

On Nonstationarity of Human Contact Networks

Salvatore Scellato*, Mirco Musolesi†, Cecilia Mascolo*, Vito Latora‡

*Computer Laboratory, University of Cambridge, United Kingdom

†School of Computer Science, University of St. Andrews, United Kingdom§

‡Dipartimento di Fisica, Università di Catania, Italy

Abstract—The measurement and the analysis of the temporal patterns arising in human networks is of fundamental importance to many application domains including targeted advertising, opportunistic routing, resource provisioning (e.g., bandwidth allocation in infrastructured wireless networks) and, more in general, modeling of human social behavior.

In this paper we present a novel and exhaustive study of the temporal dynamics of human networks and apply it to different sets of wireless network traces. We consider networks of contacts among users (i.e., peer-to-peer opportunistic networks). We show that we are able to quantify how the amount of information associated to the process evolves over time by using techniques based on time series analysis. We also demonstrate how regular patterns appear only at certain time scales: network dynamics appears nonstationary, in the sense that its statistical description is different at various time scales. These results provide a new methodology to accurately and quantitatively investigate the temporal properties of any type of human interactions and open new directions towards a better understanding of the regular nature of human social behavior.

I. INTRODUCTION

The accurate measurement and analysis of real-world contact and mobility patterns provide fascinating insights into the rhythms of people both as individuals and collective entities. These studies have many practical applications that include targeted advertising and recommendation [1], efficient opportunistic networking [13] and content sharing [26], epidemiological modeling [17], intelligent resource provisioning and balancing [34], and urban planning [29].

In the recent years, researchers have conducted large-scale measurements in Wi-Fi, cellular and ad hoc (in particular Bluetooth) environments; their efforts have led to the availability of several sets of traces for performance evaluation of wireless systems [18]. The characterization of the distribution of the contact durations (i.e., the durations of the contact times between the pairs of nodes of a network) and the inter-contact times (i.e., the average duration of the intervals between subsequent disconnections and connections of each pair of nodes) has been extensively studied [5], [15], [12]. These works, however, fail to measure or quantify the periodic aspects of human behavior. Some high-level temporal aspects of the dynamics of traces (in particular, patterns of people in campus and conference environments) have been studied by Kim et al. [16], Clauset and Eagle [7] and Scherrer et al. [32].

§This work was carried out while Mirco Musolesi was at the Computer Laboratory, University of Cambridge.

The main limitation of these approaches is that they are based on qualitative observations and are unable to help the forecast of the future dynamics of the emerging human networks.

In this paper, we present a rigorous analysis of the temporal patterns of human networks based on a novel methodology founded on results from complex network theory [4] and linear and nonlinear time series analysis [14]. We consider networks emerging from contacts among users (e.g., networks representing peer-to-peer interactions without the mediation of a central infrastructure). In these networks, the nodes are the users and an edge exists if two users are colocated (e.g., if the users are in Bluetooth transmission range or under the same access point).

In the context of this work, nonstationarity accounts for the fact that the temporal properties of human contact networks appear homogeneous at a given time scale, but, instead, they are heterogeneous across different time scales. Thus, our approach aims to unveil key aspects of these human networks, namely the *regular nature* of their temporal dynamics and the *temporal heterogeneity* of their statistical properties, such as their average number of contacts and the efficiency of the network itself. More specifically, we exploit diffusion entropy analysis [30] to investigate the amount of randomness in their temporal evolution and wavelet transform [23] to study their stationarity and uncover the presence of regular patterns at various time scales.

As a main contribution of this paper, we show that the presence of temporal structures embedded in human network dynamics can be accurately quantified and evaluated by analyzing the temporal evolution of their metrics. We describe a novel quantitative approach based on the concept of diffusion entropy to estimate how the amount of information in the dynamics of the network under consideration evolves over time. We show that statistical properties of human networks are non-homogeneous over time and are characterized instead by distinct patterns which can be isolated and unveiled with wavelet decomposition.

Most of the proposed socially-aware systems based on the evolution of social network dynamics such as opportunistic and delay-tolerant communication protocols [22], [34], [8], [5] are designed on the assumption of the stationary nature of human contacts. Our result shows instead that the assumption of stationarity for human behavior does not map to reality and, furthermore, we describe how to evaluate and identify nonstationary structures in the temporal dynamics. This quantitative analysis of nonstationarity can greatly improve the design of

such systems. In the following section we outline the key steps of our investigation mapping it to the structure of the paper.

II. THE APPROACH AT A GLANCE

Human networks can be temporally characterized by considering the sequence of contacts among people, in particular, by measuring the contact duration and the time passed between two consecutive contacts. Various studies have been conducted using wireless devices to gather data of this kind over real networks. From the analysis of these data our aim is to understand and quantify the properties of the underlying process that rules human dynamics. Our analysis proceeds through these steps:

- 1) Human networks are modeled by means of a time-varying graph, which is periodically sampled in order to obtain a sequence of static graph snapshots: a set of fundamental network metrics (such as degree, clustering coefficient, global efficiency, size of the giant component) are evaluated on the sampled graphs; for each metric, we extract a time series of values describing the evolution of that particular metric over time (Section IV).
- 2) These time series are analyzed to understand if their statistical evolution exhibits deviation from random behavior: diffusion entropy analysis [30] quantifies the degree of complexity of the network dynamics (Section V).
- 3) Finally, we show how deviation from pure randomness may be due to recurrent regular structures that are embedded in the time series: we describe how wavelet analysis [23] can be used to uncover these nonstationary patterns and to identify the appropriate time scale at which they appear (Section V).

III. DATASETS

This paper relies on the following datasets to show the performance of our analysis:

- 1) **Dartmouth Wi-Fi** This dataset was extracted from the SNMP logs of the Wi-Fi LAN of the Dartmouth College campus. Data about traffic in the access points were collected between 2001 and 2004 through three techniques: syslog events, SNMP polls, and network sniffers [12], [19].
- 2) **Ile Sans Fils** Ile Sans Fils [21] is a non-profit organization which operates a network of free WiFi hotspots in Montreal, Canada. It now counts over 45,000 users more than 180 hotspots located in publicly accessible spaces. These hotspots are deployed mostly in cafes, restaurants, bars and libraries, but also outdoor to cover parks and sections of popular commercial streets.
- 3) **Reality Mining** The Reality Mining experiment was held in 2004 at MIT [7]: 100 smartphones equipped with a sensing application were distributed to staff and students of the Media Lab and the Sloan Business School; data about their usage were collected for a period of 9 months. Collected information includes call logs, Bluetooth devices in proximity and cell tower IDs among the others.

Network	N	$\langle k \rangle$	$\langle C \rangle$	E_G
Dartmouth	5338	164.0	0.44	0.44
Ile Sans Fils	5297	11.8	0.53	0.21
Reality Mining	94	28.7	0.63	0.64

TABLE I

NETWORK METRICS COMPUTED ON THE GRAPH CONTAINING ALL THE CONTACTS AMONG USERS FOR THE NETWORKS EXTRACTED FROM THE DATASETS UNDER ANALYSIS: NUMBER OF USERS N , AVERAGE DEGREE $\langle k \rangle$, AVERAGE CLUSTERING COEFFICIENT $\langle C \rangle$, GLOBAL EFFICIENCY E_G [20].

Our analysis is based on the concept of contacts between individuals and between individuals and access points. Dartmouth and the Ile Sans Fils datasets contain associations between wireless devices and access points: for these datasets we have analyzed both contacts between people and one access point and inferred contacts between individuals by assuming that two people are in contact when they are connected to the same access point. The Reality Mining dataset contains person-to-person contacts. For each dataset, we consider a time window of 4 weeks and we analyze all the contacts within this period. In Table I we report some network properties computed on the graphs containing all the contacts within each dataset.

IV. EXTRACTING TIME SERIES FROM HUMAN NETWORKS

In this section, we describe the extraction of time series from human networks. First we model the traces with a time-varying network which can be suitably sampled to obtain a sequence of static graphs. Then, we calculate the network metrics for each graph of the sequence obtaining a time series of scalar values.

A. Sampling Temporal Graphs

Contact traces are represented as sets of entries in the form (i, j, t_k, t_{k+1}) , with the meaning that entities i and j are in reach between time t_k and time t_{k+1} .

Every set of contact traces may be modeled as a *temporal network* $\mathcal{G}(t)$, where edges between two nodes may appear and disappear as time passes, while the number of nodes N is kept constant in order to only capture the dynamics and not the network growth. This embedding graph can then be sampled with a suitable *sampling interval* Δ to obtain a sequence of static graph snapshots G_1, G_2, \dots, G_N : two nodes i and j are connected in the graph G_k if they are connected at any time and for any duration within the time interval $[k\Delta, (k+1)\Delta]$ with $k = [0 \dots N-1]$ and N equals the number of slots in the time interval.

This sequence of graphs can be analyzed with the tools of network analysis to obtain scalar series of data or global statistics about the entire sequence. We present this methodology in detail in the next section.

B. Network Metrics

We analyze contact traces from a global point of view, focusing on some properties of the time-varying network built among all the nodes.

Network	$\langle k \rangle$	$\langle C \rangle$	E_G	N_{GC}
Dartmouth	1.05	0.99	1.08	1.01
Ile Sans Fils	0.86	0.70	0.85	0.72
Reality Mining	0.88	0.83	0.90	0.86

TABLE II

DIFFUSION EXPONENTS δ FOR THE TIME SERIES EXTRACTED FROM THE THREE CONTACT NETWORKS DARTMOUTH, ILE SANS FILS AND REALITY MINING. A DIFFUSION EXPONENT $\delta = 0.5$ REPRESENTS THE DIFFUSING BEHAVIOR OF A PURELY RANDOM UNCORRELATED PROCESS, WHILE LARGER EXPONENT DENOTE CORRELATED FLUCTUATIONS.

We are interested in the following properties of each graph in the sequence:

- *node properties*: average node degree $\langle k \rangle$ and average node clustering $\langle C \rangle$ [4].
- *network properties*: the size of the giant component N_{GC} and the global efficiency E_G , which is the harmonic mean of all the shortest path lengths in the graph [25]. Such a quantity is a good indicator of the traffic capacity of a network.

After extracting the sequence of graphs from the contact traces (as explained above), we calculate the metrics for each snapshot graph and we obtain the time series for the properties of network under analysis.

We choose a sampling interval equal to $\Delta = 3600$ seconds (1 hour), since this represents a reasonable trade-off which allows us to capture the dynamical evolution of the different human networks without obtaining a large amount of redundant data due to oversampling. Contact traces with different time spans may be analyzed at larger or smaller sampling rates.

V. NONSTATIONARITY AND TEMPORAL PATTERNS

In this section we investigate how the statistical properties of the time series extracted from a human network evolve over time. We are interested in nonstationary behavior on time scales which are comparable to the time scales of typical human activities. As a consequence, we want to prove that the evolution of human contacts is not simply ruled by a stationary stochastic process but, instead, it can be better described and interpreted by means of structured and recurrent temporal patterns that are at the basis of people behavior. More specifically, we answer to the following set of questions:

- *How is the information embedded in the time series evolving over time?* We use the concept of diffusion entropy [30] to uncover scaling behavior in the fluctuations of a time series. Diffusion entropy quantifies the amount of information over time of a time series and gives an estimation of how much the underlying process is different than a purely Gaussian random process.
- *Are the properties of the time series changing with time? How is the temporal information divided across different time scales?* In order to investigate whether time series evolution exhibits nonstationarity, we make use of wavelet analysis [23] to uncover recurrent patterns embedded in the time series at different temporal scales and

to understand how various time scales convey different information about human network dynamics.

A. Diffusion Entropy Analysis

After discovering that many natural phenomena follow certain scaling laws [24], Diffusion Entropy Analysis (DEA) has been put forward as a technique for measuring the correlated variations in time series. DEA is based on the diffusion process generated by the time series itself; it measures the time evolution of the Shannon entropy of the probability density function of this process [30].

As in a Brownian-like trajectory, we interpret the values of a time series (s_0, \dots, s_N) as the steps of a diffusion process. We define the trajectories of this process by the cumulative sum of these steps, obtaining a different trajectory $x_k(t)$ for each different value of k :

$$x_k(t) = \sum_{i=0}^t s_{k+i}, \quad 0 \leq k < N - t + 1 \quad (1)$$

where t is the length of the diffusion process. Then, we compute the probability distribution function $p(x, t)$, which describes the probability that a given trajectory has a displacement of x after t steps.

For any particular t , we compute the temporal Shannon entropy of the probability distribution at time t as $S(t) = -\sum_x p(x, t) \log p(x, t)$. For diffusive processes the temporal Shannon entropy $S(t)$ of the probability distribution $p(x, t)$ is evolving according to the equation [30]:

$$S(t) = \delta \log t + A \quad (2)$$

where A is a constant and δ is the *diffusion exponent*, which can be numerically evaluated. For a random uncorrelated process with finite variance, the diffusion distribution $p(x, t)$ will converge, according to the central limit theorem, to a Gaussian distribution which exhibits $\delta = 0.5$. When dealing with discrete time series, the function $p(x, t)$ can be estimated through discrete histograms $p_{x,t}$, where the bin size should be equal to the standard deviation of the obtained diffusion process in order to normalize the results. As a consequence, the diffusion entropy can be computed as:

$$S(t) = -\sum_x p_{x,t} \log p_{x,t} \quad (3)$$

The estimated diffusion exponent *delta* can be computed through a first-order least-squares fit in a linear-log scale. Even though time series extracted from human networks may not show a pure scaling behavior as in Equation (2) but, instead, patterns with oscillations due to periodicities, we can still observe how entropy grows linearly with time and we can estimate the diffusion exponent with reasonable accuracy.

As an example, in Figure 1 we show diffusion entropy as a function of time for two different time series. We have found similar linear trends in the other time series of all networks. We report the estimated diffusion exponents in Table II. All values are larger than $\delta = 0.5$, which is the diffusion exponent

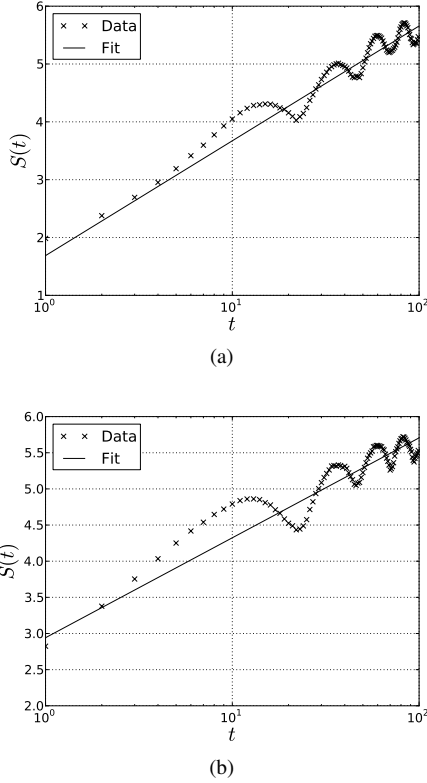


Fig. 1. Diffusion entropy as a function of time for $\langle k \rangle$ (a) and $\langle C \rangle$ (b) time series for the Ile Sans Fils contact network (sampling rate $\Delta = 3600s$). Data are fitted with Equation (2) which gives, respectively, $\delta = 0.86$ for (a) and $\delta = 0.70$ for (b), thus revealing more correlated evolution in the first case.

of a pure Brownian motion, thus the temporal evolution of the time series exhibits diffusive fluctuations that cannot be modeled or explained by a random Gaussian diffusion process. This denotes statistical properties of a time series which might require different models than purely Gaussian fluctuations. As a consequence, we know that there is a valuable amount of information to be exploited in these time series, both for modeling and prediction purposes.

Furthermore, we can quantify how much information is embedded in the temporal structures of the time series: from a general point of view, the Dartmouth contact network appears as the more regular and less random, while the Ile Sans Fils one results in lower exponents and, hence, in less regularity.

As an example, the time series of the average clustering coefficients extracted from the Ile Sans Fils contact network shows a diffusion exponent which is close to the value that indicates pure Gaussian fluctuations, probably because public access points are more likely to register lonely individuals rather than entire groups of people which meet regularly at the same place. Therefore, it is more difficult to identify clear patterns in the temporal evolution of this metric, typical of strong social structures.

On the other hand, even though the Dartmouth contact network is also generated from contacts with public access points, its temporal evolution appears more regular, with larger

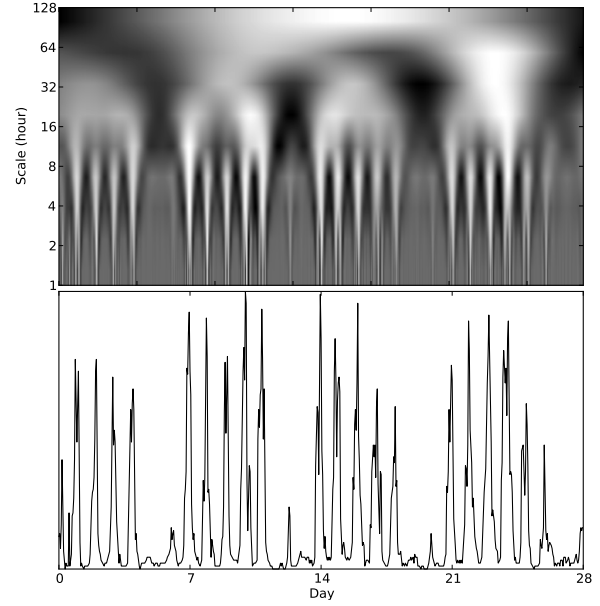


Fig. 2. Original time series (below) and DWT wavelet components at different time scales (above) for $\langle k \rangle$ for the Reality Mining network (sampling rate $\Delta = 3600s$). For each wavelet components coefficients are rendered in grayscale with bilinear interpolation: time is shown on the x-axis and different time scales are reported on the y-axis.

exponents. This may be due to the fact that students and people working at the Dartmouth campus are more likely to exhibit uniform time schedules than strangers meeting around the city. Through the calculation of the diffusion exponent we can quantitatively understand how temporal evolution of network metrics deviates from uncorrelated randomness, thus revealing correlations that can be identified and exploited.

B. Discrete Wavelet Transform Analysis

Discrete Wavelet Transform (DWT) is a mathematical technique for the analysis of scalar time series which divides the original signal into components according to both frequency and time. Wavelet analysis provides a multi-scale analysis of the time series, extracting information at different time scales and at different time instants. With respect to the Discrete Fourier Transform, which transforms a time series from the time domain to the frequency domain, DWT is still in the time domain but it results in different components at multiple time scales. Therefore, by applying this technique, we can understand not only how the information contained in the time series is divided across different time scales, i.e., at which frequency the time series contains more variability, but also precisely at what time instant the time series contains meaningful information.

The output of the wavelet decomposition of a time series is a set of smaller numerical sequences, each one half the size of the previous but with doubled time scale: hence, longer sequences represent shorter time scales (i.e., higher frequencies), while shorter sequences capture patterns emerging at larger time scales (i.e., lower frequencies). As an example, for

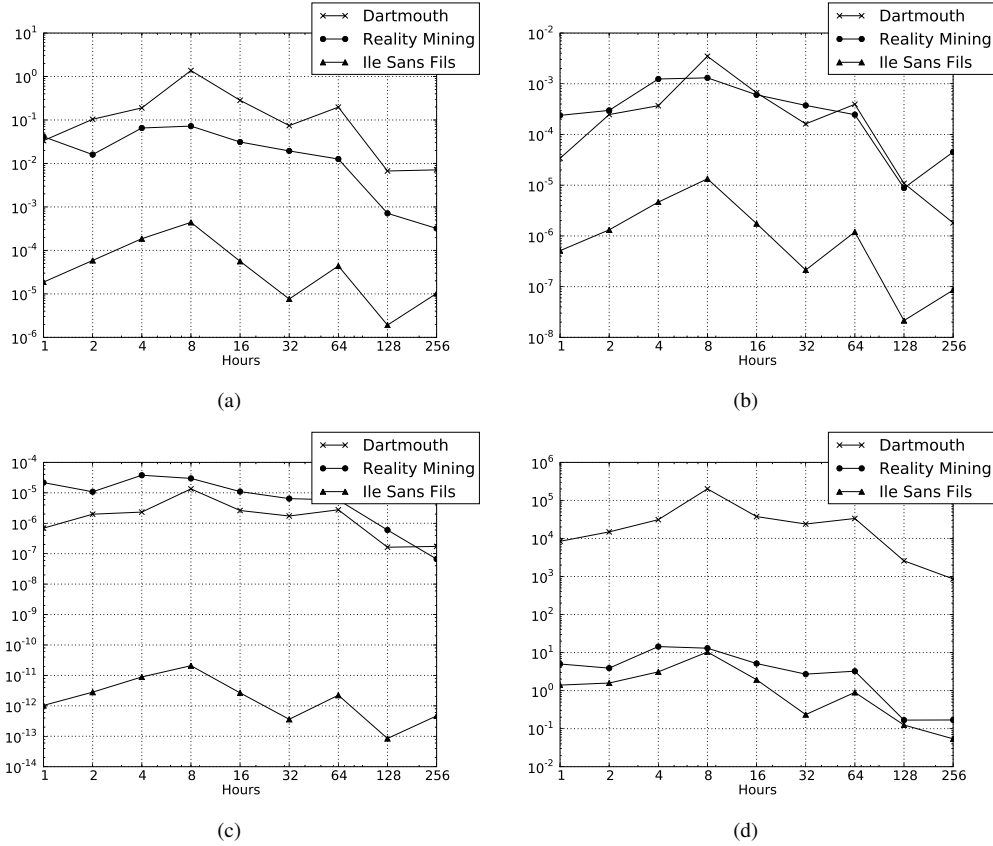


Fig. 3. Wavelet spectra for (a) $\langle k \rangle$, (b) $\langle C \rangle$, (c) E_G and (d) N_{GC} for the three contact networks. Each point represents the statistical variance contained in the corresponding time scale.

a time series of 128 values a wavelet decomposition of order 3 generates the first wavelet component (time scale 1) with 64 values, the second one (time scale 2) with 32 values and the third one (time scale 4) with 16 values. Since we adopt a sampling rate $\Delta = 3600s$ for all the time series, wavelet time scales can be directly expressed in hours. More details about this technique can be found in the appendix and in [23], [10].

In Figure 2 we show how the wavelet decomposition of a time series highlights patterns emerging only at certain time scales: in this example regular structures can be identified both at daily and weekly granularities. In this wavelet map all the wavelet components are vertically represented in grayscale next to each other, with their values aligned to the point in time they are referred to: this allows the graphical representation of regular structures at different wavelet scales. In particular, daily patterns for the time series of the average degree $\langle k \rangle$ are visible at time scales of 4 and 8 hours, while weekly patterns emerge only at higher time scales of 32 and 64 hours.

Given such a wavelet decomposition, we can also compute a wavelet power spectrum, which describes how the variance of the original series is divided across different time scales. Thus, we can investigate which time scales hold more information within the time series. For a stationary random uncorrelated time series, the wavelet spectrum exhibits a monotonic decreasing trend. In fact, stationarity implies that,

on large time scales, the statistical description is constant and, therefore, fluctuations are more evident on shorter periods of time. Local maxima in the wavelet spectrum indicate that the characteristics of the time series significantly change over time when observed at some particular time scale with respect to the adjacent scales. Hence, we can understand how and when the time series shows nonstationarity and whether we can identify temporal patterns. However, we see that this is not the case for human networks. In Figure 3 we show the wavelet power spectra of the six time series for the three contact networks. The time series have wavelet spectra which do not simply decrease with the time scale, but which instead present local maxima (or, at least, a non-decreasing trend) at the time scales 4 and 8 hours and then another one around the 64-hour time scale: this is due to the fact that these metrics exhibit a strong variation when the time series is aggregated at intervals of 8 hours. This is also in agreement with the anecdotal evidence of the division of human daily activities in segments of 8 hours [15]. Moreover, the maximum at 64 hours represents a time scale of roughly 3 days and this can be interpreted as the weekly variation of human behavior in weekdays and weekend days, since consecutive groups of three days are likely to show different properties.

Thus, by using wavelet decomposition we are able to see how regular patterns results in nonstationary characteristics

over time. The appearance of these patterns is crucial to understand and appropriately model human behavior.

VI. RELATED WORK

Analytical techniques such as wavelet transform and diffusion entropy analysis have been often applied to the analysis of time series arising from demographic data [28], physical processes [6] and biological series [31]. Investigation of nonstationarity has also been extensively done on time series of econometrics data [9] to uncover hidden trends and seasonalities.

A number of pioneering works [2], [3], [12] have focussed on contact traces and registration patterns in order to gain insight about human movement and the emerging social network. A key study in this area is the work on connectivity patterns presented by Chaintreau et al. in [5]. The main finding of this work is that contacts duration and inter-contacts time between individuals can be represented by means of power-law distributions and that these patterns may be used to develop more efficient opportunistic protocols. The work confirms the results of other studies conducted at Dartmouth [12], UCSD [27] and University of Toronto [35]. At the same time, it is interesting to note that these observed connectivity patterns are at odds with those that can be extracted from random mobility models that show an exponential decay of inter-contacts time intervals [33]. In a previous work [13], similar connectivity patterns have also been observed among the participants of INFOCOM'05.

Karagiannis et al. in [15] offered a novel perspective to the problem of the approximation of these distributions. The authors first verify the power-law decay of inter-contacts time CCDF between mobile devices. Then, they demonstrate that beyond a characteristic time of about 12 hours the CCDF exhibits exponential decay.

More recently, Scherrer et al. [32] present an analysis of the dynamical properties of two datasets characterizing contact graphs over time. A visual analysis of temporal features of the Reality Mining traces [11] is presented in [7]. However, the authors of this work focus only on a standard statistical characterization of the properties of human networks. Instead, we have investigated the nonstationarity (including distribution of information content at different time scales) and the presence of regular temporal patterns in these time series.

With respect to the state of the art, as a novel contribution, we have shown how temporal properties of human networks are heterogeneous over time and how patterns can be identified only at certain time scales. We have demonstrated how these temporal properties can be quantitatively studied and evaluated in order to better understand the dynamics of human networks.

VII. CONCLUSIONS AND FUTURE WORK

The key contributions of this work lie in three different areas. First, we have proposed a framework for the analysis of the evolution of human networks considering the dynamical properties of time series of network metrics. Then, we have proven the complex nonstationary nature of human networks

by identifying regular temporal patterns with sub-daily, daily and weekly periodicities. Finally, we have shown how these patterns can be exploited to predict future evolution of network metrics with better accuracy, also unveiling their nonlinear properties at different time scales.

These results give several insights on the dynamics of human contacts. They show how the assumption of stationarity for human behavior that is usually at the basis of most systems and models often does not map reality. Nonstationary characteristics should be instead considered in the design of systems and performance evaluation tools for mobile networking and systems. We have shown how wavelet decomposition is a powerful technique which allows for the isolation and the identification of recurrent patterns in the dynamics of the human networks at the basis of opportunistic and infrastructured wireless systems.

The applications of these results to the design of novel socially-aware and social network founded models and systems are indeed promising. Ultimately, these findings can be successfully exploited to automatically identify human temporal patterns, to model time-varying human dynamics and to enhance the design of systems based on human contacts such as online social networks and other social information platforms (e.g. blog networks, user-generated media publishers, collaborative frameworks).

Acknowledgments This work was supported through EP-SRC grants EP/C544773, EP/D077273 and EP/E012914. We thank Manlio De Domenico for insightful discussions about time series analysis methods and techniques.

REFERENCES

- [1] G. Adomavicius and A. Tuzhilin. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.
- [2] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan. Characterizing User Behavior And Network Performance In A Public Wireless LAN. In *Proceedings of SIGMETRICS'02*, pages 195–205, New York, NY, USA, 2002. ACM Press.
- [3] M. Balazinska and P. Castro. Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network. In *Proceedings of MobiSys'03*, San Francisco, CA, May 2003.
- [4] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. Complex Networks: Structure and Dynamics. *Physics Reports*, 424(4-5):175–308, February 2006.
- [5] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of Human Mobility on Opportunistic Forwarding Algorithms. *IEEE Transactions on Mobile Computing*, 6(6):606–620, 2007.
- [6] C. Chandre, S. Wiggins, and T. Uzer. Time-frequency Analysis of Chaotic Systems. *Physica D: Nonlinear Phenomena*, 181(3-4):171 – 196, 2003.
- [7] A. Clauset and N. Eagle. Persistence and Periodicity in a Dynamic Proximity Network. In *Proceedings of DIMACS Workshop on Computational Methods for Dynamic Interaction Networks*, September 2007.
- [8] E. M. Daly and M. Haahr. Social Network Analysis For Routing In Disconnected Delay-Tolerant MANETs. In *Proceedings of MobiHoc'07*, pages 32–40, New York, NY, USA, 2007. ACM.
- [9] O. Darné and C. Diebolt. Non-stationarity Tests in Macroeconomic Time Series. In *New Trends in Macroeconomics*. Springer Berlin Heidelberg, 2005.
- [10] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial & Applied Math, December 1992.

- [11] N. Eagle and A. S. Pentland. Reality Mining: Sensing Complex Social Systems. *Personal Ubiquitous Computing Journal*, 10(4):255–268, 2006.
- [12] T. Henderson, D. Kotz, and I. Abyzov. The Changing Usage Of A Mature Campus-wide Wireless Network. In *Proceedings of MobiCom'04*, pages 187–201, New York, NY, USA, 2004. ACM Press.
- [13] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pocket Switched Networks and Human Mobility in Conference Environments. In *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pages 244–251, August 2005.
- [14] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, 1999.
- [15] T. Karagiannis, J.-Y. L. Boudec, and M. Vojnović. Power Law And Exponential Decay Of Inter Contact Times Between Mobile Devices. In *Proceedings of MobiCom'07*, pages 183–194, New York, NY, USA, 2007. ACM.
- [16] M. Kim and D. Kotz. Periodic properties of user mobility and access-point popularity. *Journal of Personal and Ubiquitous Computing*, 11(6), August 2007.
- [17] J. Kleinberg. The Wireless Epidemic. *Nature*, 449(7160):287–288, 2007.
- [18] D. Kotz and T. Henderson. CRAWDAD: A Community Resource for Archiving Wireless Data at Dartmouth. *IEEE Pervasive Computing*, 4(4):12–14, October-December 2005.
- [19] D. Kotz, T. Henderson, and I. Abyzov. CRAWDAD trace dartmouth/campus/movement/01_04 (v. 2005-03-08). Downloaded from <http://crawdada.cs.dartmouth.edu/>, March 2005.
- [20] V. Latora and M. Marchiori. Efficient Behavior of Small-World Networks. *Physical Review Letters*, 87(19), October 2001.
- [21] M. Lenczner, B. Grégoire, and F. Proulx. CRAWDAD data set ileansfil/wifidog (v. 2007-08-27). Downloaded from <http://crawdada.cs.dartmouth.edu/ileansfil/wifidog>, Aug. 2007.
- [22] A. Lindgren, A. Doria, and O. Schelén. Probabilistic Routing In Intermittently Connected Networks. In *Proceedings of SAPIR'04*, August 2004.
- [23] S. Mallat. *A Wavelet Tour of Signal Processing, Second Edition*. Academic Press, September 1999.
- [24] B. B. Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman, August 1983.
- [25] M. Marchiori and V. Latora. Harmony in the Small-world. *Physica A*, 285:539–546, 2000.
- [26] L. McNamara, C. Mascolo, and L. Capra. Media Sharing Based On Colocation Prediction In Urban Transport. In *Proceedings of MobiCom'08*, pages 58–69, New York, NY, USA, 2008. ACM.
- [27] M. McNett and G. M. Voelker. Access and Mobility of Wireless PDA User. *Mobile Computing Communications Review*, 9(2):40–55, April 2005.
- [28] J. B. Ramsey and C. Lampart. Decomposition of Economic Relationships by Timescale using Wavelets. *Macroeconomic Dynamics*, 2(01):49–71, 1998.
- [29] J. Reades, F. Calabrese, A. Sevtsuk, and C. Ratti. Cellular Census: Explorations in Urban Data Collection. *IEEE Pervasive Computing*, 6(3):30–38, 2007.
- [30] N. Scafetta and P. Grigolini. Scaling Detection in Time Series: Diffusion Entropy Analysis. *Phys. Rev. E*, 66(3):036130, Sep 2002.
- [31] N. Scafetta, V. Latora, and P. Grigolini. Lévy Scaling: The Diffusion Entropy Analysis Applied to DNA Sequences. *Phys. Rev. E*, 66(3):031906, Sep 2002.
- [32] A. Scherrer, P. Borgnat, E. Fleury, J. L. Guillaume, and C. Robardet. Description And Simulation Of Dynamic Mobility Networks. *Computer Networks*, 52(15):2842–2858, 2008.
- [33] G. Sharma and R. R. Mazumdar. Scaling laws for capacity and delay in wireless ad hoc networks with random mobility. In *IEEE International Conference on Communications (ICC'04)*, pages 3869–3873, June 2004.
- [34] L. Song, U. Deshpande, U. Kozat, D. Kotz, and R. Jain. Predictability Of WLAN Mobility And Its Effects On Bandwidth Provisioning. In *Proceedings of INFOCOM'06*, April 2006.
- [35] 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 (WMCSA'04)*, pages 41–50, Washington, DC, USA, 2004. IEEE Computer Society.

Given a time series $\mathbf{X} = (x_0, x_1, \dots, x_{N-1})$, the Discrete Wavelet Transform (DWT) is formally a decomposition $\mathbf{W} = \mathbf{D}\mathbf{X}$, where $\mathbf{W} = (w_0, w_1, \dots, w_{N-1})$ is a vector containing the wavelet coefficients and \mathbf{D} is a $N \times N$ real-valued orthonormal matrix which defines the DWT and satisfies the orthogonality condition $\mathbf{D}^T \mathbf{D} = \mathbf{I}_N$.

If $N = 2^J$ then the elements in \mathbf{W} can be arranged such that the first $N/2$ elements are associated to the unit time scale changes, the next $N/4$ elements with changes on a doubled scale and so forth. There are thus exactly $N/(2\tau_j)$ coefficients in the subvector \mathbf{W}_j associated to changes on scale τ_j , where $\tau_j = 2^{j-1}$ for $j = 1, \dots, J$. It is worth noting that τ_j is an adimensional unit, but, in practical applications, we only need to consider the sampling interval Δt between consecutive values in \mathbf{X} to obtain a proper temporal scale. As a consequence, each τ_j corresponds to a physical scale of $\tau_j \Delta t$ with appropriate measure units.

The orthogonality implies that $\mathbf{X} = \mathbf{D}^T \mathbf{W}$ and $\|\mathbf{W}\|^2 = \|\mathbf{X}\|^2$, so $\|\mathbf{W}_j\|^2$ represents the contribution to the energy associated to the j -th DWT subvector. This decomposition provides a description of how the energy of the original signal is divided into various contributions for the different time scales. Thus, we can define a discrete empirical wavelet power spectrum $P_W(\tau_j) = \frac{1}{N} \|\mathbf{W}_j\|^2$ such that $\sum_{j=i}^J P_W(\tau_j) = \sigma_X^2$, where σ_X^2 is the variance of the original series \mathbf{X} .

A computationally efficient algorithm which runs in $O(N)$ is based on filter banks [23]. This is the algorithm we adopt in our work, using the wavelet D4 (Daubechies 4) [10]. The signal is processed separately with a high-pass and a low-pass filter, related to each other so that they form a quadrature mirror filter, to yield a smooth version and a detailed version of the signal. Then the outputs are downsampled by a factor of 2 so that only half of the detailed version is kept as a portion of the wavelet coefficients and half of the smooth version is used as the input for a new level. This decomposition is further repeated with the high and low pass filters and then the result is downsampled again and so forth, until all the wavelet coefficients have been determined.

Similarly, stationary wavelet components can be obtained with the same algorithm if the downsampling step is avoided, so that each component will have the same size. However, avoiding the downsampling requires to extend the values of the time series in the future. While the typical boundary treatment includes periodic or reflecting boundaries, we adopt a different approach and we extend the original signal with a nonlinear predictor [14] before computing the smooth versions at the different scales.