the second it can. We will show that this has a profound impact on the distribution of the aggregate inter-contact times. Hereafter, we denote with "Pareto" the first form, and with "Pareto0" the second form.

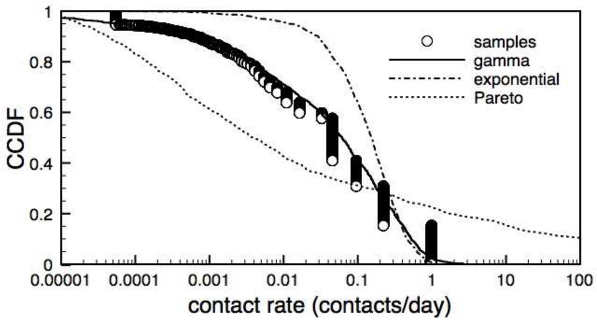

Figure 3: Fitting distributions

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 [2].

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
Pareto0	$\alpha = 0.16$, $b = 5.5\text{x}10^{-5}$	-28841.84

Table 1: AIC values for the tested distributions.

Based on this result, we study in detail the properties of the aggregate inter-contact times distribution assuming that the contact rates distribution is gamma, and the individual inter-contact times distributions are exponential. Note that the case of exponential inter-contact times of individual pairs is relevant. Most of the analytical studies in the opportunistic networking literature have been derived under this assumption. Moreover, exponential inter-contact times of individual pairs have been found in face-to-face contacts traces, e.g. [11, 15].

Lemma 2 and Theorem 2 characterises the distribution of the aggregate inter-contact times in this case.

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:*

$$\begin{cases} \mathcal{F}_l(x) \leq \frac{Re^{-\lambda_l(b+x)}}{x} & l = 1, \ldots, L-1 \\ \mathcal{F}_L(x) \simeq \frac{K}{x^{\alpha+1}} & \end{cases} \tag{9}$$

where R and K do not depend on x.

Proof. See Appendix B.

Theorem 2. *In a pervasive social 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:*

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

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.

Theorem 2 and Lemma 2 provide two interesting insights. First, the presence of aggregate inter-contact times with a 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 actually 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. This behaviour is confirmed also in the cases with Pareto contact rates.

To validate our analysis, 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.

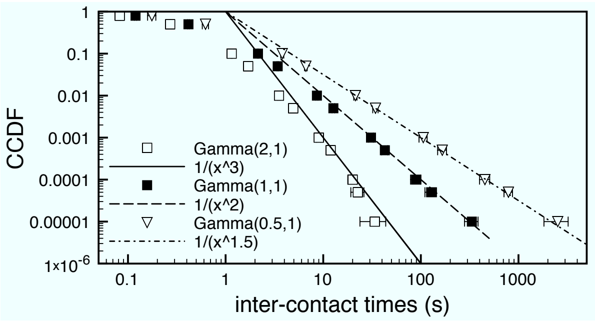

Figure 4: Aggregate inter-contact times with gamma contact rates

5.2 The role of inter-contact rates: other relevant cases

In this section we focus on the rest of the contact rates distributions considered in [25]. Specifically, as the case of an exponential distribution is a special case of a gamma distribution, we limit our analysis to Pareto distributions, in both variants, "Pareto" and "Pareto0", i.e. when contact rates arbitrarily close to 0 are not and are allowed, respectively.

The case "Pareto" is analysed in Lemma 3 and Theorem 3. With respect to the sectors corresponding to the layers of the ego networks, recall that in this case $\lambda_0 = \infty$ and $\lambda_L = b$ where b is the minimum possible value of the "Pareto" distribution.

LEMMA 3. *When contact rates follow a Pareto distribution whose CCDF is in the form $F(\lambda) = \left(\frac{b}{\lambda}\right)^\alpha$, $\lambda > b$ 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. Specifically, the following relations hold true for large x:*

$$\begin{cases} \mathcal{F}_1(x) \simeq \frac{Re^{-\lambda_1 x}}{x} \\ \mathcal{F}_l(x) \leq \frac{Ke^{-\lambda_l x}}{x} \quad l = 2, \ldots, L \end{cases} \tag{10}$$

where R and K do not depend on x.

Proof. See Appendix B.

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

$$F(\lambda) = \left(\tfrac{b}{\lambda}\right)^\alpha, F_\lambda(x) = e^{-\lambda x}$$
$$\Rightarrow \mathcal{F}(x) \leq \frac{Ke^{-bx}}{x} \text{ for large } x$$

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 related to the outer-most layer, and using the corresponding expression from Equation 10.

Lemma 3 and Theorem 3 show that even considering contact rates with an heavy tail (such as a "Pareto") may not be sufficient to obtain a heavy tail in the aggregate inter-contact times distribution. This is due to the fact that the "Pareto" distribution does not admit contact rates arbitrarily close to 0.

Finally, Lemma 4 and Theorem 4 analyse the case "Pareto0", i.e. the case where the contact rates can be arbitrarily close to 0.

LEMMA 4. *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:*

$$\begin{cases} \mathcal{F}_l(x) \leq \frac{Re^{-\lambda_l x}}{x} + \frac{Qe^{-\lambda_{l-1} x}}{x} \quad l = 1, \ldots, L-1 \\ \mathcal{F}_L(x) \simeq \frac{K}{x^2} \end{cases} \tag{11}$$

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

Proof. See Appendix B.

THEOREM 4. *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*

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

where K does not depend on x.

Proof. This comes immediately from Lemma 4 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 11.

As anticipated, in the case "Pareto0" the aggregate inter-contact times distribution presents a heavy tail. Intuitively, this is a side effect of the fact that contact rates can be arbitrarily close to 0. Again, note that this is another example (qualitatively similar to those of Theorem 2) where a heavy tail in the aggregate inter-contact times distribution is not necessarily a symptom of possible divergence of routing protocols, as it can emerge from exponentially distributed individual inter-contact times.

As a final validation check of the analytical results presented in this section, Figure 5 shows the CCDFs of aggregate inter-contact times in the "Pareto" and "Pareto0" cases, comparing analytical and simulation results (simulations where run as explained in Section 5.2). Also in this case the agreement between the analytical and simulation results is very good (remember that the analytical model describes the behaviour of the tail of the aggregate inter-contact times distribution).

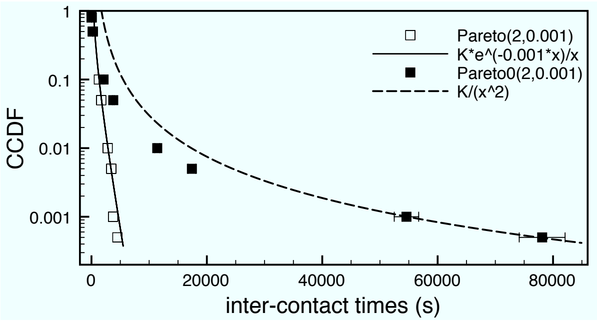

Figure 5: Aggregate inter-contact times with Pareto and Pareto0 contact rates.

5.3 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, and equal to P. Then, the following lemma holds true.

LEMMA 5. *In a network with a finite number of pairs, where there exist 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.*

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

PROOF. See Appendix B. ☐

Figure 6 provides a concrete example of the result in Lemma 5. 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 Section 5.2, this can be obtained, for example, by using exponentially distributed inter-contact times, and sampling the contact rates from a Pareto distribution. 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 (run as explained in Section 5.1) are marked with white squares, and the corresponding analytical curve (derived in Section 5.2) 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 5 is plotted with a dashed curve. According to Lemma 5, 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 exponent equal to $\eta = 1.1$. Figure 5 confirms both results obtained in Lemma 5.

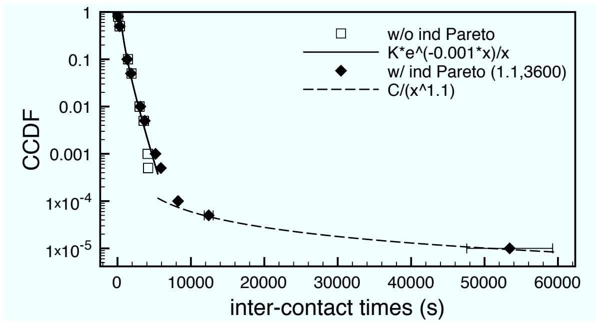

Figure 6: Aggregate inter-contact times with and without a single Pareto ICT pair.

Lemma 5 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 5. *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 5. 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 5 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 divergence of forwarding protocols found in [10], because no individual inter-contact times distribution can follow a power law. This result is dual to those found in Sections 5.1 and 5.2, 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 routing algorithms in social pervasive networks. Pervasive social networks are a possible evolution of current pervasive networks, where social network services are designed on top of a communication network that maps directly *human* social networks, i.e., a network where edges are activated between users that share social relationships, when they communicate because of their social tie. In social pervasive networks, a possible design paradigm for routing algorithms could be to exploit *contact events* between users, i.e. events of communication between users happening because of their social tie. In such a scenario, it is fundamental to characterise the properties of inter-contact times (i.e., the time between two consecutive contact events), as this has been shown to play a key role in determining convergence properties of routing algorithms in related (although different) networking environments.

In this paper we provided a mathematical model that formally characterises the dependence between the distribution of inter-contact times of individual pairs (which actually determine the convergence of routing protocols) and the distribution of aggregate inter-contact times (which is typically assumed to be the key distribution to analyse). This model can be used, as we shown in the paper, to analyse concrete cases and understand which distributions should be considered to assess the convergence properties of routing protocols for SPNs. In general, the results we derive by exploiting our model suggest a more cautionary approach to the analysis of inter-contact times with respect to the common approach largely used in the literature, which considers sufficient to analyse only the distribution of aggregate inter-contact times. Moreover, we highlight that the *heterogeneity* of the network is a fundamental aspect to take into consideration, as, together with the individual pairs distributions, it determines the distribution of the aggregate inter-contact times.

Acknowledgments

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.

7. REFERENCES

[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* NBS Applied Mathematics Series 55. National Bureau of Standards, 1964.

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

[3] S. Allen, M. Chorley, G. Colombo, and R. Whitaker. Self adaptation of cooperation in multi-agent content sharing systems. In *Self-Adaptive and Self-Organizing Systems (SASO), 2010 4th IEEE International Conference on*, pages 104 –113, 27 2010-oct. 1 2010.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

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

[10] 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.

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

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

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

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

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

[16] 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.

[17] T. Henderson, D. Kotz, and I. Abyzov. The changing usage of a mature campus-wide wireless network. In *Proceedings of the 10th annual international conference on Mobile computing and networking*, MobiCom '04, pages 187–201, New York, NY, USA, 2004. ACM.

[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 Transactions on Mobile Computing*, ((to appear)), 2011.

[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. *Communications Magazine, IEEE*, 48(9):140 –150, sept. 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. *SIGMOBILE Mob. Comput. Commun. Rev.*, 9:40–55, April 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 *Proceedings of the 6th ACM conference on Embedded network sensor systems*, SenSys '08, pages 337–350, 2008.

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

[26] 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.

[27] 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.

[28] 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.
[29] S. Roberts and R. Dunbar. Communication in social networks: Effects of kinship, network size, and emotional closeness. *Personal Relationships*, 2010.
[30] J. Su, A. Chin, A. Popivanova, A. Goel, and E. de Lara. User mobility for opportunistic ad-hoc networking. In *Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications*, pages 41–50, Washington, DC, USA, 2004. IEEE Computer Society.
[31] 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.

APPENDIX

A. PROOF OF THEOREM 1

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

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

where p_l is the probability that a social relationship of any given user is in layer l of its human social 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.

Proof. In the proof we focus on a given ego-network only. As users are supposed to be statistically equivalent as far as their social network is concerned, the distributions of inter-contact times aggregated over a given user or over all the users are the same. With respect to the expression of $\mathcal{F}(x)$ in Lemma 1, contact rates are not known, but are drawn from a set of r.v. with density $f_l(\lambda)$ (Equation 1). The expression of $\mathcal{F}(x)$ can be derived by conditioning to a specific set of rates $\lambda_1, \ldots, \lambda_N$, and applying the law of total probability. Without loss of generality, we can assume that the alters $\{1, \ldots, N\}$ are ordered according to their membership to layers, i.e., the first n_1 alters belong to the inner-most layer, etc. We thus obtain

$$\begin{aligned} \mathcal{F}(x) &= \int_{\lambda_1} \cdots \int_{\lambda_N} \mathcal{F}(x|\lambda_1,\ldots,\lambda_N) f(\lambda_1,\ldots,\lambda_N) d\lambda_1 \ldots d\lambda_N \\ &= \int_{\lambda_1} \cdots \int_{\lambda_N} \frac{\sum_{p=1}^{N} \lambda_p F_p(x)}{\sum_{p=1}^{N} \lambda_p} f_1(\lambda_1) \ldots f_N(\lambda_N) d\lambda_1 \ldots d\lambda_N \ , \end{aligned}$$

where we have assumed that rates of individual alters inter-contact times are independent. For a sufficiently large number of pairs in each layer, we can apply the law of large numbers, and approximate $\sum_{i=1}^{n_l} \lambda_i$ as $n_l E[\Lambda_l]$, and $\sum_{p=1}^{N} \lambda_p$ as $\sum_{l=1}^{L} n_l E[\Lambda_l]$. Swapping the integrals and the summations, and substituting $p_l = \frac{n_l}{N}$, we further obtain:

$$\begin{aligned} \mathcal{F}(x) &= \frac{1}{\sum_l n_l E[\Lambda_l]} \sum_p \int_{\lambda_1} \cdots \int_{\lambda_N} \lambda_p F_p(x) f_1(\lambda_1) \ldots f_N(\lambda_N) d\lambda_1 \ldots d\lambda_N = \\ &= \frac{1}{\sum_l n_l E[\Lambda_l]} \sum_{l=1}^{L} n_l \int_0^\infty \lambda F_\lambda(x) f_l(\lambda) d\lambda = \\ &= \sum_{l=1}^{L} \frac{p_l}{\sum_{l=1}^{L} p_l E[\Lambda_l]} \int_0^\infty \lambda f_l(\lambda) F_\lambda(x) d\lambda \\ &= \sum_{l=1}^{L} \frac{p_l C_l}{\sum_{l=1}^{L} p_l E[\Lambda_l]} \int_{\lambda_l}^{\lambda_{l-1}} \lambda f(\lambda) F_\lambda(x) d\lambda \ , \end{aligned}$$

where we have exploited the assumption that rates of individual pairs inter-contact times of the same layer are identically distributed, and that individual pairs inter-contact times of the same layer follow the same type of distribution, $F_\lambda(x)$.

Note that the above methodology can also be applied to show that $\mathcal{F}_l(x)$ in Equation 4 is the CCDF of the inter-contact times aggregated over layer l only. Specifically, exploiting again Lemma 1, we can condition $\mathcal{F}_l(x)$ to a known set of rates $\lambda_1, \ldots, \lambda_{n_l}$. Thus, we can write:

$$\begin{aligned} \mathcal{F}_l(x) &= \int_{\lambda_1} \cdots \int_{\lambda_{n_l}} \mathcal{F}(x|\lambda_1,\ldots,\lambda_{n_l}) f(\lambda_1,\ldots,\lambda_{n_l}) d\lambda_1 \ldots d\lambda_{n_l} \\ &= \int_{\lambda_1} \cdots \int_{\lambda_{n_l}} \frac{\sum_{i=1}^{n_l} \lambda_i F_i(x)}{\sum_{i=1}^{n_l} \lambda_i} f_1(\lambda_1) \ldots f_{n_l}(\lambda_{n_l}) d\lambda_1 \ldots d\lambda_{n_l} \ . \end{aligned}$$

By approximating $\sum_i \lambda_i$ as $n_l E[\Lambda_l]$, and by recalling that the contact rates in layer l are assumed to be identically distributed, we obtain

$$\begin{aligned} \mathcal{F}_l(x) &= \frac{1}{n_l E[\Lambda_l]} n_l \int_0^\infty \lambda f_l(\lambda) F_\lambda(x) d\lambda \\ &= \frac{C_l}{E[\Lambda_l]} \int_{\lambda_l}^{\lambda_{l-1}} \lambda f(\lambda) F_\lambda(x) d\lambda \end{aligned}$$

B. PROOF OF LEMMAS IN SECTION 5

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:*

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

where R and K do not depend on x.

Proof. First of all, it should be noted that when contact rates follow a gamma distribution, the values of λ that limits the sectors corresponding to the layers of the ego network are such that $\lambda_0 = \infty$ and $\lambda_L = 0$. Let us focus on the CCDF of aggregate inter-contact times on intermediate layers (i.e., excluding the inner- and the outer-most layers) first. From Equation 4, by substituting the expressions of

$f(\lambda)$ and $F_\lambda(x)$ we obtain:

$$\begin{aligned}\mathcal{F}_l(x) &= H\int_{\lambda_l}^{\lambda_{l-1}} \lambda^\alpha e^{-(b+x)\lambda} d\lambda \\ &= H\frac{\Gamma(\alpha+1, \lambda_l(b+x)) - \Gamma(\alpha+1, \lambda_{l-1}(b+x))}{(b+x)^{\alpha+1}}\end{aligned}$$

where H does not depend on x and $\Gamma(\cdot,\cdot)$ is the upper incomplete Gamma function. For large x, $\Gamma(s,x)$ can be approximated as $x^{s-1}e^{-x}$ [1]. We thus obtain:

$$\begin{aligned}\mathcal{F}_l(x) &\simeq \frac{x^\alpha(Re^{-\lambda_l(b+x)} - We^{-\lambda_{l-1}(b+x)})}{(b+x)^{\alpha+1}} \\ &\simeq \frac{Re^{-\lambda_l(b+x)} - We^{-\lambda_{l-1}(b+x)}}{x} \le \frac{Re^{-\lambda_l(b+x)}}{x}\end{aligned}$$

For the inner-most sector, we can write

$$\mathcal{F}_1(x) = H\int_{\lambda_1}^{\infty} \lambda^\alpha e^{-(b+x)\lambda} d\lambda = H\frac{\Gamma(\alpha+1, \lambda_1(b+x))}{(b+x)^{\alpha+1}}$$

For large x, applying the same approximation for $\Gamma(\cdot,\cdot)$, we obtain

$$\mathcal{F}_1(x) \simeq M\frac{x^\alpha e^{-\lambda_1(b+x)}}{(b+x)^{\alpha+1}} \simeq M\frac{e^{-\lambda_1(b+x)}}{x}$$

Finally, for the outer-most sector, $\mathcal{F}_L(x)$ becomes:

$$\begin{aligned}\mathcal{F}_L(x) &= H\int_{0}^{\lambda_{L-1}} \lambda^\alpha e^{-(b+x)\lambda} d\lambda \\ &= H\frac{\Gamma(\alpha+1) - \Gamma(\alpha+1, \lambda_{L-1}(b+x)}{(b+x)^{\alpha+1}}\end{aligned}$$

Approximating $\Gamma(\cdot,\cdot)$ we obtain

$$\begin{aligned}\mathcal{F}(x) &\simeq W\frac{\Gamma(\alpha+1)}{(b+x)^{\alpha+1}} - A\frac{x^\alpha e^{-\lambda_{L-1}(b+x)}}{(b+x)^{\alpha+1}} \\ &\simeq W\frac{\Gamma(\alpha+1)}{(b+x)^{\alpha+1}} \simeq \frac{K}{x^{\alpha+1}}\end{aligned}$$

Lemma 3. *When contact rates follow a Pareto distribution whose CCDF is in the form* $F(\lambda) = \left(\frac{b}{\lambda}\right)^\alpha$, $\lambda > b$ *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. Specifically, the following relations hold true for large* x:

$$\begin{cases}\mathcal{F}_1(x) \simeq \frac{Re^{-\lambda_1 x}}{x} \\ \mathcal{F}_l(x) \le \frac{Ke^{-\lambda_l x}}{x} \quad l = 2,\dots,L\end{cases}$$

where R *and* K *do not depend on* x.

Proof. Using the same methodology of Lemma 2, we obtain, for all components of $\mathcal{F}(x)$ but the one corresponding to the inner-most layer, the following expression:

$$\begin{aligned}\mathcal{F}_l(x) &= H\int_{\lambda_l}^{\lambda_{l-1}} \frac{e^{-\lambda x}}{\lambda^\alpha} d\lambda \\ &= H\frac{\Gamma(1-\alpha, \lambda_l x) - \Gamma(1-\alpha, \lambda_{l-1}x)}{x^{1-\alpha}}\end{aligned}$$

Applying the usual approximation of $\Gamma(\cdot,\cdot)$ for large x we obtain

$$\mathcal{F}_l(x) \simeq x^{-\alpha}\frac{Ke^{-\lambda_l x} - Qe^{-\lambda_{l-1}x}}{x^{1-\alpha}} \le \frac{Ke^{-\lambda_l x}}{x}$$

The component corresponding to the innermost layer can be written as

$$\mathcal{F}_1(x) = H\int_{\lambda_1}^{\infty} \frac{e^{-\lambda x}}{\lambda^\alpha} d\lambda = H\frac{\Gamma(1-\alpha, \lambda_1 x)}{x^{1-\alpha}}$$

The expression in the Lemma follows immediately by applying the usual approximation of $\Gamma(\cdot,\cdot)$ for large x.

Lemma 4. *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 outermost layer, which decays as a power law. Specifically, the following relations hold true for large* x:

$$\begin{cases}\mathcal{F}_l(x) \le \frac{Re^{-\lambda_l x}}{x} + \frac{Qe^{-\lambda_{l-1}x}}{x} \quad l = 1,\dots,L-1 \\ \mathcal{F}_L(x) \simeq \frac{K}{x^2}\end{cases}$$

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

Proof. Let us consider components of $\mathcal{F}(x)$ other than the one corresponding to the outermost layer. The following equation holds true:

$$\begin{aligned}\mathcal{F}_l(x) &= H\int_{\lambda_l}^{\lambda_{l-1}} \frac{\lambda}{(b+\lambda)^{\alpha+1}} e^{-\lambda x} d\lambda \\ &= e^{bx}\left\{\frac{\Gamma(1-\alpha, (\lambda_l+b)x) - \Gamma(1-\alpha, (\lambda_{l-1}+b)x)}{x^{1-\alpha}} + \right. \\ &\quad \left. + \; bx\frac{\Gamma(-\alpha, (\lambda_{l-1}+b)x) - \Gamma(-\alpha, (\lambda_l+b)x)}{x^{1-\alpha}}\right\}\end{aligned}$$

Applying the usual approximation of $\Gamma(\cdot,\cdot)$ for large x it is easy to obtain the following relation

$$\begin{aligned}\mathcal{F}_l(x) &\le e^{bx}\left\{\frac{[(\lambda_l+b)x]^{-\alpha}e^{-(\lambda_l+b)x}}{x^{1-\alpha}} + \right. \\ &\quad \left. + \; \frac{bx[(\lambda_{l-1}+b)x]^{-\alpha-1}e^{-(\lambda_{l-1}+b)x}}{x^{1-\alpha}}\right\}\end{aligned}$$

from which it is straightforward to derive the expression in Equation 11. As for the component of $\mathcal{F}(x)$ corresponding to the outermost layer, we obtain

$$\begin{aligned}\mathcal{F}_L(x) &= H\int_{\lambda_0}^{\lambda_{L-1}} \frac{\lambda}{(b+\lambda)^{\alpha+1}} e^{-\lambda x} d\lambda \\ &= e^{bx}\left\{\frac{\Gamma(1-\alpha, bx) - \Gamma(1-\alpha, (\lambda_{L-1}+b)x)}{x^{1-\alpha}} + \right. \\ &\quad \left. + \; bx\frac{\Gamma(-\alpha, (\lambda_{L-1}+b)x) - \Gamma(-\alpha, bx)}{x^{1-\alpha}}\right\}\end{aligned}$$

This time it is necessary to apply an approximation of $\Gamma(s,x)$ that considers higher order terms (for large x). This is derived in [1], as follows:

$$\Gamma(s,x) \simeq x^{s-1}e^{-x}\left(1 + \frac{s-1}{x}\right)$$

We thus obtain the final result shown in Equation 11:

$$\begin{aligned}\mathcal{F}_L(x) &\simeq e^{bx}\frac{(bx)^{-\alpha-1}e^{-bx}(\alpha+1) - (bx)^{-\alpha-1}e^{-bx}\alpha}{x^{1-\alpha}} \\ &= \frac{K}{x^2}\end{aligned}$$

LEMMA 5. *In a network with a finite number of pairs, where there exist 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.*

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

PROOF. With respect to the ego network model described in Section 3, it is sufficient to focus on the layer to which pair p belongs, hereafter denoted to as $\hat{l}$. Let us assume that the number of alters in that social layer is $P+1$. As we did for Theorem 1, we can exploit the result of Lemma 1 assuming that all rates of individual inter-contact times are known. Let us denote with λ_p the inter-contact time for pair p, and with $\lambda_1, \ldots, \lambda_P$ the rest of the inter-contact rates. We can write the expression in Lemma 1 as follows:

$$\begin{aligned}&\mathcal{F}_{\hat{l}}(x|\lambda_1, \ldots, \lambda_P, \lambda_p)\\ &= \textstyle\sum_{i=1}^{P} \frac{\lambda_i}{\sum_i \lambda_i + \lambda_p} F_i(x) + \frac{\lambda_p}{\sum_i \lambda_i + \lambda_p} F_p(x)\end{aligned} \quad (12)$$

Recalling that one of the rates is known a priori (the one corresponding to pair p), we can remove the conditioning on the rest of the rates, and write the distribution of aggregate inter-contact times as follows:

$$\begin{aligned}\mathcal{F}_{\hat{l}}(x|\lambda_p) &= \int_{\lambda_1} \cdots \int_{\lambda_P} \mathcal{F}(x|\lambda_1,\ldots,\lambda_P,\lambda_p) f(\lambda_1,\ldots,\lambda_P) d\lambda_1 \ldots d\lambda_P \\ &= \int_{\lambda_1} \cdots \int_{\lambda_P} \left[\frac{\sum_{i=1}^{P} \lambda_i F_i(x)}{\sum_{i=1}^{P} \lambda_i + \lambda_p} + \frac{\lambda_p F_p(x)}{\sum_{i=1}^{P} \lambda_i + \lambda_p}\right] \\ &\quad f_1(\lambda_1) \ldots f_P(\lambda_P) d\lambda_1 \ldots d\lambda_P \ .\end{aligned}$$

Denote with $E[\Lambda_{\hat{l}}]$ the average rate in layer $\hat{l}$ *excluding* pair p. By approximating, $\sum_i \lambda_i$ as $PE[\Lambda_{\hat{l}}]$ the second term of the addition in the above expression does not depend on $\lambda_1, \ldots, \lambda_P$ anymore. This terms thus result in a component of $\mathcal{F}_{\hat{l}}(x)$ equal to $\frac{\lambda_p F_p(x)}{PE[\Lambda_{\hat{l}}] + \lambda_p}$. As far as the first term of the addition, recall that we assume that inter-contact rates in layer $\hat{l}$ (but the rate of pair p) are iid. Applying the usual approximation for $\sum_i \lambda_i$, and swapping the summation with the integrals we can write

$$\begin{aligned}&\int_{\lambda_1} \cdots \int_{\lambda_P} \frac{\sum_{i=1}^{P} \lambda_i F_i(x)}{\sum_{i=1}^{P} \lambda_i + \lambda_p} f_1(\lambda_1) \ldots f_P(\lambda_P) d\lambda_1 \ldots d\lambda_P = \\ &\frac{1}{PE[\Lambda_{\hat{l}}] + \lambda_p} \sum_i \int_{\lambda_1} \cdots \int_{\lambda_P} \lambda_i F_i(x) f_1(\lambda_1) \ldots f_P(\lambda_P) d\lambda_1 \ldots d\lambda_P = \\ &\frac{P}{PE[\Lambda_{\hat{l}}] + \lambda_p} \int_0^\infty \lambda f_{\hat{l}}(\lambda) F_\lambda(x) d\lambda\end{aligned}$$

By recalling the expression of the aggregate distribution over an individual layer derived in the proof of Theorem 1, and by denoting with $\hat{\mathcal{F}}(x)$ the distribution of the aggregate inter-contact times over layer $\hat{l}$ excluding pair p, it is straightforward to see that the above term can be written as follows:

$$\frac{P}{PE[\Lambda_{\hat{l}}] + \lambda_p} \int_0^\infty \lambda f_{\hat{l}}(\lambda) F_\lambda(x) d\lambda = \frac{PE[\Lambda_{\hat{l}}]}{PE[\Lambda_{\hat{l}}] + \lambda_p} \hat{\mathcal{F}}(x)$$

Putting everything together, and by defining α as $\frac{PE[\Lambda_{\hat{l}}]}{PE[\Lambda_{\hat{l}}] + \lambda_p}$, we can write the CCDF of the aggregate inter-contact times as follows:

$$\mathcal{F}_{\hat{l}}(x|\lambda_p) = \alpha \hat{\mathcal{F}}(x) + (1-\alpha) F_p(x) \quad (13)$$

To complete the proof it is sufficient to distinguish between two possible cases, i.e., whether $\hat{\mathcal{F}}(x)$ presents or not a heavy tail. In the latter case, by definition, for any $\beta > 0$ it holds true that $\lim_{x\to\infty} \frac{\hat{\mathcal{F}}(x)}{x^{-\beta}} = 0$, i.e. $\hat{\mathcal{F}}(x)$ decays, for large x, faster than any power law. Therefore, recalling that $F_p(x)$ can be approximated with $x^{-\eta}$ for large x, we obtain:

$$\lim_{x\to\infty} \frac{\mathcal{F}_{\hat{l}}(x|\lambda_p)}{x^{-\eta}} = \lim_{x\to\infty} \alpha \frac{\hat{\mathcal{F}}(x)}{x^{-\eta}} + (1-\alpha) \frac{F_p(x)}{x^{-\eta}} = 1 - \alpha$$

In this case, therefore, $\mathcal{F}_{\hat{l}}(x|\lambda_p)$ can be approximated with $(1-\alpha)x^{-\eta}$, which means that the distribution of the aggregate inter-contact times presents a heavy tail with exactly the same shape of pair p.

In the case where $\hat{\mathcal{F}}(x)$ presents a heavy tail, we can approximate it, for large x, as $x^{-\beta}$ for some $\beta > 0$. If $\beta > \eta$ then $\hat{\mathcal{F}}(x)$ decays faster than $F_p(x)$, and we can thus again approximate $\mathcal{F}_{\hat{l}}(x|\lambda_p)$ with $(1-\alpha)x^{-\eta}$. If $\beta < \eta$ then $F_p(x)$ decays faster than $\hat{\mathcal{F}}(x)$, which means that we can approximate $\mathcal{F}_{\hat{l}}(x|\lambda_p)$ with $\alpha x^{-\beta}$. As $\beta < \eta$, $\mathcal{F}_{\hat{l}}(x|\lambda_p) \geq \alpha x^{-\eta}$ holds true for large x.

This concludes the proof. $\square$