Rhythm and Randomness in Human Contact

Mervyn P. Freeman, Nicholas W. Watkins
British Antarctic Survey
Cambridge, United Kingdom
Email: {mpf, nww}@bas.ac.uk

Eiko Yoneki, Jon Crowcroft
University of Cambridge
Cambridge, United Kingdom
Email: {eiko.yoneki, jon.crowcroft}@cl.cam.ac.uk

Abstract—**There is substantial interest in the effect of human mobility patterns on opportunistic communications. Inspired by recent work revisiting some of the early evidence for a Lévy flight foraging strategy in animals, we analyse datasets on human contact from real world traces. By analysing the distribution of inter-contact times on different time scales and using different graphical forms, we find not only the highly skewed distributions of waiting times highlighted in previous studies but also clear circadian rhythm. The relative visibility of these two components depends strongly on which graphical form is adopted and the range of time scales. We use a simple model to reconstruct the observed behaviour and discuss the implications of this for forwarding efficiency.**

I. Introduction

Digital traffic flows not only over the wired backbone of the Internet or network of mobile phone masts, but also in small leaps through physical space as people pass one another on the street [14]. Thus opportunities for a new communication paradigm via wireless-enabled devices are emerging, which communicate directly with other devices within their range and without a costly and inflexible planned infrastructure (e.g., [9]). To improve communication efficiency and prevent the spread of wireless viruses in this new generation of communication requires new insights and quantitative models of human interaction. Of fundamental importance in this case is the time sequence of human contacts, as well as other properties of complex networks, such as small-worldness, etc. (e.g., the special issue of Science on Complex Systems and Networks, July 24, 2009).

Recently, the emergence of human interaction traces from online and pervasive environments is allowing us to understand details of human activities. For example, the MIT Reality Mining project [6] collected proximity, location and activity information, with nearby nodes being discovered through periodic Bluetooth scans and location information from cell tower IDs. Several other groups have performed similar studies. Some have used Bluetooth to measure device connectivity [6], [9], [18], while others rely on WiFi [11], GPS [22], [23], [15], or the position of cell towers [10]. The duration of experiments has varied from 2 days to over one year, and the numbers of participants has also varied from ~ 10 to $\sim 100,000$.

It has been suggested that the probability density function (pdf) $p(t)$ of times between human contact is well approximated by a truncated power law i.e. $p(t) \sim t^{-(1+\alpha)}$ over some range. This is so whether the contact is by physical proximity (i.e., detectability of wireless access points or Bluetooth devices, or closeness of GPS locations [3], [13], [22]) or by telecommunication (i.e., mobile phone call [10] or e-mail [16]), and whether one or both contacting devices are in motion (e.g., both Bluetooth, one Bluetooth and fixed wireless access points, mobile phone and fixed masts).

A summary is given in Table I of studies in which the stability exponent α has been inferred from an inter-contact time (ICT) distribution, together with the approximate range of applicability. From the quoted values, α is inferred to be in the interval $[\approx 0, 0.9]$ which is within the allowable range ($0 < \alpha \leq 2$) for the tails of a Lévy (stable) distribution [20], [17] (except possibly for the marginal case of the Europe study of mobile phone contact which could actually be a gamma distribution.) Consequently it has been argued that human mobility patterns resemble truncated Lévy walks (TLW). The TLW paradigm represents a development of the Lévy flight, which was a random walk comprising steps drawn from a Lévy distribution, rather than a Gaussian as occurs in the more familiar Brownian random walks [24]. The first modification, to a finite constant velocity, was dubbed a Lévy walk. Subsequently the limitation to a finite domain was described as truncation [17]. More recently some researchers have also considered the velocity to be a variable (e.g. [23]).

Similar movement patterns have also been inferred for animals [25], and it has been proposed that Lévy

User population	Intel	Cambridge 1	INFOCOM 2005	Toronto	UCSD	Dartmouth	Europe
Source	Chaintreau et al. (2006) [3]						Gonzalez et al. (2009) [10]
Device	iMote	iMote	iMote	PDA	PDA	Laptop/PDA	Mobile phone
Network type	Bluetooth	Bluetooth	Bluetooth	Bluetooth	WiFi	WiFi	Mobile phone
Granularity	120 seconds	120 seconds	120 seconds	120 seconds	120 seconds	300 seconds	N/A
Duration	3 days	5 days	3 days	16 days	77 days	114 days	6 months
Devices participating	8	12	41	23	273	6648	100,000
Number of internal contacts	1,091	4,229	22,459	2,802	195,364	4,058,284	16,364,308
Approximate extent of power law region	4min - 14min	10min - 30min	10min - 10h	2min - 6min	20min - 1day	10min - 1h	100s - 8h
Quoted power law exponent	-0.9	-0.9	-0.4	-0.9	-0.3	-0.3	-0.9 +/- 0.1
Type of distribution plotted	Tail df (ccdf)	Tail df (ccdf)	Tail df (ccdf)	Tail df (ccdf)	Tail df (ccdf)	Tail df (ccdf)	Log-binned pdf
Inferred stability exponent α	**0.9**	**0.9**	**0.4**	**0.9**	**0.3**	**0.3**	**-0.1 +/- 0.1**

TABLE I: Summary of studies in which the stability exponent α has been inferred from an inter-contact time distribution

foraging is an optimal strategy under at least some circumstances [26]. Debate continues as to the extent to which a Lévy strategy could be universal and insensitive to the details of the environment and of the physiology and motivation of the individual (e.g. [8], [21], and references therein). However, the statistical analysis methods which have most frequently been used to infer empirical support for the truncated Lévy walk hypothesis have recently been criticised, both in the ecology literature and more generally [8], [7], [4], [27]. Key problems identified have included: (1) The widespread inference of power law pdfs by the graphical method of straight line fitting to histograms with double logarithmic axes; (2) the difficulty of inferring power laws over very limited ranges; (3) the use of intrinsically biased methods (such as (1)) for estimating the power law exponent; and, perhaps most importantly, (4) inadequate, or even a complete lack of, alternative hypotheses.

In the light of this, it is worthwhile to consider how these problems might apply to the human mobility studies cited above and summarised in Table I. For example, some simply compare their distributions with a straight line on a log-log plot with unavoidable bias and spread for the inferred power law exponent [3], [13]. In addition, referring to Table I, the inference of a possible power law region is very weak for the Intel, Cambridge 1 and Toronto experiments because the region is so limited ($\sim 1/3$ decade), presumably related to the small samples ($\sim$ 1000 contacts). The evidence is more convincing for the larger samples (INFOCOM 2005, UCSD, Dartmouth and Europe) with wider apparent power law regions. Alternative hypotheses to the pure power law null model have been considered, such as the exponentially-truncated power law [13], [10], but only one study [15] has actually fitted and quantitatively compared several alternative models to ICT distributions (albeit simulated), using the less biased maximum likelihood estimate to infer the model parameters such as the power law exponent and Akaike weights [1] to compare the goodness of fits. Thus, at present, the inference of a truncated power law ICT distribution directly from experiment is limited.

Indeed, it would be surprising if a truncated power law was a complete description of human ICT considering our prior knowledge about the social habits and structures of humans, such as the working day and family and community responsibilities [16]. In fact it has been recognised that the ICT distribution is not stationary and changes with the time of day [13]. Spatial movement distributions also exhibit daily patterns [10] and Fourier analyses of proximity edges have daily and weekly periodicities [6]. Similarly a fundamental semi-diurnal periodicity was identified in an early study claiming a Lévy strategy for animal foraging [25]. This suggests that alternative models combining non-trivial randomness and periodic rhythms should be investigated. At present such more complicated models are challenging to test rigorously (e.g., by MLE) but progress can nevertheless be made by closer examination of the experimental ICT distribution using different graphical methods and modelling.

In this paper we consider three similar human contact experiments of varying durations (section 2). We analyse and model them to identify regularities that modify the underlying Lévy walk behaviour (section 3). Then we briefly compare these analyses with others in the literature and discuss how this hybrid behaviour may be modelled and will affect the efficiency of ad-hoc communication (section 4).

II. Datasets

We analyse trace data from the Haggle project [9] and Crawdad database [5], collected using Bluetooth communication in a conference environment and two university study environments. The configuration of data collection is summarised in Table II.

MIT: in the MIT Reality Mining project [6], 100 smart phones were deployed to students and staff at MIT over a period of 9 months. These phones were running software that logged contacts.

Cambridge 2: in the Cambridge Haggle project [9], 36 iMotes (Intel Mote ISN100-BA) were deployed to 1st year and 2nd year undergraduate students for 11 days to detect proximity using Bluetooth. The iMote runs TinyOS and is equipped with an ARM7TDMI processor operating at 12MHz, with 64kB of SRAM, 512kB of flash storage, and a multi-coloured LED, and a Bluetooth 1.1 radio, which has a radio range around 30 meters.

INFOCOM 2006: also in the Cambridge Haggle project, 77 iMotes were deployed at the INFOCOM 2006 conference for 3 days.

Experimental data set	MIT	Cambridge 2	INFOCOM 2006
Device	Phone	iMote	iMote
Network type	Bluetooth	Bluetooth	Bluetooth
Duration L (days)	246	11	3
Granularity Δ (seconds)	600	100	100
Number of Devices	97	36	77
Number of Contacts	54,667	10,873	191,336
Average # Contacts / Day	0.024	0.345	6.7

TABLE II: Characteristics of the experiments

The logged data from the above experimental studies are used to build time-dependent network information to study the distribution of contact times, inter-contact times, community structure and their statistical properties, where we constructed discrete event traces of pair interactions of 10 to 600 seconds intervals. We have aggregated raw data within 100 or 600 second time windows to avoid uncertainty of device detection from a complex Bluetooth communication protocol.

A complex operation is required to collect accurate connectivity traces using Bluetooth communication, as the device discovery protocol may limit detection of the devices in radio proximity. Bluetooth uses a special physical channel for devices to discover each other. A device becomes discoverable by entering the inquiry substate where it can respond to inquiries from other devices. The inquiry scan substate is used to discover surrounding devices. The discovering device iterates (hops) through all possible inquiry scan channel frequencies in a pseudo-random fashion. For each frequency, it broadcasts an inquiry and listens for responses. Therefore, a Bluetooth device cannot scan for other devices when the device cannot be in discoverable. Bluetooth inquiry can only happen in 1.28 second intervals. It is reported that an interval of $4 \times 1.28 = 5.12$ seconds gives a more than 90% chance of finding a device. However, there is no available data for situations where many devices are present, and no precise study has been reported. The Bluetooth standard recommends being in the inquiry scan substate for 10.24 seconds in order to collect all responses in an error-free environment. A 10.24 seconds alternation may cause missing links, and we therefore deploy 5.12 seconds for inquiry. The power consumption of Bluetooth is also a critical limitation for the scanning interval. The iMote connectivity traces in Haggle [9] use a scanning interval of approximately 2 minutes, while the Reality Mining project in MIT [6], with cell phones, uses 5

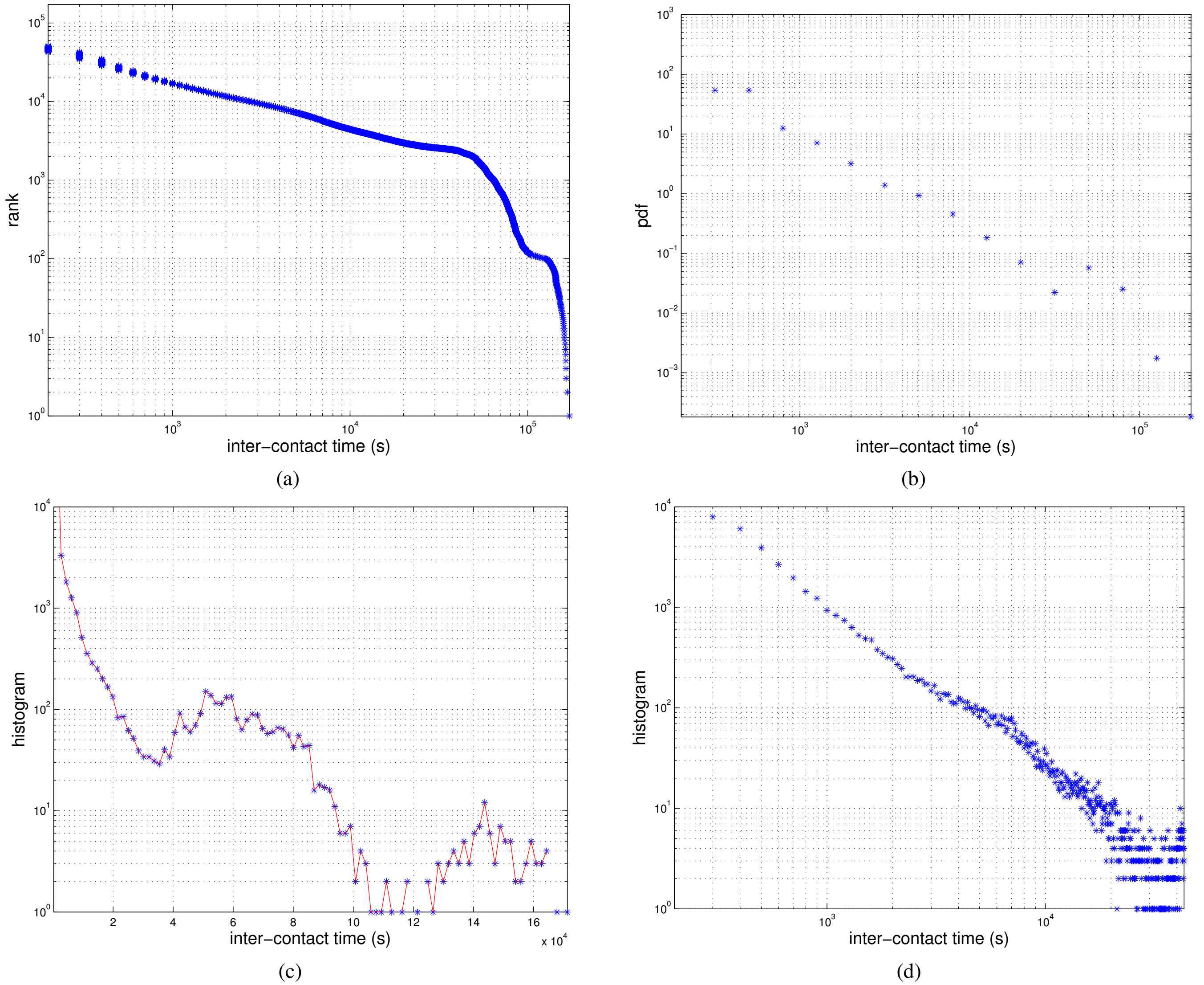


Fig. 1: INFOCOM 2006: (a) Rank order plot, (b) pdf, (c) semilog histogram with linear bins, and (d) loglog histogram for times less than 12 hours, with linear bins

minutes. The ratio of devices with Bluetooth enabled to the total number of devices is around only an average 15% - 20% of population. The range of Bluetooth varies between 10m and 80m, which depends on the device class such as cell phones or laptops. In cell phones, the Bluetooth range is usually 5 - 10m. We have observed that the devices can be detected in a 20m range if there are no obstacles, while with obstacles such as a thick wall the range drops to 5m (see more detail in [18][19]).

III. Rhythm and randomness

In each of the experiments we calculated all possible inter-contact times T between any two nodes, where ICT is defined as the time between the end of contact between two nodes and the start of next contact between the same two nodes. Figures 1-3 summarise the ICT distribution for the three experiments. In each case, the distribution is plotted as (a) a rank order plot with double logarithmic axes, (b) a probability density function with logarithmic co-ordinate (probability density) axis and logarithmic ordinate (inter-contact time) axis using exponentially spaced bins (i.e., equal bin width in logarithm space = 0.1 decade), (c) a histogram with logarithmic co-ordinate (frequency) axis and linear ordinate (inter-contact time) axis using 100 equally-spaced bins (equivalent to a pdf with linearly spaced bins to within a constant), and (d) a histogram for inter-contact times up to 12 hours with logarithmic co-ordinate (frequency) axis and logarithmic ordinate

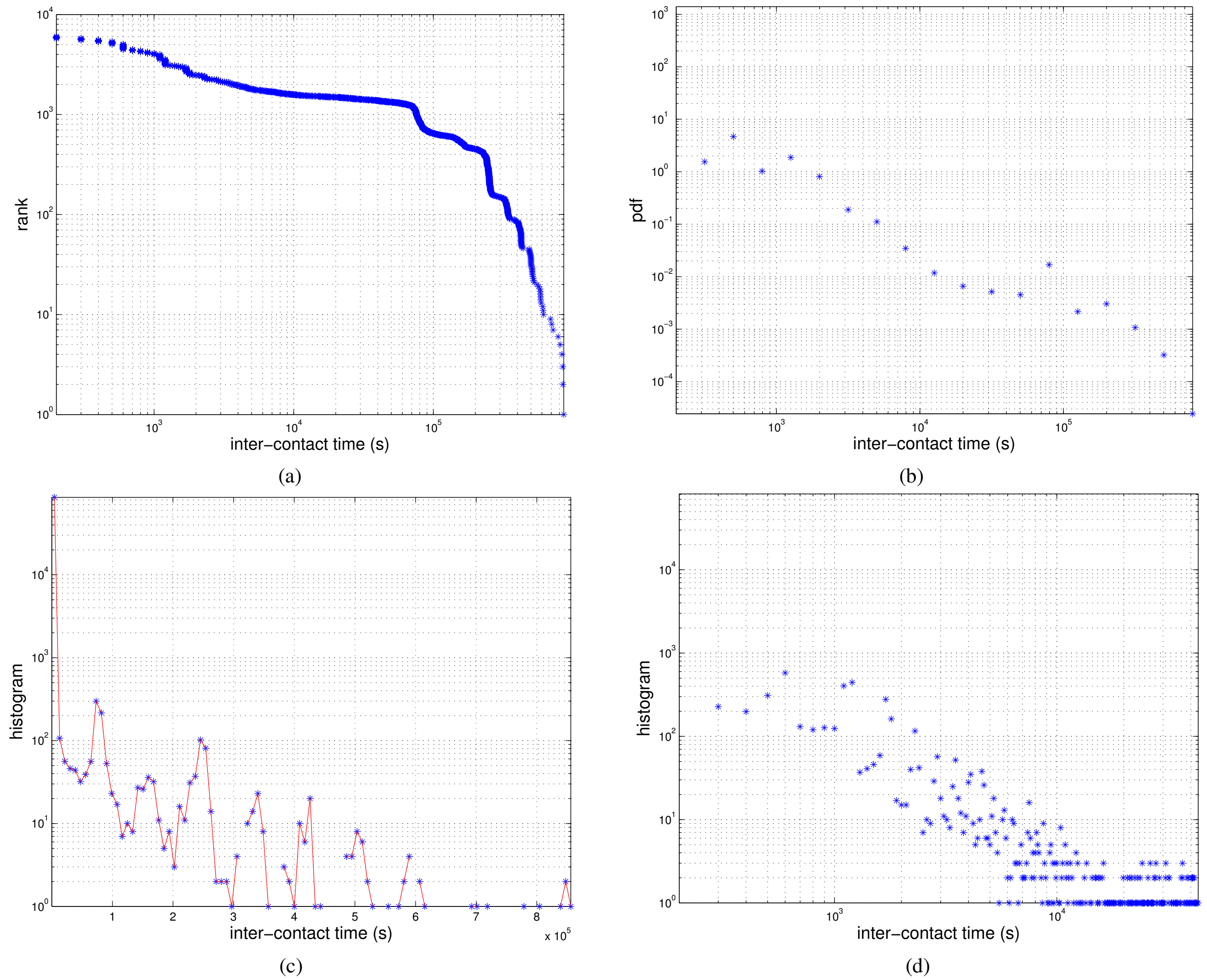


Fig. 2: Cambridge 2: (a) Rank order plot, (b) pdf, (c) semilog histogram with linear bins, and (d) loglog histogram for times less than 12 hours, with linear bins

(inter-contact time) axis using equal bin widths at the granularity $\Delta = 100$ s or 600 s. (The inset in Figure 3c shows a double logarithmic histogram using equal bin widths of 1800 s.)

A. **Truncated power law distribution**

Considering the rank order plots in Figure 1-3, we might suggest as others have done that the ICT tail distribution of all three experiments roughly resembles a restricted range power law with exponent < 1 (cf Figure 1 and 2 of [3]). To illustrate this, we performed the following simulation:

1) A set of contact times is calculated for Lévy walks in a domain bounded by the duration of the experiment L (see Table II). Specifically we calculate the cumulative sum, $t_i = \sum_{j=1}^{i} X_j$, where X is a set of N iid samples chosen from the Pareto distribution with pdf $p(x) \sim x^{-(1+a)}$ in the range Δ to $100L$. The samples are generated by picking iid samples C_i from the uniform distribution in the range (0,1] and then inverting the analytical equation for the Pareto cumulative probability distribution to find the value $x = X_i$ that yields the value $C = C_i$.
2) Divide the contact times into individual trials (i.e., trial number $= t_i$ modulo L).
3) Calculate the set of inter-contact times T from the time differences between neighbouring contact times, $T_i = t_{i+1} - t_i$, omitting inter-contact times

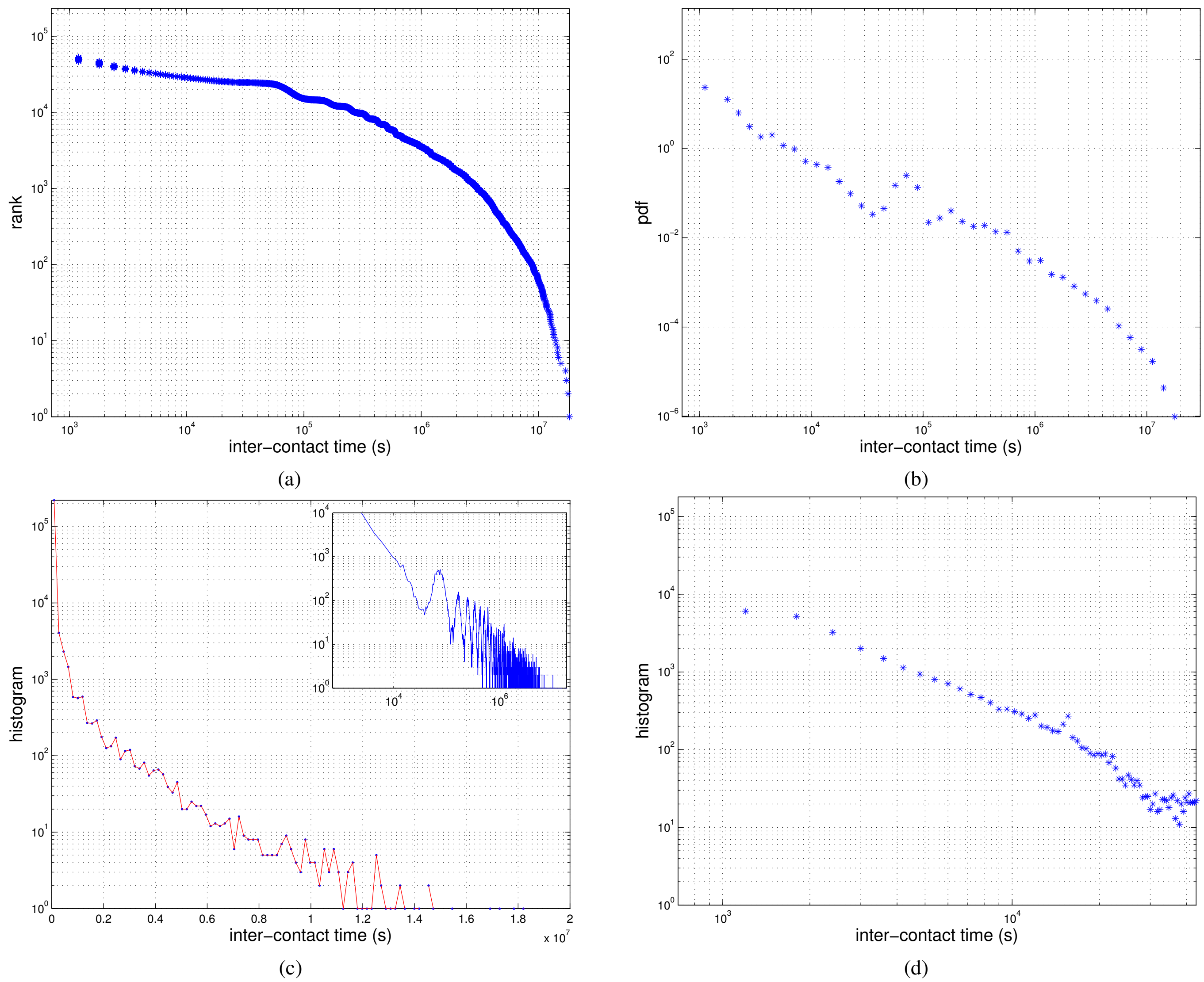


Fig. 3: MIT: (a) Rank order plot, (b) pdf, (c) semilog histogram with linear bins (inset shows loglog histogram with linear bins), and (d) loglog histogram for times less than 12 hours, with linear bins

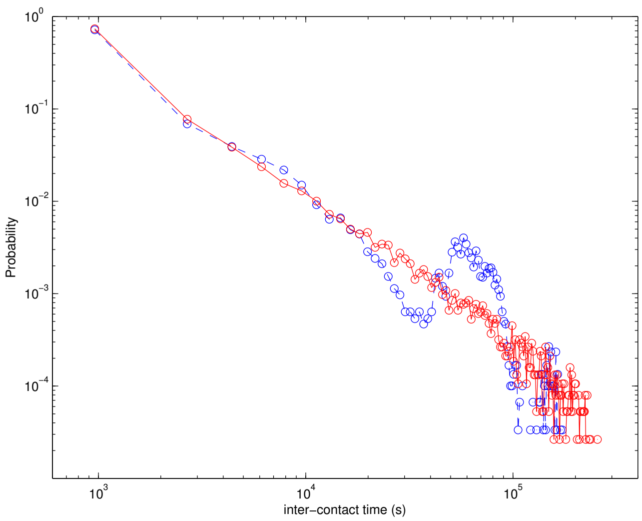
that straddle trials.

Figure 4a shows a simulated ICT probability distribution (solid line) choosing $a = 0.4$ and other parameters corresponding to the configuration of the INFOCOM 2006 experiment – $\Delta = 100$ s, $L = 3$ days, and $N = 10,000$. It is clear that the simulated distribution is only a crude approximation to the actual INFOCOM 2006 distribution (dashed line) and other structure is evident.

B. Circadian rhythm

This is also obvious in the other experiments (e.g., the histograms in figures 1c, 2c and 3c.) where there are significant deviations about any candidate monotonic function. Closer inspection reveals much of this deviation to be associated with a circadian rhythm, as evidenced by the alignment of peaks in the histogram/PDF at integer multiples of 24 hours. (Note also a weekly rhythm in Figure 3.)

Nevertheless, the INFOCOM 2006 and MIT distributions are well approximated by a power law on time scales much less than a day (e.g., $<$ 12 hours, see Figure 1d and 3d). (This is less obvious in the Cambridge 2 experiment (Figure 2d) due to a ≈ 10 min periodicity which is likely an experimental artefact). This suggests that a better null model of ICT in these experiments is a Lévy walk in a periodic domain. To investigate this we performed the following simulation:



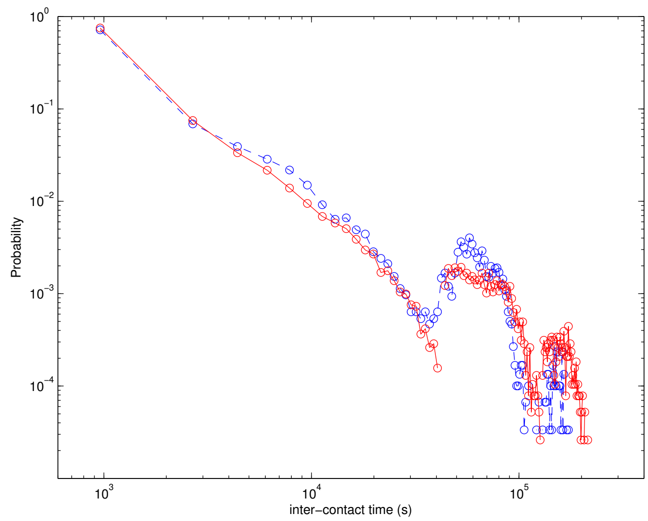


Fig. 4: Comparison of Lévy flight simulation of inter-contact times without (left), and with (right), the presence of circadian periodicity.

1) A set of contact times is calculated for Lévy walks in a domain bounded by the duration of the experiment L (see Table II). Specifically we calculate the cumulative sum, $t_i = \sum_{j=1}^{i} X_j$, where X is a set of N iid samples chosen from the Pareto distribution with pdf $p(x) \sim x^{-(1+a)}$ in the range Δ to L. The samples are generated by picking iid samples C_i from the uniform distribution in the range (0,1] and then inverting the analytical equation for the Pareto cumulative probability distribution to find the value $x = X_i$ that yields the value $C = C_i$.
2) Divide the contact times into days and retain only contact times that fall within a working day, defined to start at t_s h and end at t_e h (i.e., $t_s \leq d_i = t_i \text{ modulo } 24\text{hours} \leq t_e$).
3) Divide the contact times into individual trials (i.e., trial number $= t_i \text{ modulo } L$).
4) Calculate the set of inter-contact times T from the time differences between neighbouring contact times, $T_i = t_{i+1} - t_i$, omitting inter-contact times that straddle trials.

Figure 4b shows a simulated ICT probability distribution (solid line) choosing $a = 0.4$ and other parameters corresponding to the configuration of the INFOCOM 2006 experiment: $\Delta = 100$ s, $L = 3$ days, and $N = 10,000$. The simulated distribution compares favourably with the actual INFOCOM 2006 distribution (dashed line), supporting the null model over a Lévy walk confined within the domain L but not within the working day.

IV. Conclusions and Implications

The distribution of human inter-contact times from three experiments of differing durations has been analysed using different graphical presentations. This has revealed three essential properties of human contact:

Random, scale-free. On sufficiently short time scales, the ICT distribution is approximated by a power law consistent with the return times of a Lévy flight. The value of the stability exponent ($\alpha < 1$) implies no characteristic ICT in the absence of other constraints.

Truncated. At some time scale the power law component is truncated by a constraint on inter-contact time. One artificial constraint is the experiment itself which prohibits recording ICTs longer than the experiment duration. This is demonstrated in the simulated ICT distribution in Figure 4a and should be considered in comparing results from experiments of differing durations. More significantly, another constraint is the removal of agents from the contact domain. An example of this is movement from work to home which suppresses ICTs between agents in the same work group on times scales beyond the working day. This is demonstrated in the simulated ICT distribution in Figure 4b by the truncation of the power law component at $ICT \sim 10^4$ s.

Periodic. Environmental, biological, and social constraints may have rhythms that encourage repeated encounters such as the daily to-ing and fro-ing between work and home. This is demonstrated in the simulated ICT distribution in Figure 4b by the peak at $ICT \sim 6 \times 10^4$ s and $\sim 15 \times 10^4$ s (i.e., separated by 24 hours).

These three properties have been previously surmised by various different means but evidence of their co-existence in the ICT distribution has been overlooked. In particular, closer examination of previously published ICT distributions (e.g., [10]) reveals deviations about a truncated power law consistent with a circadian rhythm. Recognition of this rhythm in the empirical distribution is important otherwise models of human movement and behaviour may be unrealistically modified to generate only the scale-free property (e.g., [2]). It also has significant implications for building efficient routing algorithms and functionality on top of opportunistic networks. As a very simple example, clearly a rhythm of period P that removes agents from each other for a time $P/2$ reduces the average number of contacts by 50% over multiple cycles. But its determinism might also be exploited to increase communication efficiency. For example, the time of the next encounter could be estimated at the node and thus selection of the next hop could be determined based on the expected shortest time to the next encounter. The periodic behaviour of nodes could indicate moving from one network partition to another and this could be used for temporal clustering of nodes, where temporal-based communities could be used as a backbone of logical network structure for forwarding [12]. By these means, mobility-assisted forwarding can take advantage of patterns arising in the distribution of nodes in time and space. One alternative movement model is suggested that combines the Lévy walk model with models such as the Home Cell Mobility Model (e.g., [2]) that incorporate the influence of social structure. However the development of more complicated models will also present challenges in testing them and distinguishing between competing models.

Acknowledgement

The research is part funded by the EU grants for the Haggle project, IST-4-027918, the SOCIALNETS project, 217141, and the EPSRC DDEPI Project, EP/H003959. MPF and NWW are supported by the Natural Environment Research Council.

References

[1] D. R. Anderson and K. P. Burnham. Understanding information criteria for selection among capture-recapture or ring recovery models, S14-21. *Bird Study*, 46 (suppl.), 1999.

[2] C. Boldrini, M. Conti, and A. Passarella. The sociable traveller: human travelling patterns in social-based mobility. In *Proc. MOBIWAC09*, 2009.

[3] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on the design of opportunistic forwarding algorithms. In *Proc. INFOCOM*, April 2006.

[4] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51, 2009.

[5] Dartmouth College. A community resource for archiving wireless data at dartmouth, http://crawdad.cs.dartmouth.edu/index.php, 2007.

[6] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, V10(4):255–268, May 2006.

[7] A. M. Edwards. Using likelihood to test for Lévy flight search patterns and for general power-law distributions in nature. *Journal of Animal Ecology*, 77, 2008.

[8] A. M. Edwards, R. Phillips, N. W. Watkins, M. P. Freeman, and Others. Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, 449, 2007.

[9] EU FP6 Haggle Project. http://www.haggleproject.org, 2010.

[10] M. Gonzalez, C. Hidalgo, and A. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196), 2008.

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

[12] P. Hui, J. Crowcroft, and E. Yoneki. BUBBLE Rap: Social Based Forwarding in Delay Tolerant Networks. In *MobiHoc*, 2008.

[13] T. Karagiannis, J.-Y. L. Boudec, and M. Cojnovic. Power law and exponential decay of inter contact times between mobile devices. In *Proc. MOBICOM*, 2007.

[14] J. Kleinberg. The wireless epidemic. *Nature*, 449(20), 2007.

[15] K. Lee, S. Hong, and Others. Slaw: A mobility model for human walks. In *Proc. INFOCOM*, 2009.

[16] R. D. Malmgren, D. B. Stouffer, A. E. Motter, and L. A. N. Amaral. A Poissonian explanation for heavy tails in e-mail communication. *PNAS*, 105, 2008.

[17] R. N. Mantegna and H. E. Stanley. *An introduction to econophysics: Correlations and complexity in finance*. Cambridge University Press, 2000.

[18] T. Nicolai, E. Yoneki, N. Behrens, and H. Kenn. Exploring social context with the wireless rope. In *Proc. Workshop MONET: LNCS 4277*, 2006.

[19] E. Nordstrom, C. Diot, R. Gass, and P. P. Gunningberg. Experiences from measuring human mobility using bluetooth inquiring devices. In *Proc. MobiEval*, 2007.

[20] W. Paul and J. Baschnagel. *Stochastic processes from physics to finance*. Springer, 1999.

[21] A. M. Reynolds and C. J. Rhodes. The Lévy flight paradigm: random search patterns and mechanisms. *Ecology*, 90, 2009.

[22] I. Rhee, M. Shin, and Others. Human mobility patterns and their impact on routing in human-driven mobile networks. In *Proc. Hotnets-VI*, 2007.

[23] I. Rhee, M. Shin, and Others. On the Lévy-walk Nature of Human Mobility. In *Proc. INFOCOM*, 2008.

[24] M. F. Shlesinger, G. M. Zaslavsky, and U. Frisch. *Lévy flights and related topics in physics: Lecture Notes in Physics 450*. Springer, 1995.

[25] G. M. Viswanathan and Others. Lévy flight search patterns of wandering albatrosses. *Nature*, 381, 1996.

[26] G. M. Viswanathan and Others. Optimizing the success of random searches. *Nature*, 401, 1999.

[27] E. P. White, B. J. Enquist, and J. L. Green. On estimating the exponent of power-law frequency distributions. *Ecology*, 89, 2008.