

Uncovering Properties in Participatory Sensor Networks

Thiago H. Silva
Universidade Federal de
Minas Gerais - UFMG
Computer Science
Belo Horizonte, Brazil
thiaghos@dcc.ufmg.br

Pedro O. S. Vaz de Melo
Universidade Federal de
Minas Gerais - UFMG
Computer Science
Belo Horizonte, Brazil
olmo@dcc.ufmg.br

Jussara M. Almeida
Universidade Federal de
Minas Gerais - UFMG
Computer Science
Belo Horizonte, Brazil
jussara@dcc.ufmg.br

Antonio A. F. Loureiro
Universidade Federal de
Minas Gerais - UFMG
Computer Science
Belo Horizonte, Brazil
loureiro@dcc.ufmg.br

ABSTRACT

A fundamental step to achieve the Ubiquitous Computing vision is to sense the environment. The research in Wireless Sensor Networks has provided several tools, techniques and algorithms to solve the problem of sensing in limited size areas, such as forests or volcanoes. However, sensing large scale areas, such as large metropolises, countries, or even the entire planet, brings many challenges. For instance, consider the high cost associated with building and managing such large scale systems. Thus, sensing those areas becomes more feasible when people collaborate among themselves using their portable devices (e.g., sensor-enabled cell phones). Systems that enable the user participation with sensed data are named participatory sensing systems. This work analyzes a new type of network derived from this type of system. In this network, nodes are autonomous mobile entities and the sensing depends on whether they want to participate in the sensing process. Based on two datasets of participatory sensing systems, we show that this type of network has many advantages and fascinating opportunities, such as planetary scale sensing at small cost, but also has many challenges, such as the highly skewed spatial-temporal sensing frequency.

Categories and Subject Descriptors

J.4 [Computer Applications]: Social and Behavioral Sciences; C.2 [Computer-Communication Networks]: Distributed Systems; G.3 [Mathematics of Computing]: Probability and Statistics—*Statistical computing*

General Terms

Measurement

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HotPlanet'12, June 25, 2012, Low Wood Bay, Lake District, UK.
Copyright 2012 ACM 978-1-4503-1318-6/12/06 ...\$10.00.

Keywords

sensor networks, participatory sensing, characterization, ubiquitous computing

1. INTRODUCTION

The future world envisioned by Weiser, called Ubiquitous Computing (ubicom or pervasive computing), consider a computing environment in which each person is continually interacting with many wirelessly interconnected devices [16]. Weiser believed that the most powerful things are those that are effectively invisible in use. The essence of this vision is make everything easier to do, with fewer mental gymnastics [15, 17].

A fundamental step to achieve Weiser's vision [15] is to sense the environment. The research in Wireless Sensor Networks has provided several tools, techniques and algorithms to solve the problem of sensing in limited size areas, such as forests or volcanoes [18]. However, sensing large scale areas, such as large metropolises, countries, or even the entire planet, brings many challenges. For instance, consider the high cost associated with building and managing such large scale systems.

Sensing large areas becomes more feasible when people carrying their portable devices (e.g., smart phones) collect data and collaborate among themselves. Smart phones are taking center stage as the most widely adopted and ubiquitous computer [7]. It is also worth noting that smart phones are increasingly coming with a rich set of embedded sensors, such as GPS, accelerometer, microphone, camera, gyroscope and digital compass [8].

Systems that enable sensed data in this way are named participatory sensing systems [2]. We consider that the shared data is not limited to sensor readings passively generated by the device, but also includes proactively user observations. It is possible to find several examples of participatory sensing systems already deployed, such as Waze¹ and Weddar². Waze allows users to report real-time traffic conditions. For its part, Weddar allows users to share weather conditions in a particular location. Moreover, there

¹<http://www.waze.com>

²<http://www.weddar.com/>

are the city location tagging applications, such as Gowalla³ and Foursquare⁴. This kind of application allows users to share their actual location associated with a specific place (e.g., restaurant).

Based on participatory sensing systems, a new type of network is derived, namely Participatory Sensor Network (PSN). In this type of network, nodes are autonomous mobile entities (users) and the sensing activity depends on whether they want to participate in the sensing process. PSNs have particular properties that differ them from traditional Wireless Sensor Networks (WSNs). The objective of this work is to characterize and analyze these properties using two datasets of participatory sensing systems. We show that this type of network has many advantages and fascinating opportunities, such as planetary scale sensing at small cost, but also has many challenges, such as the highly skewed spatial-temporal sensing frequency.

The rest of the work is organized as follows. In Section 2 we present some related proposals. In Section 3, we further discuss participatory sensing systems. In Section 4, we present participatory sensor networks, including their particularities and advantages. In Section 5, based on two datasets of participatory sensing systems, we discuss in details the pros and cons of participatory sensor networks. Finally, in Section 6, we present some concluding remarks and discuss some future steps.

2. RELATED WORK

In the literature there are different studies dedicated to the participatory sensing. Several of them propose participatory sensing systems, including traffic monitoring [5, 6] and noise level monitoring [11]. In order to guarantee the success of participatory sensing systems, it is necessary to ensure that the participation is sustained over time. Thus, there are research groups dedicated to study incentive mechanisms [12] and the quality of the shared data [10].

There are also proposals dedicated to the study of social and spatial properties of data shared in location sharing services [3, 4, 13]. All of them aim to study user mobility patterns and their implications. For example, Cho et al. [4] were interested in answering where and how often users move, and how social ties interact with movement. Our work differs from the aforementioned studies since we are interested in analyzing PSN properties. In particular, we analyze a participatory sensing system as a sensor network.

3. PARTICIPATORY SENSING

Participatory sensing is the process where individuals use mobile devices and cloud services to share sensed data [2]. Usually participatory sensing systems consider that the shared data is generated automatically (passively) by sensor readings from the device, but in this work we consider also manually (proactively) user-generated observations. Sometimes, participatory sensing with this characteristic is called ubiquitous crowdsourcing [10]. Figure 1 shows an overview of the essential components of a participatory sensing application: sensing, processing, and application analysis.

The sensing component is the element that exhibits more particularities. Given the widespread adoption of sensor and

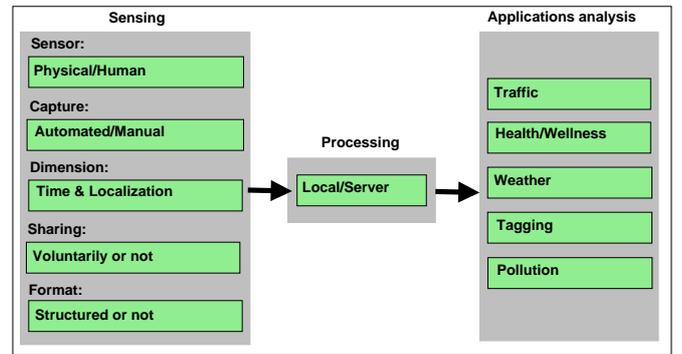


Figure 1: Overview of typical components of participatory sensing systems

Internet-enabled cellphones, these devices create an important tool for this component. They have become a powerful platform that encompasses sensing, computing and communication capabilities able to capture both manual (on-demand) and pre-programmed data. As depicted in Figure 1, a sensed data in a participatory sensing application is:

- obtained from physical sensors (e.g. accelerometer) or human observations (e.g., accident in the road);
- defined in time and space;
- acquired automatically or manually;
- structured or unstructured;
- shared voluntarily or not.

To illustrate this type of system, consider an application for transit monitoring, like Waze. Users can share observations about accidents or potholes manually. Additionally, an application can calculate and share automatically a car speed based on GPS data (several portable devices are capable of being programmed for automatic data capture). With the speed of different cars at the same time and area, it is possible to infer, for instance, congestions. Since users use an application designed for a specific purpose the sensed data is structured. Instead, if a user uses a microblog (e.g. Twitter), the sensed data would be unstructured (e.g., message sent by user X: “traffic now is too slow near the main entrance of campus”)

Location sharing services, such as Gowalla and Foursquare, are also examples of participatory sensing applications. The sensed data is an observation (check-in) of a particular place that indicates, for instance, a restaurant in a specific place. Analyzing a dataset from this service, it is possible to discover what is around you, or receive recommendations of places to visit.

4. PARTICIPATORY SENSOR NETWORKS

In a participatory sensor network (PSN) (Figure 2), a consumer portable device forms a fundamental building block. In this scenario, users carry these devices that can help them to make important observations at a personal level. The sensed data is, then, sent to a server, which we could also call the “sink node”. This leads to particular characteristics of a participatory sensor network:

³<http://www.gowalla.com>

⁴<http://www.foursquare.com/>

- nodes are autonomous mobile entities;
- sensing depends on the nodes that will participate in the sensing process;
- nodes transmit the sensed data directly to the sink;
- sink only receives the data and does not have control over the nodes; and
- there is no communication between nodes.

To analyze this type of network we consider two location sharing services: Gowalla and Brightkite. The main reasons for choosing this kind of service are due to their popularity, and the availability of public datasets [4]. Associated with a check-in we can track the user coverage in specific areas, and also their sharing patterns. Since other data could be aggregated into the check-in data (e.g., temperature), the obtained results upon analyzing these systems are relevant for other participatory sensing applications.

To explain the network analyzed in this work, consider Figure 3. This figure represents four users and their actuation in three different times. Locations shared by users at each time are pointed with dashed arrows. Note that users not necessarily participate all time. We can represent all shared locations in the samples as a graph, where nodes represent shared location, and edges connect shared locations by the same user (this is represented in the figure with the label “Total Time”). With this graph we can extract many valuable information, such as the user trajectory.

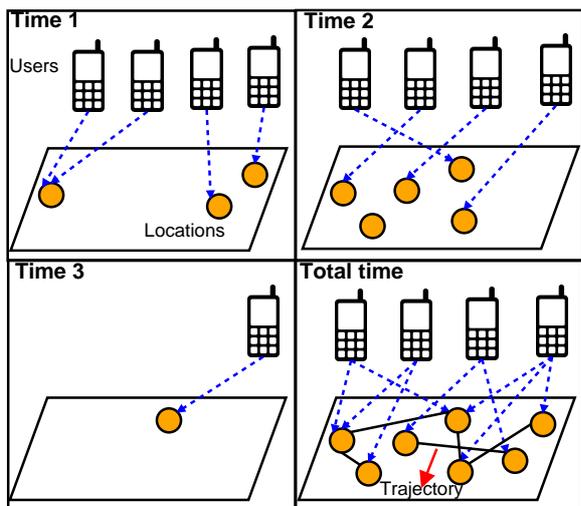


Figure 3: PSN analyzed: location sharing services

Given the ubiquity of cellphones, it is possible to include people with different interests, providing a remarkably scalable and affordable infrastructure, as we can see in Figure 4 that shows the plot of all shared locations in Gowalla, and Brightkite. In Section 5, we present and discuss more details of the PSN properties.

5. PSN CHARACTERISTICS

In this section we discuss pros and cons of participatory sensor networks. Figure 4 depicts the coverage in PSNs,

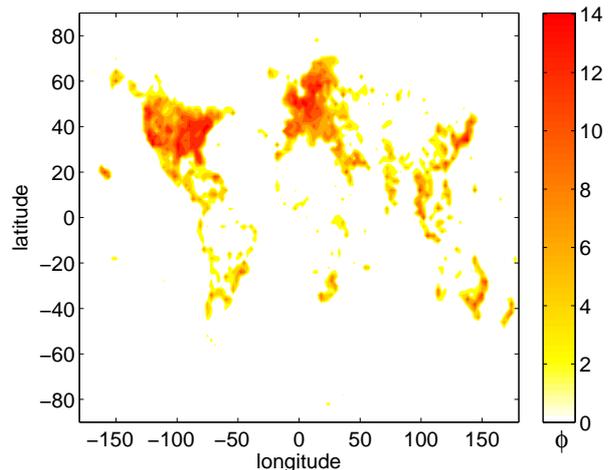


Figure 4: All sensed locations. The number of locations n per pixel is given by the value of ϕ displayed in the colormap, where $n = 2^\phi - 1$.

which can be very comprehensive in a planetary scale. Despite the global magnitude of the coverage, it is important to analyze the total number of sharing data per region as shown in Figure 5. Observe that for the Gowalla network, the vast majority of the participation is concentrated in North America and in European locations. Note that Brightkite had its popularity decreasing after a certain period, but we can still see that most of the contributions come from a single region: North America.

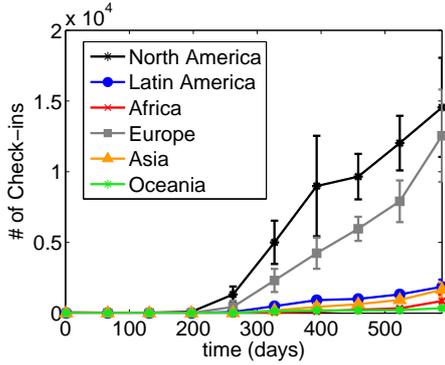
Participatory sensor networks are very scalable because their nodes are autonomous, i.e., users are fully responsible for their own functioning. Since the cost of the network infrastructure is distributed among the participants, this enormous scalability and coverage are achieved without significant costs. The key challenge to the success of this type of network is to have sustained and high quality participation. In other words, the sensing is efficient as long as users are kept motivated to share their resources and sensed data frequently.

Figure 6 presents the complementary cumulative distribution function (CCDF) of the number of check-ins per area. First, observe that a power law fitting is appropriate to explain this distribution. Second, note that for both datasets most of the locations have only a handful of check-ins and there are few locations with thousands of them. As we are analyzing location sharing systems it is natural that some locations are shared more than others. For example, locations representing a restaurant or a coffee shop are more likely to be shared than a post office, despite the fact that post offices are usually very popular as well. If our application needs a homogeneous contribution per area, we have to incentive users to participate in places that usually they would not. A punctuation system is one of many types of incentive that might work in this case. Thus, it would be interesting to compare the characteristics of the systems we analyze in this paper to systems that give a “reward” for those who share their locations no matter where they are.

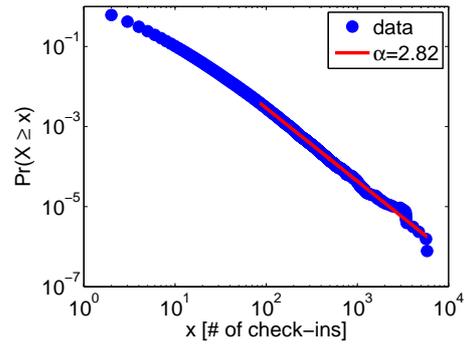
We have seen that PSNs can cover a planetary scale area. On the other hand, we have also seen that most of the check-



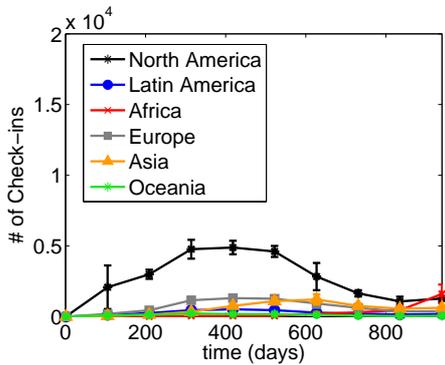
Figure 2: Participatory sensor network



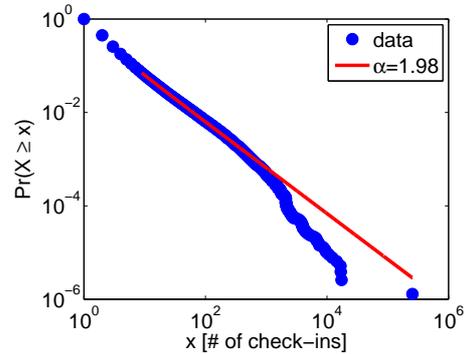
(a) Gowalla



(a) Gowalla



(b) Brightkite



(b) Brightkite

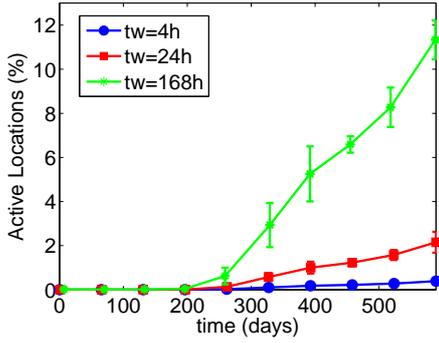
Figure 5: Number of check-ins per region

ins are concentrated in North America and in Europe. Now we verify, in Figure 7, the percentage of locations that are active in a given time window tw . For instance, when $tw = 1$ day, we verify the percentage of locations that were active at each day of the analysis. Naturally, observe that as we increase tw , the coverage also increases. However, even when $tw = 1$ week, the percentage of locations that were shared by users is still small, maximum of $\approx 12\%$ for Gowalla and $\approx 3\%$ for Brightkite. This shows that the instant coverage of PSNs is very limited considering all locations they can

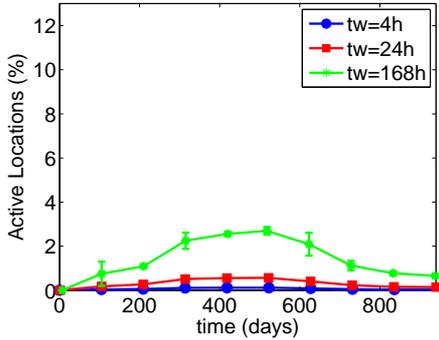
Figure 6: The complementary cumulative distribution function of the number of check-ins per area

reach, i.e., the probability of a random location to be active in a given day is very small.

Now we look at individual locations of our datasets and observe the frequency in which users perform check-ins in them. Figures 8-a and 8-c show the histogram of the inter-event times Δ_t between consecutive check-ins. Observe the bursts of activity and the long periods of inactivity in both areas, i.e., a large number of check-ins separated by a few minutes and also consecutive check-ins separated by several



(a) Gowalla



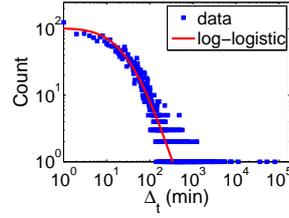
(b) Brightkite

Figure 7: The average percentage of locations that were active in a given day and their standard deviation

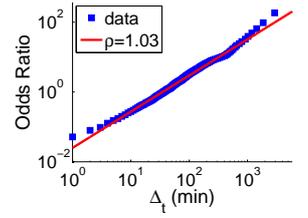
days. This may suggest that most of the data sharing, in these particular places, happens in specific intervals of time, probably related to the time that people usually visit them (e.g., in restaurants people check-in for lunch and dinner mostly). If, for instance, an application depends on sensed data from a beach area (e.g., real-time weather), it has to be aware that very few people go to the beach at night, so the sensing data will be rare.

Another interesting observation related to the inter-event times Δ_t can be drawn from Figures 8-b and 8-d. In these figures, we show the Odds Ratio (OR) function of the inter-event times Δ_t . The OR is a cumulative function where we can clearly see the distribution behavior either in the head or in the tail, and its formula is given by $OR(x) = \frac{CDF(x)}{1-CDF(x)}$, where $CDF(x)$ is the cumulative density function. As in [14], the OR of the inter-event times between check-ins also show a power law behavior with slope $\rho \approx 1$. This is fascinating, since it suggests that the mechanisms behind human activity dynamics may be more simple and general than we know [1, 9].

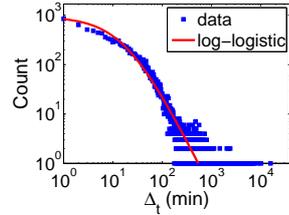
An application that naturally arises from the analysis we have shown in this section is area classification. Given the large variety of places available and all the information we



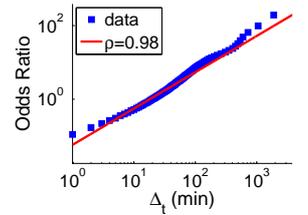
(a) Gowalla, Histogram



(b) Gowalla, Odds Ratio



(c) Brightkite, Histogram



(d) Brightkite, Odds Ratio

Figure 8: The distribution of the inter-event times between consecutive check-ins in two popular areas

can extract from the check-ins, one can expect to see very distinct sensing activities from location to location. For instance, the check-in activity in a bar may be significantly different from the check-in activity in a park. Thus, in order to illustrate this idea, Figure 9 shows the heatmap of locations considering two features. First, we consider the median of the inter-event times Δ_t of the location. Second, we consider the ratio of the number of distinct users who performed a check-in to the total number of check-ins in the location.

In Figure 9 we can clearly see three different groups, or clusters, of areas, named: A, B, and C. These groups represent different behavioral sharing patterns. Group A contains popular locations, because the median Δ_t is low, where most of the users do not return frequently. An international airport could be in group A, for example. On the other hand, group B contains locations that belong to the users' routine, like schools or gyms, since the users who perform check-ins in these areas tend to repeat this activity. Finally, group C contains most of the locations. It contains areas where it is common to have a significant time between two consecutive check-ins. Moreover, users who already performed a check-in are not likely to return and check-in again. Touristic locations could be in group C, since they are very popular and users usually go only once. We can see that these results may indicate that the coverage of the network is linked to the users' social behavior, and this must be taken into account when developing algorithms and techniques for PSNs.

6. CONCLUSIONS AND FUTURE WORK

In this paper we uncovered properties of participatory sensor networks (PSNs), a new type of network comprised of

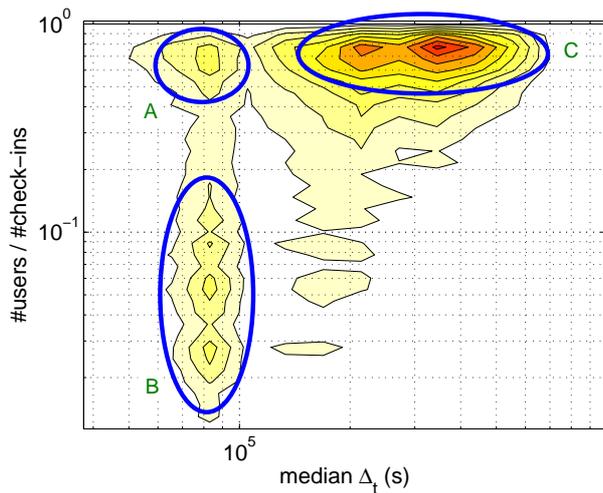


Figure 9: Heatmap of inter-event times between consecutive check-ins by the ratio of the number of users and the number of check-ins

autonomous mobile entities with sensing capability. One of the main differences between PSNs and wireless sensor networks is that in PSNs the sensing process depends on whether nodes will participate. We analyzed two datasets of a particular type of participatory sensing system, the location sharing services Gowalla and Brightkite. We showed that data from participatory sensor networks brings fascinating opportunities for the problem of sensing large scale areas. This is true mainly because it can achieve high coverage (planetary scale) without significant costs. However, we also showed many challenges of this emerging type of network, such as the highly skewed spatial-temporal sensing frequency.

At this time we are working in two main directions. First, we are analyzing other types of participatory sensing systems to complement our analysis on location sharing services. Second, we are studying actual incentive mechanisms for participatory sensor systems and their implications on the participation rate of the users.

7. REFERENCES

- [1] A. Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435:207–211, May 2005.
- [2] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory sensing. In *In: Workshop on World-Sensor-Web (WSW'06): Mobile Device Centric Sensor Networks and Applications*, pages 117–134, 2006.
- [3] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui. Exploring Millions of Footprints in Location Sharing Services. In *Proceedings of the Fifth International Conference on Weblogs and Social Media*, Menlo Park, CA, USA, July 2011. AAAI.
- [4] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, pages 1082–1090, New York, NY, USA, 2011. ACM.
- [5] S. B. Eisenman, E. Miluzzo, N. D. Lane, R. A. Peterson, G.-S. Ahn, and A. T. Campbell. Bikenet: A mobile sensing system for cyclist experience mapping. *ACM Trans. Sen. Netw.*, 6(1):6:1–6:39, Jan. 2010.
- [6] R. K. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. F. Abdelzaher. Greengps: a participatory sensing fuel-efficient maps application. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, MobiSys '10, pages 151–164, New York, NY, USA, 2010. ACM.
- [7] J. Krumm. *Ubiquitous Computing Fundamentals*. Chapman & Hall/CRC, 1st edition, 2009.
- [8] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell. A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(9):140–150, sept. 2010.
- [9] R. D. Malmgren, D. B. Stouffer, A. E. Motter, and L. A. N. Amaral. A poissonian explanation for heavy tails in e-mail communication. *Proceedings of the National Academy of Sciences*, 105(47):18153–18158, November 2008.
- [10] A. J. Mashhadi and L. Capra. Quality control for real-time ubiquitous crowdsourcing. In *Proceedings of the 2nd international workshop on Ubiquitous crowdsourcing*, UbiCrowd '11, pages 5–8, New York, NY, USA, 2011. ACM.
- [11] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu. Ear-phone: an end-to-end participatory urban noise mapping system. In *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, IPSN '10, pages 105–116, New York, NY, USA, 2010. ACM.
- [12] S. Reddy, D. Estrin, M. Hansen, and M. Srivastava. Examining micro-payments for participatory sensing data collections. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pages 33–36, New York, NY, USA, 2010. ACM.
- [13] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo. Socio-spatial Properties of Online Location-based Social Networks. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011.
- [14] P. O. S. Vaz de Melo, C. Faloutsos, and A. A. Loureiro. Human dynamics in large communication networks. In *SDM*, pages 968–879, 2011.
- [15] M. Weiser. The Computer in the 21st Century. *Scientific American*, 265(3):94–104, 1991.
- [16] M. Weiser. Some computer science issues in ubiquitous computing. *Commun. ACM*, 36:75–84, July 1993.
- [17] M. Weiser. Ubiquitous computing. *Computer*, 26:71–72, October 1993.
- [18] J. Yick, B. Mukherjee, and D. Ghosal. Wireless sensor network survey. *Computer Networks*, 52(12):2292 – 2330, 2008.