A Complex Network Analysis of Human Mobility

Theus Hossmann
Communication Systems Group
ETH Zurich, Switzerland
lastname@tik.ee.ethz.ch

Thrasyvoulos Spyropoulos
Mobile Communications
EURECOM, France
firstname.lastname@eurecom.fr

Franck Legendre
Communication Systems Group
ETH Zurich, Switzerland
lastname@tik.ee.ethz.ch

***Abstract*—Opportunistic networks use human mobility and consequent wireless contacts between mobile devices, to disseminate data in a peer-to-peer manner. To grasp the potential and limitations of such networks, as well as to design appropriate algorithms and protocols, it is key to understand the statistics of contacts. To date, contact analysis has mainly focused on statistics such as inter-contact and contact distributions. While these pair-wise properties are important, we argue that *structural properties* of contacts need more thorough analysis. For example, communities of tightly connected nodes, have a great impact on the performance of opportunistic networks and the design of algorithms and protocols.**

In this paper, we propose a methodology to represent a mobility scenario (i.e., measured contacts) as a weighted *contact graph*, where tie strength represents how long and often a pair of nodes is in contact. This allows us to analyze the structure of a scenario using tools from complex network analysis and graph theory (e.g., community detection, connectivity metrics). We consider four mobility scenarios of different origins and sizes. Across all scenarios, we find that mobility shows typical small-world characteristics (short path lengths, and high clustering coefficient). Using state-of-the-art community detection, we also find that mobility is strongly modular. However, communities are not homogenous entities. Instead, the distribution of weights and degrees *within* a community is similar to the *global* distribution of weights, implying a rather intricate intra-community structure.

To the best of our knowledge, this is the most comprehensive study of structural characteristics of wireless contacts, in terms of the number of nodes in our datasets, and the variety of metrics we consider. Finally, we discuss the primary importance of our findings for mobility modeling and especially for the design of opportunistic network solutions.

I. Introduction

The rapid proliferation of small wireless devices creates ample opportunity for novel applications [1], as well as for extending the realm of existing ones. *Opportunistic* or *Delay Tolerant Networking* (DTN) [2] is a novel networking paradigm that is envisioned to complement existing wireless technologies (cellular, WiFi) by exploiting a *niche* performance-cost tradeoff. Nodes harness unused bandwidth by exchanging data whenever they are within mutual wireless transmission range of each other (*in contact*).

Since every contact is an opportunity to forward content and bring it probabilistically closer to a destination (or a set of destinations), understanding statistical properties of contacts is vital for the design of algorithms and protocols for opportunistic networks. To this end, a number of efforts have been made to collect mobility traces at various scales using different methods [3], [4], [5], [6].

Analysis of such traces has led to several important findings: On an individual level, mobility patterns exhibit time-of-day periodicity and strong location preference [7]. The amount of regularity observed implies high (statistical) predictability of these patterns [8]. On a pairwise level, most trace analysis research has focused on *inter-contact* and *contact duration* statistics [9], [10] and it is still debatable whether these distributions are power-law, have exponential tail, or have qualitatively different behavior from on pair to another. More recently, studies have focused on the *macroscopic* structure of contacts such as tightly knit communities of nodes that meet each other frequently. Such community structure is generally assumed because of the social nature of human mobility. In fact, some state-of-the-art mobility models [11] explicitly model community structure, and most recent DTN routing protocols [12], [13] exploit structural characteristics. However, there are few studies that systematically measure the macroscopic structure of mobility. Notable exceptions are [12], reporting high modularity of contacts (i.e., strong community structure), [14] finding short average path length in opportunistic networks, and [15] analyzing the time-dependence of communities.

In this paper, we provide a thorough and extensive analysis of the *structural properties of contacts*. To do so, we represent contacts in a compact and tractable way as a weighted *contact graph*, where the weights (i.e., tie strengths) express how frequently and how long a pair of nodes is in contact. Given such a contact graph, we can use tools and metrics from social network analysis and graph theory (e.g., connectivity metrics, community detection, etc.) to quantify the amount of structure in the underlying mobility scenario. Our main findings and contributions can be summarized as follows:

i) Our study is based on 4 contact traces ranging in size from ~ 100 to ~ 1000 nodes, hence, we add one order of magnitude compared to prior works (which have only analyzed contact graphs of ~ 100 nodes). We also present a completely new trace reporting the whereabouts of users using a popular location-based smartphone application (Sec. II).

ii) We provide thorough evidence that the structure of mobility has *small-world* properties typically observed in social networks. While previous work [14] has only shown that there exist short space-time paths (possibly formed by random, *unpredictable* contacts), our methodology allows to conclude that the strong and *predictable* structure is small-world (Sec. III).

iii) We show that contacts are strongly modular, i.e., that there are close-knit communities of people with strong mobility ties. We study and compare the distributions of tie strengths and node degrees *within* and *across* communities. We find high

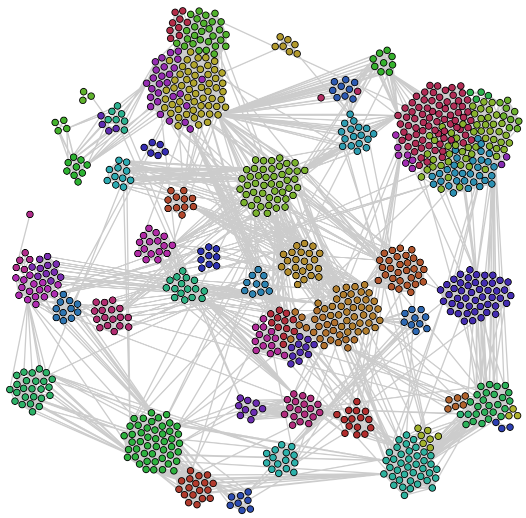


Fig. 1: Contact Graph of DART trace.

variance and heavy tails on both levels, implying that communities are by no means homogenous entities. In all traces, we find surprisingly similar distributions within communities and on a global level, suggesting that a community itself has properties similar to the entire network (Sec. IV).

iv) We discuss the implications of these findings for the design of opportunistic routing protocols and mobility modeling (Sec. V).

II. Datasets and Contact Graph

In this section, we detail the mobility scenarios we use in our study, and how we derive contact traces from them (II-A). Finally, we explain our methodology of creating an aggregated contact graph from such a contact trace (II-B).

A. Mobility Scenarios and Contact Definition

We define a *contact* as the period of time during which two devices are in mutual radio transmission range and can exchange data. For our analysis, we use four mobility traces collected in different contexts and with different methods: WLAN Access Point associations from the Dartmouth and ETH Zurich campuses, Bluetooth contacts from the MIT campus and self-reported "check-ins" from a popular geographic social network, Gowalla. In the following we describe these scenarios and how we derive contact traces from them.

Dartmouth (DART) We use 17 weeks of the WLAN Access Point (AP) association trace [3], between Nov. 2nd 2003 and Feb. 28th 2004. We choose the 1044 nodes which have activities at least 5 days a week on average i.e., the nodes have associations at least $5 \times 17 = 85$ days. The trace is preprocessed to remove short disconnections ($< 60s$) which we attribute to interference and other non-mobility related effects, as well as the well known ping-pong effect where devices jump back and forth between different APs in reach. We assume that two nodes are in contact when they are associated to the same AP at the same time.

ETH Campus (ETH) Using the same methodology as for Dartmouth, we process a trace collected at the ETH campus [4] during almost 15 weeks between Oct. 25th 2006 and Feb. 3rd 2006. Similarly, we choose 285 nodes which connect to the WLAN AP network at least 5 days a week (i.e., 75 days).

Gowalla (GOW) Gowalla[1] is a location-based service where users check-in to close-by spots (e.g., restaurants, office buildings), thereby logging their position. We use the publicly available location data of 473 heavy-users who, during the 6 months from Apr. to Sept. 2010, check-in at least 5 days a week somewhere in the State of Texas in the United States. Since users only check-in and do not check-out, we cannot infer the stay duration at a spot. Therefore, we assume users are in contact when they check-in less than 1 hour apart at the same spot. As we do not know the duration of a contact, we assume all contacts have the same duration (1 hour[2]).

MIT Reality Mining (MIT) The MIT trace [6] logs contacts between 92 campus students and staff, detected by Bluetooth scans with a scanning interval of 5 minutes. We take a 3 months long piece of the trace. It is the only trace we use where contacts are measured directly (and not inferred from location). However, supposedly, many short contacts are not registered due to the relatively long scanning interval.

Note that these traces differ vastly in their nature, and different traces capture different aspects of mobility. For instance the Gowalla trace, by the nature of the application, mainly captures the mobility of users while they go out and socialize. The DART trace captures students and staff at home and at the university, whereas the ETH trace captures only work behavior, since there are no APs in residential buildings. MIT is the smallest trace but captures work, home and leisure equally. We believe that due to the variety of the traces, our results are general even though individual traces may be biased. Details about all these traces are listed in Table I.

B. Aggregation of Contact Traces to Contact Graphs

Contacts happen due to the mobility of the people carrying the devices and reflect the complex structure in people's movements: meeting *strangers* by chance, *colleagues, friends and family* by intention or *familiar strangers* because of similarity in their mobility patterns. Our goal is to represent the complex resulting pattern of who meets whom, how often and for how long, in a compact and tractable way. This allows us to quantify structural properties beyond pairwise statistics such as inter-contact and contact time distributions.

To represent the structure of a mobility scenario, we aggregate the entire sequence of contacts of a trace to a static, *weighted contact graph* $G(N, \mathbf{W})$ with weight matrix $\mathbf{W} = \{w_{ij}\}$. Each device (or rather person carrying a device) is a node of this graph and a link weight w_{ij} represents the strength of the relationship between nodes i and j.

A key question is how to derive the tie strength between two nodes, i.e., what metric to use for w_{ij}, based on the observed contacts. This weight should represent the amount of mobility correlation (in space and time) between two nodes. Various metrics, such as the age of last contact [16], contact frequency [12] or aggregate contact duration [12] have been used as tie strength indicators in DTN routing.

[1]http://gowalla.com

[2]With the tie strength described in Sec. II-B the chosen duration does not affect the contact graph if all contacts have the same duration.

	DART	ETH	GOW	MIT
# People and context	1044 campus	285 campus	473 Texas	92 campus
Period	17 weeks	15 weeks	6 months	3 months
Type	AP associations	AP associations	Self-reported location	Bluetooth scanning
# Contacts total	4′200′000	99′000	19′000	81′961
# Contacts per dev.	4′000	350	40	890

TABLE I: Mobility traces characteristics.

In our study, we consider *both*, contact frequency and aggregate contact duration. They capture different aspects, both of which are important for opportunistic networking (e.g., for data dissemination). Frequent contacts imply many meetings and hence many forwarding opportunities (short delays) and long contacts imply meetings where a large amount of data can be transferred (high throughput)[3].

Since most network analysis metrics require one-dimensional tie strengths, we map these two features to a scalar weight. We first assign each pair of nodes $\{i, j\}$ a two-dimensional feature vector, $\mathbf{z_{ij}} = \left(\frac{f_{ij}-\bar{f}}{\sigma_f}, \frac{t_{ij}-\bar{t}}{\sigma_t}\right)$, where f_{ij} is the number of contacts in the trace between nodes i and j, and t_{ij} is the sum of the durations of all contacts between the two nodes. $\bar{f}$ and $\bar{t}$ are the respective empirical means, and σ_f and σ_t, the empirical standard deviations. We normalize the values by their standard deviations to make the scales of the two metrics comparable. We then transform the two-dimensional feature vector to a scalar feature value, using the *principal component*, i.e., the direction in which the feature vectors of all node pairs $\mathbf{Z} = \{\mathbf{z_{ij}}\}$, $i, j \in N$ has the largest variance. This is the direction of the eigenvector $\mathbf{v_1}$ (with the largest corresponding eigenvalue) of the 2×2 covariance matrix of frequency and duration. We then define the tie strength between i and j as the projection of $\mathbf{z_{ij}}$ on the principal component $w_{ij} = \mathbf{v_1}^T \mathbf{z_{ij}} + w_0$, where we add $w_0 = \mathbf{v_1}^T \left(-\frac{\bar{f}}{\sigma_f}, -\frac{\bar{t}}{\sigma_t}\right)$ (the projection of the feature value for a pair without contacts) in order to have positive tie strengths. The obtained weight is a generic metric that combines the frequency and duration in a scalar value and captures the heterogeneity of node pairs with respect to frequency and duration of contacts[4].

III. Small-world Structure

Fig. 1 shows as an example the DART contact graph. We observe that there is strong non-random structure. To quantify this structure, we first measure some standard metrics such as average shortest path lengths and clustering coefficients.

The graphs of all scenarios are connected, i.e., there is a path between all node pairs. However, they vary in terms of density (the density D being the percentage of node pairs for which $w_{ij} > 0$):

$D_{\text{DART}} = 0.12$, $D_{\text{ETH}} = 0.09$, $D_{\text{GOW}} = 0.04$, $D_{\text{MIT}} = 0.68$.

We first compute average path lengths and clustering coefficients. With these properties we can examine the graphs for small-world characteristics. Small-world networks, according to [17], manifest short paths between nodes (a typical property of random, Erdős-Rényi graphs) *and* high clustering coefficient (tendency of relations to be transitive). The clustering coefficient of node i is defined as (e.g., [17])

$$C_i = \frac{\text{number of triangles connected to } i}{\text{number of triples connected to } i}.$$

It ranges from 0 to 1, indicating the percentage of triangles which are "closed". The clustering coefficient of a graph is the average of the nodes' clustering coefficients. The average path length is the shortest path length, averaged over all connected node pairs.

	Clustering Coefficient				Avg. Path Length			
	1%	**2%**	**3%**	**4%**	**1%**	**2%**	**3%**	**4%**
DART	0.71	0.63	0.57	0.54	7.4	3.7	2.9	2.6
ETH	-	0.66	0.57	0.53	-	6.1	5.6	4.0
GOW	0.28	0.27	0.27	0.26	4.5	3.4	3.0	2.8
MIT	-	-	0.56	0.57	-	-	4.6	3.8

TABLE II: Clustering Coefficients and Average Path Lengths using different graph densities. Missing values in cases where there is no giant component.

While clustering coefficient and average path length provide meaningful and comprehensible information about the structure of a *binary* graph, their respective generalizations (e.g., [18]) to *weighted* graphs are much less easily interpretable. To maintain the interpretability, we dichotomize our contact graphs by using a threshold to extract the strongest ties (i.e., set them to one and the rest to zero). By doing so, we extract the regular and predictable "backbone" of the contact graph, and dismiss the random unpredictable part (c.f. [19]). To ensure that our results are not distorted by the threshold, we show that the qualitative relations of path length and clustering coefficient stay the same with different thresholds.

Table II shows the average clustering coefficients for different weight thresholds, chosen such that the binary graph densities are fixed to 0.01, 0.02, 0.03 and 0.04. Note that for a random graph (Erdős-Rényi), the clustering coefficient increases linearly with density from 0 to 1. Thus, in a graph where 10% of the node pairs are connected, the expected clustering coefficient is 0.1. The values show that all scenarios are considerably more clustered, strongly suggesting non random connectivity. We observe that the clustering coefficient of DART, ETH and MIT are very high and strikingly similar, whereas the GOW trace is a bit less clustered. We attribute this to the different nature of the traces: Transitivity of ties in work and home environments is stronger than in social activities captured by Gowalla.

[3]Note that the age of last contact is not suitable for our purpose, since we need aggregate properties over the trace duration.

[4]Note that this framework implicitly assumes stationarity of the underlying mobility process, which is not always true in some traces. In practice (e.g., for protocol design), one can implement a sliding window mechanism.

Trace/Model	# Comm.	Q
DART	23	0.84
ETH	21	0.81
GOW	29	0.7
MIT	6	0.52

TABLE III: Number of communities and modularity (Q).

Looking at the average shortest path length, we see that paths are only few hops long on average. *Thus, we observe the small-world behavior, typical for social networks, also in the network of physical encounters.*

This finding is related to the report of short opportunistic paths by Chaintreau et. al. [14]. However, there are two main differences to this study which we discuss in the following.

i) [14] measures path length in terms of number of relays using epidemic dissemination (i.e., messages are copied at every encounter). Their result means that short paths exist, when accounting for *both*, strong and weak ties (pairs which meet often and ones that only meet randomly). Here, we limit the edges to *strong* ties, and find that paths are still short. This is an important distinction for designing dissemination protocols, where forwarding decisions must be made based on regular encounters and not random, unpredictable ones [19].

ii) The scenarios in [14] are one order of magnitude smaller in number of nodes (MIT, which we also consider, is the largest trace considered there). It is not obvious that the results in [14] would also hold for networks with more nodes (DART) and broader geographic range (GOW).

Note also, that [14] does not account for clustering.

IV. Community Structure

Additionally to the described small-world characteristics, we are interested in the existence of *communities* [20] in the contact graph. Communities, informally defined as subsets of nodes with stronger connections between them than towards other nodes, are typical for the structure of social networks[5].

The existence of strong communities in the contact graph has various implications for opportunistic networks: On one hand, it implies high potential for node cooperation and community-based trust mechanisms. On the other hand, it may also imply high convergence times for distributed algorithms since there may be strong bottlenecks between communities.

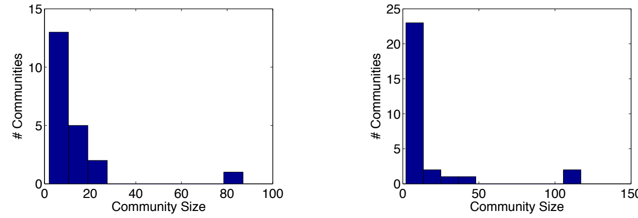
A. Community Detection Results

To detect communities in the contact graph, we apply the widely used Louvain community detection algorithm [22][6]. To measure how strongly modular the resulting partitioning is, we use Newman's Q *function* [20]:

$$Q = \frac{1}{2m} \sum_{ij} \left(w_{ij} - \frac{d_i d_j}{2m} \right) \delta(c_i, c_j),$$

[5]Note that the existence of community structure is related to a high clustering coefficient, however, strong clustering can have origins other than community structure. For instance in the ring lattice (where the nodes are arranged in a ring and connected to their K (with constant, small K, $K \geq 2$) neighbors on the left and right side), without re-wiring, the clustering coefficient is $\frac{3}{4}$ [21], even though there is no community structure.

[6]Since the partitioning of communities depends on the algorithm, we used *spectral clustering* as a second algorithm for detecting communities. The results are very similar, hence we do not report them here.



(a) ETH (b) GOW

Fig. 2: Community sizes.

where $d_i = \sum_j w_{ij}$ is the degree[7] of node i and $m = \frac{1}{2}\sum_j d_j$ is the total weight in the network. c_i denotes the community of node i thus, the Kronecker delta function $\delta(c_i, c_j)$ is one if nodes i and j share the community and zero otherwise. $Q = 0$ is the expected quality of a random community assignment. [20] reports modularities of above $Q = 0.3$ for different networks (social, biological, etc.).

The number of identified communities and the respective Q values are reported in Table III. We observe that in all cases, we get high modularity values between 0.52 and 0.84. Note that high modularity values of the graph of aggregate contact durations have also been reported in [12] (for smaller traces and different community detection algorithms), thus our findings are in agreement. Relating the modularities to clustering coefficients (see Table II), we observe on one hand that for DART and ETH, where we measure high clustering coefficients, also modularities are strong. On the other hand, MIT having similarly high clustering coefficient, has the lowest modularity, and GOW has a high modularity despite having the lowest clustering coefficient among all traces. This confirms that clustering coefficient and modularity measure different aspects of clusteredness: While the social nature of the GOW contacts makes them less transitive in terms of forming triangles, they are still grouped into larger communities.

To get an impression of the sizes of such communities, we show two examples of community size histograms in Fig. 2. In both, ETH and GOW we observe that the majority of communities are small (≤ 10 nodes), yet, there are few very large communities. The same observation holds for DART, which we do not show due to space limitations. MIT is, with only 6 communities, too small to draw conclusions on the community size distribution. These large communities, some of which have 100 and more nodes, raise the question about the community internal structure: Are communities homogenous entities within which all nodes have similarly strong connections to all other community members, or do they manifest more complex internal structure?

B. Intra-Community Structure

A community, by definition, is a strongly connected subgroup of nodes. Hence, we expect intra-community weights (where intra-community weights w_{ij} are those, for which $c_i = c_j$) to be stronger on average than the average of all

[7]This metric is sometimes also called *node strength*, to distinguish it from the degree (number of edges of a node) in binary graphs.

		DART	ETH	GOW	MIT
Density	Global	0.12	0.10	0.04	0.68
	Community	0.65	0.47	0.13	0.92
Median	Global	0.02	0.16	0.24	0.1
	Community	0.74	0.34	0.24	0.94
Coeff. of Variation	Global	3.4	2.7	7.2	2.5
	Community	1.8	2.2	5.9	1.4
Skewness	Global	8.7	5.8	33	7.3
	Community	5.1	4.5	21	4.5
Kurtosis	Global	188	54	1431	105
	Community	72	34	607	44

TABLE IV: Weight statistics, comparing weight distributions of *intra-community* weights to distributions of *all weights* in the network.

		DART	ETH	GOW	MIT
Median	Global	0.10	0.09	0.02	0.41
	Community	1.7	0.7	0.08	1.8
Coeff. of Variation	Global	0.9	1.2	1.9	0.7
	Community	0.9	1.5	3.5	0.7
Skewness	Global	2.2	2.1	5.0	1.0
	Community	2.6	2.3	6.3	1.4
Kurtosis	Global	11.5	9.0	37	4.2
	Community	15.1	9.1	52	6.4

TABLE V: Statistics for all normalized global and intra-community degrees.

weights. Indeed, Table IV reports that the median[8] of the weights is much stronger *within* communities than globally. However, we also notice that communities are not completely meshed entities: The density (we define the community density as the percentage of intra-community weights > 0) is far from 1. This suggests high heterogeneity of weights within communities, which is confirmed by the complementary cumulative distribution function (CCDF) plots of the global and intra-community weights in Fig. 3.

We also observe that the distributions differ between the traces. The straight line in log-log scale implies a power law distribution of the GOW weights. For the other traces, the plots suggest a somewhat thinner tail with two regimes or log-normal shape. Yet, note that the distribution of the *global weights* is in all cases qualitatively very similar to the distribution of the *intra-community weights*[9]. This is an important observation, as it suggests that there is no fundamental difference between community weights and global weights, other than intra-community weights being stronger on average.

To confirm this visual conclusion, and to characterize the distributions further, we report statistics related to their second (Coefficient of Variation), third (Skewness) and fourth (Kurtosis) moments [23] in Table IV. The high Coefficients of Variation (> 1 means more variation than an exponential distribution) confirm high heterogeneity of weights, both within communities and globally. Further, we notice high Skewness values (2 for exponential, higher positive values imply higher asymmetry towards the right of the mean) and high Kurtosis values (9 for exponential, higher values imply flatter distributions) implying a “fat” tail of weight distributions, both globally and within communities.

Even if a node does not meet all of its community peers, there must still be a reason that it is placed in a certain community: We expect it to have a high average weight towards other nodes in its communities. To verify this, we define the *normalized global degree* of node i as its node degree $d_i = \sum_{j \in N} w_{ij}$, divided by the number of nodes in the network $|N|$. Similarly, we define the *normalized community degree* of node i as the sum of its weights to other nodes of its community $d_i^{c_i} = \sum_{c_i = c_j} w_{ij}$, divided by the number of nodes in community c_i.

Table V indeed shows that the median is significantly higher for community degrees. Fig. 4 further plots the CCDFs of the normalized degrees in log-linear scale (notice the difference in scale from Fig 3). From the almost straight lines (particularly in DART and ETH), we visually conclude that the distributions are close to exponential. This is confirmed with the Coefficients of Variation, Skewness and Kurtosis values reported in Table V, which (except for GOW) are close to the values for exponential distributions. Note also that the statistics for global and community degrees follow each other closely, suggesting that the distributions are of the same type. We will discuss the implications of these findings in the following section.

V. Discussion and Conclusion

We have presented an extensive measurement study of structural properties of contact traces collected in different mobility scenarios. In the following, we summarize our findings and discuss the implications for *mobility modeling* and for *routing protocols* for opportunistic networks.

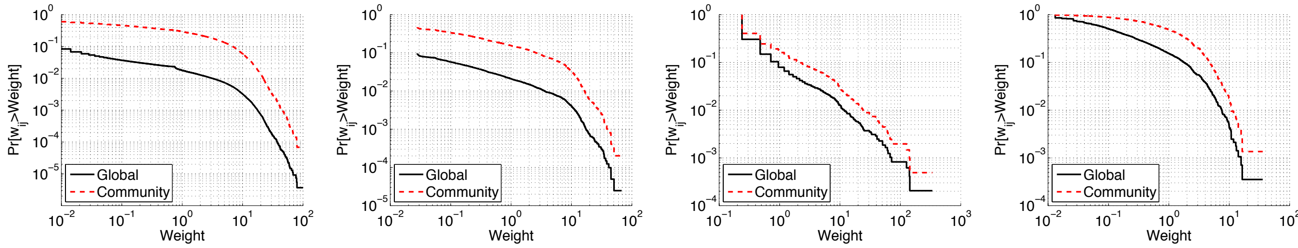
Small-world: Short average path lengths and high clustering coefficients show that the graph of wireless contacts has small-world structure. Unlike [14], which only proves the *existence* of short space-time paths (possibly formed by random, *unpredictable* contacts), our findings suggest that short paths can in fact be *found* by complex network analysis based routing protocols [12], [13] that can only infer and exploit strong mobility ties and *predictable* contacts. This result suggests that the contact graph has good *navigability* properties [24], an aspect we plan to analyze in the future.

Heterogenous weights: Weight distributions show high variance and heavy tails, both, globally and within communities. While this result for global weights is related to heavy tailed inter-contact times reported before [9], [10], it is more surprising for the intra-community weights. Such heterogeneity within communities should be considered when modeling social (group) mobility. Further, it has implications for opportunistic routing protocols: It is not enough to find the community of the message destination. Intra-community routing could help to find a node within a community.

Degree distributions: Heterogeneity of degrees is much smaller than heterogeneity of weights and the degree distributions are clearly not scale-free. This raises some questions as

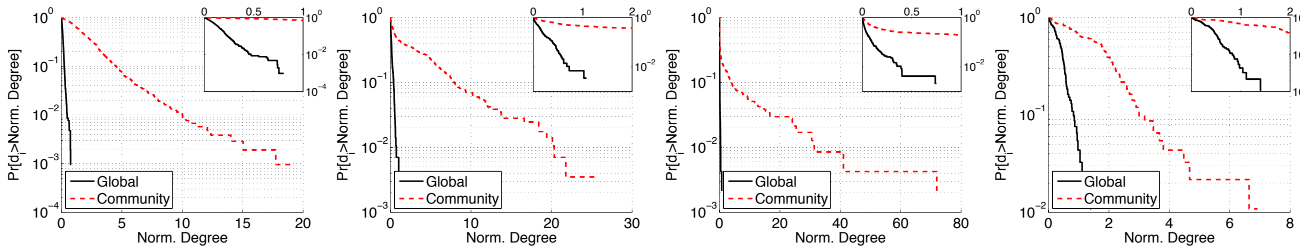
[8]Because of the heavy tail of weight distributions (see below), we use the robust median as an average.

[9]Notice that the intra-community weights are subsets of the global weights, containing 29% (DART), 59% (ETH), 42% (GOW) and 26% (MIT) of the weight values. However, because these subsets are not picked randomly but such that the weights must fall within a community, it is not obvious that the subsets should have the same distribution as the total sets of weights.



(a) DART (b) ETH (c) GOW (d) MIT

Fig. 3: Weight CCDFs, global and intra-community weights in log-log scale.



(a) DART (b) ETH (c) GOW (d) MIT

Fig. 4: CCDFs for normalize global and community degree in log-linear scale. Insets show the same values with different x-axis range.

to the efficiency of intra- and inter-community routing schemes based on increasing degree centrality of the relay (as a typical search strategy for networks with scale-free degrees). In future work, we intend to quantitatively investigate, whether such approaches can efficiently navigate the contact graph.

Similarity of global and community scale: The similarity of the weight and degree distributions, globally and within communities, suggests that similar routing strategies could be applied on both levels. This finding also hints towards a *self-similar* structure of the contact graph across different scales, a property which has already been reported for different complex networks [25]. We plan to further investigate the similarity of global and intra-community weight and degree distributions.

Further, we plan to compare the characteristics we found here to those of contact graphs created from synthetic mobility models. Since opportunistic networks research is heavily based on simulation, it is important to have models that accurately reproduce realistic contact structure.

REFERENCES

[1] The Aka Aki Network. [Online]. Available: http://www.aka-aki.com/
[2] Delay tolerant networking research group. [Online]. Available: http://www.dtnrg.org
[3] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *ACM MobiCom*, 2004.
[4] C. Tuduce and T. Gross, "A mobility model based on WLAN traces and its validation," in *Inocom*, 2005.
[5] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *WDTN*, 2005.
[6] N. Eagle, A. Pentland, and D. Lazer, "Inferring Social Network Structure using Mobile Phone Data," *PNAS*, 2009.
[7] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, June 2008.
[8] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, 2010.
[9] 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 *IEEE Infocom*, 2006.
[10] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnovic, "Power law and exponential decay of inter contact times between mobile devices," in *ACM MobiCom*, 2007.
[11] M. Musolesi and C. Mascolo, "A community based mobility model for ad hoc network research," in *ACM REALMAN*, 2006.
[12] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble Rap: Social-based forwarding in delay tolerant networks," in *ACM MobiHoc*, 2008.
[13] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant MANETs," in *ACM MobiHoc*, 2007.
[14] A. Chaintreau, A. Mtibaa, L. Massoulie, and C. Diot, "The diameter of opportunistic mobile networks," in *CoNEXT 07*, 2007.
[15] A. Scherrer, P. Borgnat, E. Fleury, J. L. Guillaume, and C. Robardet, "Description and simulation of dynamic mobility networks," *Elsevier Computer Networks*, 2008.
[16] H. Dubois-Ferriere, M. Grossglauser, and M. Vetterli, "Age matters: efficient route discovery in mobile ad hoc networks using encounter ages," in *ACM MobiHoc*, 2003.
[17] M. E. J. Newman, "The structure and function of complex networks," March 2003.
[18] J. Saramäki, M. Kivelä, J. P. Onnela, K. Kaski, and J. Kertész, "Generalizations of the clustering coefficient to weighted complex networks," *Physical Review E*, 2007.
[19] T. Hossmann, T. Spyropoulos, and F. Legendre, "Know thy neighbor: Towards optimal mapping of contacts to social graphs for DTN routing," in *Infocom*, 2010.
[20] M. E. J. Newman, "Modularity and community structure in networks," *PNAS*, 2006.
[21] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks." *Nature*, 1998.
[22] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J.STAT.MECH.*, 2008.
[23] W. Feller, *An Introduction to Probability Theory and Its Applications*. Wiley, 1968.
[24] J. M. Kleinberg, "Navigation in a small world," *Nature*, 2000.
[25] C. Song, S. Havlin, and H. A. Makse, "Self-similarity of complex networks," *Nature*, 2005.