SniffMob: Inferring Human Contact Patterns using Wireless Devices

Eric Anderson, Caleb Phillips, Harold Gonzales,
Kevin Bauer, Douglas Sicker, and Dirk Grunwald
Computer Science Department
University of Colorado, Boulder
[first.last@colorado.edu]

ABSTRACT

The size of existing data sets regarding human mobility and person-to-person contact has been limited by the labor-intensive nature of the data collection techniques employed. In this paper, we propose a practical data collection system which is automatic and transparent to the user, requires only installing new software, and uses the multiple sensing capabilities provided by current commodity mobile devices. This approach allows the scale and duration of these human contact studies to increase by several orders of magnitude and allows for the collection of location and contact information about individuals who do not install our data collection software. We present an analysis of the expected coverage of our data collection system drawing from existing data sets and random graph theory. To illustrate the type of application enabled by the availability of human contact data, we present "personalized epidemiology," a novel application that provides its users with information about their exposure to illness and offers advice on how to remain healthy.

1. INTRODUCTION

Large-scale data regarding human mobility and person-to-person contact has the potential to enable new frontiers of research in fields such as epidemiology and behavioral research as well as guide new mobile technologies and applications. We propose to use electronically-collected data to approximate the complete contact graph. Participants can run software on their existing mobile devices to detect their patterns of movement and contact with other individuals. Because participants' devices can infer location and contact information about non-participants as well, the size of the population studied can be dramatically larger than the number of participants.

Our principle goal is to suggest a system that is capable of extracting the contact patterns of a very large population using a relatively small number of sensors. To achieve this scalability, these sensors are deployed organically on a volunteer basis and use a combination of existing technologies to infer contact information. To justify our approach, we provide an *a priori* analysis of the ability of a subset of realistically connected observers to document the connectivity of the population as a whole. We also discuss the limitations of this framework, not least of which are the inherent privacy implications, which we do not attempt to resolve here.

As an example of the kind of application enabled by this type of rich contact data, we propose *personalized epidemiology*. By combining traditional disease modeling techniques with knowledge of *individuals'* contact patterns and *real-time* information, it may be possible to identify risks and propose interventions appropriate to each person at any moment.

2. RELATED WORK

Although many researchers have considered modeling mobility and contact patterns, few data sets exist that are sufficient in their resolution and size. The simplest models are purely theoretical. Placing way-points at random, or as determined by Brownian motion, is a common naïve approach which has been shown to poorly predict reality [31]. In [25], the authors study human mobility in tens of kilometer wide areas (*i.e.*, college campuses, amusement parks, etc.). They find human patterns to be statistically similar to Levy walks, a type of motion studied in particle physics.

There have been several attempts to gather empirical data. Many data sets used in the social sciences provide complete records of participants' contacts with anyone; these have been produced by hand and therefore cover a small number of people for a short period. Good introductions to these methods and specific data sets are provided in [11, 20]. Automatic but coarse mobility traces have been produced using a relatively small number of fixed sensors to track a large number of mobile users. The data collected at the access points (APs) on the Dartmouth campus [14] is probably the best example, although the work of Balazinska [1] on modeling a corporate wireless network and that of researchers at UCSD [17] is also worth mentioning. Because these data sets can only describe mobility on a fixed way-point based resolution, some more recent work has gone into using mobile sensors to track mobility and contact on small scales, such as [9] and the PDA and iMote experiments in [6]. These are generally limited to analyzing contact *among participants*. Other studies have inferred proximity from information such as student classroom schedules [28], mobile phone call records [23] or workplace computer log-ins [8].

Our proposal primarily differs from this prior corpus on the basis of scope. Whereas previous projects have focused on small domains, used specialized hardware, or required

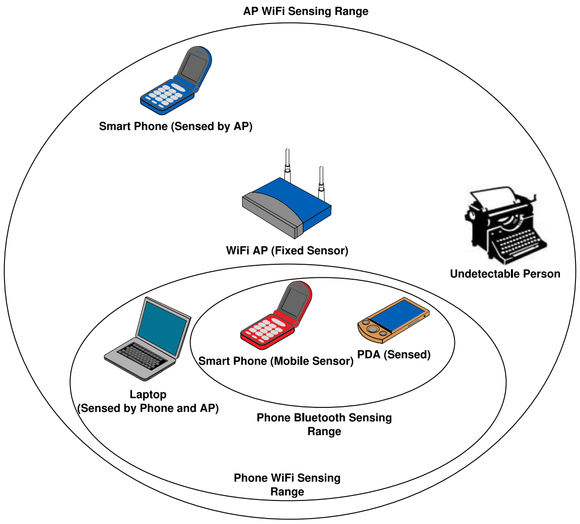

Figure 1: System example

total participation, our proposal is able to scale to any size population with any proportion of participant sensors. We accomplish this by using a simple design which is based on already ubiquitous and inexpensive technology whose power we demonstrate with a novel epidemiology application.

3. TOWARD A GLOBAL DATA SET

Contemporary mobile electronic devices include mobile telephones, laptop computers, music players, digital cameras, PDAs, and many more. Such devices are nearly ubiquitous in wealthy countries, and are only increasing in their global reach, number, and capabilities. Because of their ubiquity, there has been significant interest in using such devices as sensing platforms.

3.1 System Concept

We propose to take advantage of two capabilities that are both wide-spread and mature in wireless consumer devices: the ability to communicate with – and therefore also detect and uniquely identify – other devices, and the ability to determine their own location. By regularly recording their location and the other devices' identifiers in the area, participants' devices can produce the raw traces from which mobility and contact models can be built.

Sensor Types – We envision a system involving a mix of mobile sensors, such as smart phones, and fixed sensors such as access points and wireless sniffers. The use of fixed sensors is motivated by the observation that there are common meeting and aggregation points for individuals (*i.e.*, coffee shops, cafeterias, classrooms, etc.) [12]. An obvious candidate for this task is wireless APs, which are already deployed with great density and have similar capabilities to embedded platforms used in mobile devices with the additional benefit of having a reliable power supply and a stable location. In our concept, individuals would volunteer to run sensing software on their APs or mobile devices. This software would then collect traces and transmit them to an aggregation point where they would be combined, filtered, and stored.

Unique Identifiers – Most devices can be uniquely identified with a fixed global identifier advertised by the wireless communication protocol. For instance, both Bluetooth and IEEE 802.11 transmit their hardware (MAC) address in every frame. Even in the absence of explicit identifiers, radio frequency (RF) and traffic characteristics may uniquely identify devices in some cases [3].

Proximity Granularity – Every wireless communication technology, from the infrared signals used by remote controls to the various radio systems of Bluetooth, WiFi, and cellular telephony, has its own detection capabilities and limitations. There is a significant body of work on using such detection capabilities for other purposes or on smaller scales (*e.g.*, [9, 15, 19]). In general, the longer-range technologies provide the greatest radius of *coverage*, but with the least *resolution* of location. For example, if an 802.11 device is able to detect two devices at the same time, they are likely within 100 meters of each other. In the case of Bluetooth, it is more likely 10[1]. By using many technologies, it is possible to build a multi-resolution data set, in which the researchers interested in gathering contact data can trade-off between precision and completeness.

Inferring Contacts – Human contact (i.e, proximity in space and time) must be inferred from the measured data. This is simple for direct detections: if participating device a senses neighboring device b using communication technology t, then we can say with fair certainty that a and b were within the sensing radius of t of each other, and can therefore connect a and b on the graph at that point in time. The observed signal strength can give an imperfect further hint about their distance. A more difficult scenario arises when a participant detects several devices: it can be inferred that each device was within a known range of the sensor, but they could be adjacent to or opposite each other. The possible separation between nodes in such inferred contacts is up to twice that of directly-measured contacts. Unless the measurements are simultaneous, there is also temporal uncertainty. This is especially a concern with fixed sensors, which always must infer connections between the nodes it senses. *Contact is not boolean: it includes physical and temporal proximity.*

Sensing Example – Consider Figure 1. In this example, there are two sensors, one mobile - the phone pictured at center, and one fixed - the wireless access point. Here, the mobile sensor is able to detect the PDA using its Bluetooth sensor, and the laptop using its 802.11 sensor. The access point may have greater sensitivity and be able to sense an additional mobile device that is outside the reach of the mobile sensor. The typewriter is undetectable since it does not produce a directly-measurable RF identifier. In this scenario, contacts between the non-sensor devices, such as the PDA and laptop, must be inferred. Because the inferred PDA-laptop contact is noticed by two sensors, we can assign it a higher confidence, and using signal strengths and possibly additional sensors, we can start to use trilateration to pinpoint the devices to a higher degree of confidence.

Localization – Although it is not essential for generation of strict "contact" graphs, localization provides a means of mapping the contacts observed into the physical environment. This creates greater opportunity for connection inference as well as tracking movement. For fixed devices, the lo-

---

[1] Communication range depends on the antennas, power levels, and modulation schemes in use, as well as many environmental factors. Some of these may be known to the system, or reasonably estimated, but significant uncertainty is unavoidable.

calization problem is trivial - it can be computed accurately once and stored. For mobile devices there are a few localization technologies that are currently available and in general use. The Global Positioning System (GPS) allows for fine grain localization in outdoor environments, but is ineffective in indoor environments, which are of interest for many purposes. Global System for Mobile communications (GSM) localization is based on multilateration[2] and provides coarse grain localization. 802.11 based localization services such as SkyHook [27] claim resolutions of 10 to 20 meters, but suffer from their data collection methods relying on manually collected and contributed AP observations.

Sensor Fusion – Producing a global view of the location and contact record entails combining qualitatively different types of data, with differing confidence levels, acquired from many devices. Some data, such as GPS location, have well-defined confidence intervals. Others, such as signal strength ranging, require subjective judgment by the system designers. It consequently makes sense to reason in terms of degrees of belief as in Dempster-Shafer theory [26]. Suppose, for example, that participating devices a and b estimate themselves to be in the same place, and they detect non-participants c and d respectively, but neither detects both. Our belief that c and d are within two detection radii of each other depends on our degree of confidence in the following statements: 1) a and b are in the same place, 2) c is near a 3) d is near b, 4) d is not near a, and 5) c is not near b. If a and b have GPS devices and are sensing with 802.11 devices on different channels, we have a strong belief in 1, and very weak belief in 4 and 5, so we end up with a reasonable belief that c and d are close.

3.2 Challenges

Simplicity usually comes with a cost, and our proposal is no different. There are a few significant limitations, assumptions, and concerns which must be considered. By its very nature, using a mobile device as a unique identifier for a person has some issues.

Device ↔ Person Mapping – So far, we have spoken as though detecting a device was equivalent to detecting a person. It is important to handle several cases: a person with multiple devices, a device with no person (infrastructure), a device which is passed between multiple people, and a person without a device. It is not likely that any of these can be handled completely, but good-enough filtering via heuristics and optimization strategies may be possible. For instance, infrastructure devices can be differentiated from people, because people are observed to move and therefore generate contacts in different locations whereas infrastructure presumably never moves. Individuals with multiple devices will often move their devices in concert. Hence, a group of closely correlated devices (in space and/or the connectivity graph) are most likely a single individual. Users without devices pose a difficult problem, which cannot be addressed directly.

Selection Bias – There are two issues here. First of all, the sort of electronic devices that we can detect are not uniformly distributed throughout the population. Telephones may be sufficiently universal, but other equipment is probably correlated with wealth and other influences. Second, within the group of electronics-bearing people, participants in a large-scale data gathering system are likely to be self-selecting. There is no reason to believe that this group will be representative. While we cannot remove this bias from the data, we can hope to account for it by comparing our results with separately compiled demographic information on technology usage [24].

Privacy – There are obvious privacy issues which we will not pretend to resolve here. Devices' unique identifiers can be obfuscated to prevent directly linking measurements to specific people, but the risk remains that sensitive information could be inferred from context. The information recorded would be minimal relative to other network monitoring, but the proposed scope increases the importance of what is revealed. We leave this debate to those more qualified to discuss the legal ramifications of the system we propose.

3.3 Completeness

One goal of the system we propose is to provide information about not only the participants' contact patterns, but those of the non-participants they encounter. The quality of that information will depend on the *fraction of the population participating* (f), the *number of other people* within each participant's area of observation at any given moment, and the statistical distribution of their contacts. These cannot be known in advance, but we propose some plausible *a priori* models, and tests for *ex post facto* evaluation of a data set.

3.3.1 Analytical Projections

One reasonable measure of completeness is the population coverage, that is to say the fraction of the population included in the data in any given time window. If we model device contact as a graph, $G = (V, E)$, every device is a vertex, and there exists an edge between two vertices if they can detect each other. The *devices* covered are then the participating devices plus their immediate neighbors in the graph. Without knowing the actual contact patterns for the proposed system, we are limited to making educated guesses based on prior research. Analytically, the expected coverage as the *total* number of nodes n increases is given by Equation (1), where f' is the fraction of nodes *not* participating and $P(x)$ is the probability of a node having degree x:

$$\lim_{n\to\infty} \text{coverage} = 1 - \left(f' \cdot \int f'^{x} P(x)\, dx \right) \tag{1}$$

Exact results depend on the degree sequence $P(x)$, but different probability sequences with similar *expected mean degrees* ($\hat{d}$) behave similarly for our purpose, especially when the set of participating nodes is sparse. Existing work supports the idea that social graphs can be modeled reasonably well as random graphs with particular degree distributions, especially power-law distributions [2, 22, 10]. Similar results have been found for device-level contacts [5, 16]. For the simple case of a regular random network, Figure 2 shows the expected coverage relative to the fraction of nodes participating and the mean degree. Coverage is bounded from above by f plus the product $f \cdot \hat{d}$, which represents the case that all of the participating nodes' connections are with non-participants. That bound is relatively tight for small values of f. Note, for instance, that for 1% of the population participating and degree 100, the actual coverage is 64%. It is important to note that device detections, and therefore the edge set, degree distribution, and all derived graph proper-

[2] Determination of location by computing the time difference of arrival (TDOA) between signals from multiple cell towers.

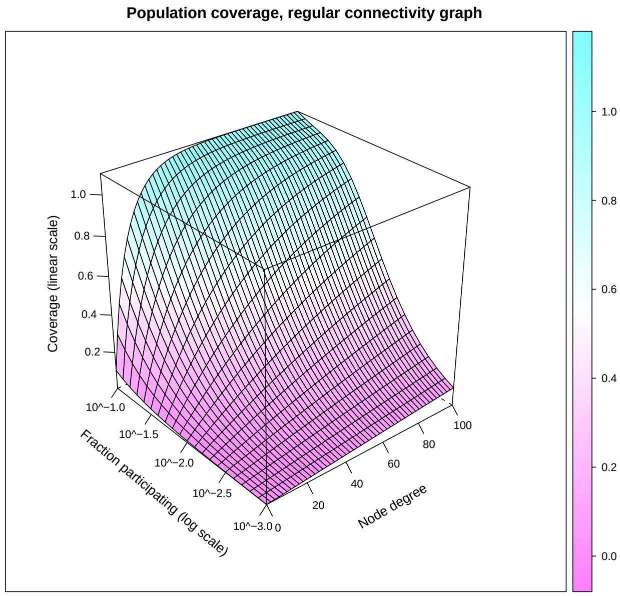

Figure 2: Population coverage relative to the fraction participating and the number of individuals each node can see; uniform regular contact graph.

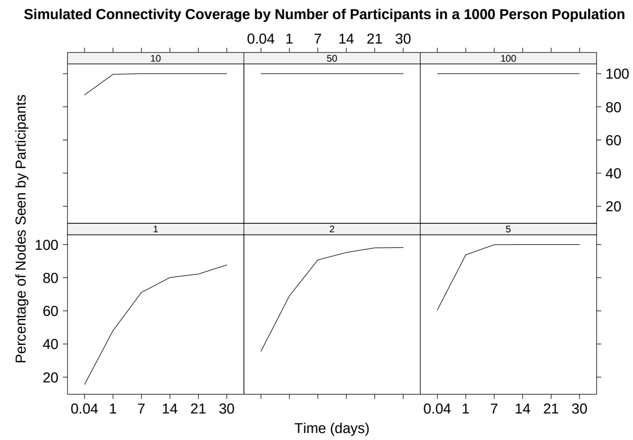

Figure 3: Simulated population coverage (percentage of people seen by participants) for a 1000 user population and $k = 0.35$ in the P_c calculation as a function of the number of participants and time period.

ties, are a function of time. The effects of time span are discussed further in the next section.

3.3.2 Naïve Empirical Simulation

Using random graphs along with empirical estimates of inter-contact times (ICT) from the literature, we can take this a step further by simulating the expected average coverage of a population as a function of the time-period of measurement and the proportion of the population participating. In [25], Rhee *et al.* tie together mobility and contact results from different scales and claim that they all are well modeled by a power-law distribution. Indeed, in [6], UCSD researchers find similar slope power-law fits to the inter-contact times (ICT) in 6 trace sets from various measurement studies. Most of the data they study fits with $k = 0.3$ to 0.4 in the tail, and a few deviant traces have $k = 0.9$. Using these empirical estimates of ICT distributions, we can simulate the contact dynamics using a random graph and study the way coverage converges as a function of the number of participants. The ICT distribution, which is modeled as $P_c(t)$, the probability of two randomly chosen nodes p and q contacting within time t:

$$P_c(t) = at^k \tag{2}$$

Where a is a scaling factor for the distribution:

$$a = \frac{1}{t_{max}^k} \tag{3}$$

We can then trivially define the mean degree $\hat{d}$ over some time t as follows, where n is the population size:

$$\hat{d}(t) = (n-1) * P_c(t) \tag{4}$$

A Monte Carlo style simulation follows easily. For a population of size n, there are $n(n-1)$ possible unique connections, each of which is connected with probability $P_c(t)$. By placing the fraction of participants f at random in the population, we can study how well they are connected as t and f vary. Because the mean degree is not directly a function of the population size, this model is independent of the population size. Figure 3 plots these results for a representative population of 1000. We can see how the percentage of the population seen by the participants grows as a function of time and saturates for higher proportions of participants. For instance, with 10 participants out of 1,000 (top-left in figure 3) we can see that the coverage converges to 100% after only one day. Although this is clearly an upper bound since it does not consider asymmetries in the graph and tries to model the complexities of human movement with a straight-line fit through data, it is certainly motivating as a first estimate.

3.3.3 Evaluating an Existing Data Set

The previous sections were concerned with estimating the quality of data which might be obtained, a fundamentally perilous task. Once the data is in hand, however, there are a number of reasonable tests one may apply to either the whole set or selected subsets. Most of the proposed tests are based on the assumption that participants are representative of the population as a whole. With that assumption, for a characteristic which can be directly measured for the participant set and inferred for non-participants, differences between the two sets suggest data completeness problems.

Population Coverage – The number of devices seen can be compared with an estimated true population size. For cellular telephones, "ground truth" is theoretically available: carriers already know the state and very rough location of every active phone. For other devices, an estimate must be based on human population, market penetration, estimated duty cycles, and other approximations as in surveys such as those conducted by [24].

Temporal Coverage Quality – For those devices seen at all, an obvious question is "what fraction of the time were they being observed?" This can be answered directly from the data, but to be useful it needs to be normalized against how often a *perfect* observer would see them. Suppose that an average phone is active and near anyone x hours per day. The quality of temporal coverage should be evaluated relative to the perfect case of seeing non-participants $x/24$ of the time, not all the time.

Connection Coverage – Observing contact between non-participants is important to the overall quality of the data. To infer such contact, both parties must be detected. If detection and non-participant contact are perfectly independent, one would expect the square of the population coverage. If meeting each other is correlated with also meeting a participant (that is, being detected) – which is not unreasonable if both types of meeting occur in popular places [12] – the connection coverage would approach the population coverage. The actual quality can be estimated by comparing the number of connections known to occur on participating nodes with the number observed on non-participating nodes.

Relative Entropy – An alternative to estimating the quality of the data set is to estimate the quality of the observers. For some measure of interest, the *marginal* information in any observation can be regarded as the Kullback-Leibler distance between the distributions of the measure taken over the data sets with and without that observation [7]. If the relative entropy from the addition of participants converges to zero, that strongly suggests that the distribution of the measure has converged.

4. PERSONALIZED EPIDEMIOLOGY

One application that could be deployed using the sensing model we describe is a system to give users information about their potential exposure to disease. Traditional (Kermack-McKendrick) infection models assume homogeneous contact between individuals. Better information about contact patterns allows more accurate modeling [21]. For example, it is believed that the 2002-2003 SARS epidemic started among groups with atypically high connectivity (poor families living in close proximity). Based on the initial spread, researchers estimated the disease to be far more infectious that it actually was and predicted an epidemic at least an order of magnitude worse than what occurred [4]. Structure-aware network models have been applied *retrospectively* to outbreaks in hospitals, where patients' and caregivers' contacts can be reasonably estimated [18].

By combining such techniques with knowledge of *individuals'* contact patterns and *real-time* information gathering, we propose an application to identify risks and propose interventions appropriate to each person at each moment. Using our application and sensing framework, it would be possible to estimate the risk that people whom an individual sees regularly are sick. This application provides not only an example of how the data sets collected using our proposed method could be used, but also a possible incentive for an individual to participate in a data gathering system. Additionally, because this application operates on local networks of individuals and doesn't necessarily need a complete graph to make predictions, it is robust to sparseness in the contact graph early in the deployment.

The proposed application consists of two parts: the mobile sensors that are collecting contact and health data about the users and a centralized processing service that sends health alerts to keep users informed of their risk level. The next section describes various methods for collecting health information and the subsequent section describes how the centralized service produces health alerts for users.

4.1 Obtaining Health Information

Direct Reporting – The simplest approach is to allow users to provide information about their own health. This could be accomplished by a mobile application that runs on devices participating in the contact monitoring system. The application would allow the user to report symptoms of common illnesses and save them as their current health state. The user-specified health state along with the sensor-based contact data will then be delivered to the centralized processing service.

Sensing – Additionally, it might be possible for the application to sense symptoms that the user is experiencing or has been exposed to and automatically record them as part of the user's health state. Some sensing capabilities of mobile devices, such as a mobile phone's microphone, may hold the potential for such automated detection. Ordinary microphones on low-power sensor nodes have been used to analyze gun fire and identify the source location, direction, and type of weapon involved [29]. One can imagine using existing microphones to identify coughing, for example, and possibly even categorize the type of cough in order to estimate probability of infection. Although this is an extremely noisy data source, from which it is difficult to make useful predictions, we are motivated by other successes in extracting health information from noisy data. For instance, several groups have recently reported success inferring aggregate health information from online behavior. Notably, very strong correlations were found between influenza-related searches and actual disease prevalence [13], and web searches were a *leading* indicator for the 2008 Canadian listeria outbreak [30].

4.2 Health Alerts

Health alerts would keep users of the system informed about their level of contact with potentially sick individuals and allow for a proactive approach to disease prevention. The specifics of such a processing and notification system to combat the spread of disease is beyond the scope of this work and would require the collaboration of medical professionals. Whatever the disease model(s) employed, the output would be fairly simple. Users would be given a qualitative assessment of their disease exposure risk at that time, and specific guidance if appropriate. It is our belief that appropriately timed situation-specific advice could lead to better compliance – and better outcomes – than "standing orders" to which people may become desensitized. Also, many prophylactic treatments do not make sense for the entire population because of side effects, cost, or limited availability. In such cases, contact information may help to identify the most appropriate candidates [8].

5. CONCLUSIONS

This paper describes a system for gathering a global-scale database of human mobility and contact information, and an example of a health application enabled by such data. This approach replaces a labor-intensive process with an automated process which can be continuous in time and ubiquitous in scope. We argue that this can be achieved by leveraging the capabilities of existing mobile devices, so that no effort is required from study participants beyond the initial installation of software. In a world with billions of mobile electronic devices, the participation of even a tiny fraction of users could produce a data set of unprecedented detail and breadth.

There are significant unresolved issues, both social and technical. Foremost among these are the privacy questions. Even so, we suggest that the scientific and practical potential of such data and applications more than justifies their continued exploration.

6. REFERENCES

[1] Balazinska, M., and Castro, P. Characterizing mobility and network usage in a corporate wireless local-area network. In *Proc. MobiSys '03* (New York, NY, USA, 2003), ACM, pp. 303–316.

[2] Barabási, A.-L., and Albert, R. Emergence of scaling in random networks. *Science 286* (October 1999), 509 – 512.

[3] Bauer, K., McCoy, D., Greenstein, B., Grunwald, D., and Sicker, D. Physical layer attacks on unlinkability in wireless LANs. In *Proceedings of the 9th Privacy Enhancing Technologies Symposium* (August 2009).

[4] Brauer, F. *Mathematical Epidemiology.* Springer-Verlag Berlin Heidelberg, 2008, ch. An Introduction to Networks in Epidemic Modeling, pp. 133 – 146.

[5] Chaintreau, A., Hui, P., Crowcroft, J., Diot, C., Gass, R., and Scott, J. Pocket switched networks: Real-world mobility and its consequences for opportunistic forwarding. Tech. Rep. UCAM-CL-TR-617, University of Cambridge Computer Laboratory, 15 JJ Thomson Avenue, Cambridge CB3D 0FD, United Kingdom, Feb. 2005.

[6] Chaintreau, A., Hui, P., Crowcroft, J., Diot, C., Gass, R., and Scott, J. Impact of human mobility on the design of opportunistic forwarding algorithms. In *In Proc. IEEE Infocom* (2006).

[7] Cover, T. M., and Thomas, J. A. *Elements of information theory.* Wiley-Interscience, New York, NY, USA, 1991.

[8] Curtis, D., Pemmaraju, S., Hlady, C., Fries, J., Herman, T., Segre, A., and Polgreen, P. Vaccination strategies for healthcare workers based on social networks. In *International Meeting on Emerging Diseases and Surveilance (IMED)* (Vienna, Feb 2009), International Society for Infectious Diseases. Poster.

[9] Eagle, N., and Pentland, A. Reality mining: Sensing complex social systems. *Journal of Personal and Ubiquitous Computing* (2005).

[10] Eubank, S., Guclu, H., Kumar, V. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z., and Wang, N. Modeling disease outbreaks in realistic urban social networks. *Nature 429* (May 2004), 180 – 184. Letters.

[11] Fu, Y.-C. Contact diaries: Building archives of actual and comprehensive personal networks. *Field Methods 19*, 2 (2007), 194 – 217.

[12] Ghosh, J., Philip, S., and Qiao, C. Sociological orbit aware location approximation and routing in manet. In *Broadband Networks, 2005. BroadNets 2005. 2nd International Conference on* (Oct. 2005), pp. 641–650 Vol. 1.

[13] Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. Detecting influenza epidemics using search engine query data. *Nature 457* (Feb 2009), 1012 – 1015. Correspondence.

[14] Kotz, D., and Essien, K. Analysis of a campus-wide wireless network. *Wirel. Netw. 11*, 1-2 (2005), 115–133.

[15] Lawrence, J., Payne, T. R., and Roure, D. D. Co-presence communities: Using pervasive computing to support weak social networks. In *WETICE '06* (2006), IEEE Press, pp. 149 – 156.

[16] Lindgren, A., Diot, C., and Scott, J. Impact of communication infrastructure of forwarding in pocket switched networks. In *Proc. SIGCOMM '06 Workshops* (Sept. 2006), ACM, pp. 261 – 268.

[17] McNett, M., and Voelker, G. M. Access and mobility of wireless PDA users. *SIGMOBILE Mob. Comput. Commun. Rev. 9*, 2 (2005), 40–55.

[18] Meyers, L. A., Newman, M. E., Martin, M., and Schrag, S. Applying network theory to epidmeics: Control measures for outbreaks of Mycoplasma Pneumoniae. *Emerging infectious diseases 9* (2003), 204 – 210.

[19] Miklas, A. G., Gollu, K. K., Chan, K. K., Saroiu, S., Gummadi, K. P., and de Lara, E. Exploiting social interactions in mobile systems. In *UbiComp 2007* (2007), J. K. et al., Ed., no. 4717 in LNCS, Springer-Verlag, pp. 409 – 428.

[20] Mikolajczyk, R. T., and Kretzschmar, M. Collecting social contact data in the context of disease transmission: Prospective and retrospective study designs. *Social Networks 30* (2007), 127 – 135.

[21] Newman, M. E. The spread of epidemic disease on networks. *Physical Review E 66*, 016128 (2002).

[22] Newman, M. E., Watts, D., and Strogatz, S. Random graph models of social networks. *Proc. National Academies of Science 99* (2002), 2566 – 2572.

[23] Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., de Menezes, A. M., Kaski, K., Barabási, A.-L., and Kertész, J. Analysis of a large-scale weighted network of one-to-one human communication. *New J. Phys. 9*, 6 (June 2007), 179+.

[24] Pew Internet & American life project. http://www.pewinternet.org/, May 2009.

[25] Rhee, I., Shin, M., Hong, S., Lee, K., and Chong, S. On the levy-walk nature of human mobility. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE* (April 2008), pp. 924–932.

[26] Shafer, G. *A Mathematical Theory of Evidence.* Princeton Univ. Press, 1976.

[27] Skyhook wireless. `http://www.skyhookwireless.com`.

[28] Srinivasan, V., Motani, M., and Ooi, W. T. Analysis and implications of student contact patterns derived from campus schedules. In *Proc. MobiCom '06* (New York, NY, USA, 2006), ACM, pp. 86–97.

[29] Volgyesi, P., Balogh, G., Nadas, A., Nash, C. B., and Ledeczi, A. Shooter localization and weapon classification with soldier-wearable networked sensors. In *Proc. MobiSys '07* (New York, NY, USA, 2007), ACM, pp. 113–126.

[30] Wilson, K., and Brownstein, J. Early detection of disease outbreaks using the Internet. *Canadian Medical Assn. Journal 180*, 8 (April 2009), 829 – 831.

[31] Yoon, J., Liu, M., and Noble, B. Random waypoint considered harmful. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE* (March-3 April 2003), vol. 2, pp. 1312–1321 vol.2.