Philosophical Transactions of Royal Society, 370 ser. A (1958), 2012, pp. 176-197.

Human-centric Sensing

By Mani Srivastava†, Tarek Abdelzaher‡, and Boleslaw Szymanski¶

Abstract

The first decade of the century witnessed a proliferation of devices with sensing and communication capabilities in the possession of the average individual. Examples range from camera phones and wireless GPS units to sensor-equipped, networked fitness devices and entertainment platforms (such as Wii). Social networking platforms emerged, such as Twitter, that allow sharing information in real time. The unprecedented deployment scale of such sensors and connectivity options usher in an era of novel data-driven applications that rely on inputs collected by networks of humans or measured by sensors acting on their behalf. These applications will impact domains as diverse as health, transportation, energy, disaster recovery, intelligence, and warfare. This paper surveys the important opportunities in human-centric sensing, identifies challenges brought about by such opportunities, and describes emerging solutions to these challenges.

1. Introduction

Our work is motivated by the recent surge in sensing applications characterized by distributed collection of data by either self-selected or recruited participants for the purpose of sharing local conditions, increasing global awareness of issues of interest, computing community statistics, or mapping physical and social phenomena. This type of applications has recently been called *participatory*, *opportunistic*, or *human-centric* sensing [1]. Examples of early applications include CarTel [2], BikeNet [3], MMM2 [4], and ImageScape [5], among others.

A confluence of technology trends has precipitated the advent of such sensing applications, where the focus of sensing processes is more personal or social [6]. The first set of technology trends has to do with the proliferation of a wide variety of sensors in the possession of the average individual. The second set of trends lies in the proliferation of options for ubiquitous and real-time data sharing, as exemplified in the ubiquity of smart phones with network connectivity and the increasing popularity of social networking sites (e.g., Twitter) for information publishing.

On the sensor front, RFID tags, smart residential power meters (with a wireless interface), camera cell-phones, in-vehicle GPS devices, accelerometer-enhanced entertainment platforms (e.g., Wii-fit), and activity monitoring sportsware (e.g., the Nike+iPod system) have all reached mature market penetration, offering unprecedented opportunities for data collection. Major industry initiatives, such as

† Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, CA 90095, USA

‡ Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

¶ Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

HealthVault, automate collection of and access to information. A significant number of vendors announced wearable health and biometric monitoring sensors since 2008 that automatically upload user data to HealthVault.

On the social and networking fronts, the ubiquitous proliferation of cell-phones and social network cites offers ample opportunities for real-time data sharing. Modern cell phones are equipped with a non-trivial collection of sensors, in addition to Bluetooth, WiFi, 4G, and near-field communication options that turn the device into a portal for connecting physical world instrumentation to the Internet. Vehicular Internet access, offered in some new car models (e.g., Chrysler's Uconnect Web, and BMW's ConnectDrive) enables new applications that exploit network connectivity to export sensory information on the move. For example, services such as OnStar have long since exported on-board diagnostics (OBD-II) measurements to offer remote access to a large number of vehicle sensors and gauges.

The availability of sensing devices, Internet connectivity options, and social forums for information sharing open up an important new category of distributed applications in energy, health, environmental, and military domains that rely on individual and community sensing. New research challenges emerge from the involvement of human populations in a sensory data collection and decision-making loop. They include incentives, recruitment, privacy, trust, data accuracy, system modeling, and interpretation of social sensing dynamics. This paper categorizes human-centric sensing applications, surveys the aforementioned challenges, and discusses emerging solutions.

2. Dimensions of Human-centric Sensing

In traditional sensor networks, the emphasis has been on unattended and autonomous system operation, with the run-time role of humans limited to being end-consumers of information products (using information products from the sensor network to make decisions and take actions). By contrast, the distinguishing aspect of human-centric sensing systems is a larger involvement of humans along other points in the data-to-decision path. This path generally consists of *sensing* (i.e., acquisition of sensor measurements by observing a target of interest), and *information processing* (e.g., extracting relevant information and metadata from sensor measurements, and fusing and analyzing such information from multiple sources to derive knowledge that forms the basis of decisions and actions). Human involvement is particularly useful in *sensing* various processes in complex personal, social, and urban spaces where traditional embedded sensor networks suffer from gaps in spatiotemporal coverage, limitations in making complex inferences, inability to adapt to dynamic and cluttered spaces, and aesthetic and ergonomic problems. By taking advantage of people who already live, work, and travel in these spaces, and their adaptability and intelligence, human-centric sensing makes it feasible to get information that otherwise is not possible. While human-centric sensing systems are quite diverse, one can classify them in terms of the extent and role of human participation, which falls under one or more of the following categories:

- **Humans as targets of sensing:** Perhaps the most obvious form of human-centric sensing is one where the humans are the target of sensing. While such sensing systems have existed for a long time, particularly in the security domain, the advent of more pervasively deployed sensor technologies have resulted in an exponentially increased interest in applications which have as their goal the sensing of human activities, behaviors, and patterns at scales ranging from individuals (often oneself) to larger groups and communities. In this paper, we shall focus on examples of applications where individuals are a knowing and willing target of sensing (e.g., longitudinal health monitoring) as opposed to ones where human involvement is not voluntary and perhaps even adversarial (such as tracking enemy dismounts in a battlefield).

- **Humans as sensor operators:** A second role for humans is to participate in community sensing campaigns, either via explicit recruitment or implicitly by downloading a sensing application. Such campaigns typically exploit sensors that individuals own (such as cameras) to collect and share raw measurement data and media streams. The advent of powerful consumer-grade mobile smartphones equipped with embedded or wirelessly connected sensors has suddenly enabled billions of individuals to collect geo-tagged sensor measurements and media streams about their immediate spaces, such as an image or a sound clip or a temperature reading. Other objects besides smartphones that are associated with humans, such as the vehicles they operate, may also embed sensors collecting measurements. Not only does such sensing naturally provide sensor coverage where interesting processes are happening, but also the human expertise in intelligently operating the sensor is useful in gathering higher quality measurements (e.g., capturing high quality images in a cluttered space with poor lighting).
- **Humans as data sources:** Humans regularly act as data sources themselves, acquiring and disseminating information on their own, without the aid of sensing devices. Indeed, humans are versatile and unique sources of information about processes and relationships that exist in their spaces. In the defense and security arena there is a long history of information gathered via HUMINT (HUMan INTelligence) as opposed to electronic sensors. In social, behavioral, and medical sciences, Ecological Momentary Assessments (EMAs) of human subjects are commonly used to acquire information that is hard to get from physical sensor sources. In this paper, we focus on emerging social sensing and information dissemination that arises thanks to our ubiquitous Internet connectivity, coupled with the increasing popularity of social networks and media services, such as Facebook, YouTube, and Twitter that offier scalable tools for human-sourced information dissemination. These developments have motivated research that aims at a deeper understading of the emergent aggregate behavior of such self-organized social sensing systems and networks.

Note that, these different roles that a human can play in sensing are in general not mutually exclusive. For example, human-assisted sensor measurement (where humans are operators of sensors) may have as its target a physical process, such as mapping pollution, or a human such as monitoring state of health. It should also be noted that the above human roles with respect to *sensing* are orthogonal to roles that humans (and machines) may play with respect to *data processing*. We view data processing challenges as cross-cutting issues. Data processing may have different purposes such as data modeling, privacy enforcement, security, and trust management.

There as several examples of human participation in data processing, downstream from sensing. For example, humans may be involved in *annotating sensor data*, where information features are extracted and annotated with metadata describing the context. These tasks are often difficult to do algorithmically and autonomously at the sensor, particularly for complex phenomena and rich sensor data types. At the same time, raw sensor measurements are usually too noisy and not semantically rich enough to disseminate unprocessed further downstream along the data-to-decision path. Human participation and assistance can simplify them considerably. For instance, having an individual who captured an image with a smartphone camera to also classify, triage, and tag the image with information identifying the scene can aid subsequent analysis of that image. This results in a significantly more useful and efficient sensing system than one where all the raw images captured were blindly sent to further up the stack. Crowd-sourcing platforms such as Amazon Turk can be leveraged to engage a large number of human participants in processing raw sensor data [7]. Besides annotation, humans may play a crucial role in *data analysis and fusion*, analyzing and fusing data from diverse sensors and other sources, and extracting semantically rich and actionable inferences. In some cases, sensor-sourced information has imperfections, such as information gaps, errors, uncertainty, bias, obfuscation, and willful falsification, making automated inference hard. In other cases, such as natural language and images, the sensory information may be too complex for current machine learning algorithms. Human analysts can alleviate these shortcomings by bringing knowledge about the social, political, economic, and cultural context in which the sensory information was obtained to assess its overall trustworthiness, and derive useful knowledge despite the imperfections.

Finally, in addition to where in the data-to-decision pipeline the human participation occurs, another dimension of human-centric sensing is the *nature and purpose* of human participation. The nature of participation can span a range of possibilities that include voluntary, opportunistic, incentivized, directed, and organized, while the purpose may include collecting sensory information for self-analysis, a top-down directed sensing campaign for a director's purpose, or bottom-up data collection that emerges naturally from the participants' cause.

In the sections below, we focus on the roles humans play in sensing. We first explore applications where data are collected about individuals for self-analysis. We then describe humans as sensor operators in coordinated community sensing applications, where a community of self-interested or incentivized parties join a sensing compaign or an otherwise coordinated effort to collect information. Finally, we cover humans as information sources, and explore human-centric sensing that emerges naturally, for example, when communities propagate information for a shared cause in social spaces such as Twitter. Cross-cutting concerns are introduced where appropriate in the above sections.

3. Humans as Targets of Sensing

In this section, we focus on a category of human-centric sensing where an individual collects sensory information about themselves for their own use. Owing their inspiration to "lifelogging" applications (capturing and archiving memories of one's life in the form of a continuous time-series of data [8]) these sensing systems provide individuals with information about their activity, health, and lifestyle, and enable them to introspect about the choices they made, analyze their consequences, and take actions. A good example of such an application is the PEIR system [9] that enables individuals with mobile phones to learn the impact of their transportation choices on the environment due to vehicular emissions, as well as the exposure they get to environmental pollution. In addition to letting individuals introspect about their data, the PEIR system lets them selectively share it with others, and compare it against group statistics.

While the sensors providing data for such applications may be embedded in the spaces we live in, more common is to use sensors that are always carried on one's person, either built into one's mobile phone or in separate wearable sensors. For example, measurement traces from accelerometers, gyros, and GPS embedded in the mobile phones can be used to obtain a geo-stamped time series of one's activity and transportation state (such walking, running, sitting, sleeping, biking, and driving), and make inferences such as computing one's physical energy expenditure. Separate sensors, embedded in personal and social spaces, are often necessary for a variety of reasons. Proper placement of sensors on the body may not be possible in some cases with a mobile phone that is carried in the pocket or held in the hand. Sensing modalities such as an ECG and SPO2 sensors are not typically embedded in mobile phones. Size and battery life optimization considerations might further dictate the choice and location of sensors. Finally, other items of frequent personal use, such an individual's car, may be instrumeted and wirelessly connected (e.g., to a mobile phone) for real-time retrieval of sensory information over a sensor area network, or may log the data in a local memory for later retrieval when connected to a personal computer. At the back end, these applications typically use software running on the user's mobile phone or personal computer, or increasingly more commonly as a cloud service, for archiving, visualization, analysis, and sharing of the sensory information with social contacts. While human-centric sensing, as a tool for capturing and reflecting on one's life, is becoming increasingly commonplace, several technical challenges present hurdles to wide adoption, some of which we discuss here.

(*a*) *The Energy Challenge*

The applications described above seek to use the smartphone either as a sensing device, or as a communication gateway for wearable wireless sensors. However, modern smartphones are designed primarily as devices for sporadic use of personal communication, mobile applications, and web services, and not for continual sensing. Sensors such as light sensors, accelerometers, gyros, and magnetic compass were incorporated primarily for the purposes of offering richer user interfaces, such as display adaptation to screen orientation and lighting condition, and gesture based control. As human-centric sensing applications have begun to use these sensors to make continual measurements and inferences about the user's context, the limitations of the platform in terms of battery life become evident. Even seemingly simple modalities such as the accelerometer turn out to be quite energy constrained because of the high sampling rates needed for inferring physical context and the lack of architectural support in the I/O subsystems for handling sensor data streams. Complex modalities such as the GPS and imager are even more energy hungry, as are the wireless radios needed for the phone to communicate with wearable sensors.

The end result is that smart phones which may last for a day or two when used in their intended role as personal communication and computing devices, barely last for 3-4 hours when used for continual sensing.

In the short term, the key to at least partially meeting this energy challenge lies in smarter selection, activation, duty cycling, and sampling of the energy-hungry sensors such as GPS [10], while making use of information such as the current contextual state, model of expected behavior, and external constraints. Although measurements from multiple sensors may contribute to inferring a contextual variable of interest such as location, the sensitivity of the inference to the sensors may vary over time, and can be exploited algorithmically to selectively shutdown or lower the sampling rate of sensors. An example is the SensLoc system described in [11] that actively controls a GPS receiver, a WiFi scanner, and an accelerometer, and fuses their measurements to detect commonly visited places and commonly traversed paths. Additionally, prior knowledge of a road map or building layout may be used to constrain possible evolution of future location, and thus further limit the sensor samples needed [12].

In the long run, however, the smartphone platform architecture may need to evolve to support more energy efficient sampling of sensors. For example, dedicated hardware that can deposit sensor data samples to the main memory and perform simple processing on the data without waking up the main processor can significantly reduce the energy overhead [13]. Wearable sensors, used external to the smartphones, suffer from their own energy challenge, primarily due to their small size and weight that severely limits the battery size. This is particularly true for sensors designed for high rate sensing modalities, such as an ECG signal, where there is little opportunity to duty cycle naively. Instead, compute intensive local processing that would predict the occurrence of an event of interest, and specialized circuits that would detect their start at an early stage, may be used to activate and shutdown the sensors smartly. Additionally, a major source of power consumption in wearable sensors for physiological signals is the analog frontend that is used to amplify and filter the tiny signals, and, worse, these circuits are hard to duty cycle because of long time constants associated with the filters. More optimized analog-to-digital pathways together with the use of emerging compressive and event-driven sampling mechanisms, instead of the conventional Nyquist sampling, would be crucial to meeting the energy challenge. The issue of energy-efficient sensing on cell-phones has been the topic of several recent publications [14, 15, 16].

(*b*) *Challenge of Inferring Rich Context*

The utility of human-centered sensing applications comes from their ability to make inferences about individuals' contexts can be made for purposes such as personal awareness, individual behavior adaptation, personal health management, and population-level studies. For example, prior research [17] shows how the sensory data available on a typical smartphone can be algorithmically processed and fused to make complex inferences about one's mobility pattern and transportation modes. However, experience with the first generation of such applications on mobile platforms has exposed both systemic and algorithmic limitations in making complex inferences about physical, physiological, behavioral, social, environmental, and other contexts that the applications demand. The mobile platform hardware and system software typically do not perform any inferences on the measurements, and instead expose raw data or provide hooks for simple notifications and accuracy control. With little system support for processing raw sensor measurements and deriving higher-level inferences, each application is custom developed leading to a higher development effort as well as run-time inefficiencies and performance problems.

The challenge of making the smartphone into an effective platform for human-centered sensing would require solving several problems. The system software must provide semantically rich interfaces for applications to express interest in complex patterns of sensory observables, and permit composition of high-level context inferences from primitive operations. Energy-scalable lightweight machine learning algorithms than can be embedded on resource-constrained platforms are needed, along with robust and reusable primitives for rich context inferences based on them. Compact and easy to train personalized models of user behavior and preferences that tailor the inference algorithms to specific individuals are essential for improved performance [18]. Additional complications come from the wide variations in the way users hold and carry their smartphones, and addressing them would require both algorithms that are robust to such variations as well as learning and adapting to specific usage patters of a user.

4. Humans as Sensor Operators: Collection Campaigns

The next class of human-centric sensing applications we discuss are those where individuals (who operate sensors) contribute sensor measurements about themselves or about the spaces they visit as part of a larger-scale effort to collect data about a population or a geographical area. The effort is coordinated explicitly, for example through active participant recruitment, or implicity, for example by making a new application available on a cell-phone app-store where users can download it, thereby implicitly joining a sensing campaign. This is the most common human sensing model, featuring participatory and opportunistic sensing, where humans are used as sensor operators. Under this umbrella also fall applications such as: monitoring spread of a disease in a community; human subject research conducted by medical and social scientists on groups of individuals; documenting the state of the physical infrastructure in an area such as the quality of the roads [19] or level of pollution in a city [9], or the state of garbage cans on a campus [20]; and, monitoring spread of an invasive plant or animal species in an ecosystem. These applications, which have been termed "sensing campaigns" in the literature, can range in sophistication. At the one extreme are applications which involve simply collecting sensor measurements that are stamped with the time and location, and presenting them to end users who may do further analysis manually or with the assistance of visual analytics tools. At the other extreme are sophisticated applications that use the collected spatiotemporal data to compute aggregate statistics and models to assist users in identifying patterns and make predictions. These applications share several important challenges as discussed below.

(a) *Participant Recruitment Challenge*

Developers of a sensor data collection campaign face the challenge of identifying the appropriate set of individuals who would collect the data, for example, using their mobile phones. In most cases, the participation by individuals is voluntary, although there may be applications where an organization may have its employees be the data contributors as part of their jobs. The problem lies in identifying what subset of individuals who are interested and meet the basic requirements of being data contributors (i.e., have the right type of sensors and reside in the geographical area where the data collection is to be done in a window of time) is actually selected to contribute. A possible solution would be for the application to simply accept data contributors from whichever individuals decide to do so in response to an open call. Given problems such as self-selection bias in such a bottom-up approach, usually more appropriate is a top-down approach where the set of actual contributors is actively shaped to have appropriate characteristics to prevent statistical bias in the data collection. Additionally, cost and resource constraints may also place a limit on the number of participants, for example when the participants are paid or if there are opportunity costs associated with the time they would spend collecting the data. This "participant recruitment" problem has similarities to traditional employee recruitment and to sensor selection in traditional sensor networks, but needs to consider sensing needs of the data collection campaign and diverse attributes associated with the participants. Moreover, it is not a one-time selection process but an on-going one that requires methods to keep the participants engaged.

The difficulty in participant recruitment [20] comes from the human element. On the one hand, it is the human element that makes data collection campaigns so powerful. The intelligence, mobility and flexibility of participants is leveraged in making difficult measurements and rich inferences that may not be possible with unattended sensors. On the other hand, the human element introduces complicating factors. Potential participants may have different motivations, availability, diligence, skill, timeliness, phone capabilities, and privacy constraints that would affect the amount and quality of data they collect. Moreover, participants' availability and the quality of data they contribute during the actual campaign may differ from what was predicted at the time of initial recruitment, and may also vary over time due to distractions, boredom, and other human factors. This presents the need to rate contributions so as to assign reputation to participants, which may be used to modify the set of participants during a campaign, to modify the incentives offered to them, and to assess their suitability as participant in future campaigns.

Initially, the authors and organizers of a campaign provide requirements on spatiotemporal coverage and extent, sensing modalities, and budget constraints. Additionally, the recruitment process may have available to it the participants' profiles, where they are modeled in terms of their capabilities (e.g. the type of phone and sensors they possess), availability (when and where can they collect data), reputation (their experience and prior performance in data collection), and cost (the incentives they would need or the opportunity cost of their participation). Using data collected during the campaign, the recruitment is adapted to evolving campaign needs and actual data collection performance of the participants, and the participant profiles are updated as well.

(*b*) *Challenge of Learning Context-annotated Mobility Profiles*

Perhaps the most crucial metric in recruiting is a participant's availability. It would be futile to select a participant whose mobility patterns do not intersect in both space and time with the regions of space and the intervals of time that the campaign is being conducted in. Formally, given mobility profiles of potential participants, one needs to find a subset of participants that would maximize coverage in space and time while keeping the cost of running the campaign within budget. In reality, the problem is even more complex since just because a participant is at a place and time where data needs to be collected does not mean that the participant is actually in a position to do so. For example, a participant may be in a fast moving vehicle, precluding certain types of data collection. To accommodate this, the mobility profiles of participants need to be annotated with contextual information, such as transportation modality and activity state, which have an effect on participants' ability to make sensor measurement. For reasons of practical system implementation, the context-annotated mobility profiles need to be both compact and in a form that can be effectively queried by the recruitment algorithms to select participants. One approach, proposed in [21], is to maintain a participant's probability of occupancy in a spatial grid as a function of the time of the day, day of the week, and holidays. An alternative approach might be to represent mobility profile in terms of a graph composed of important places and routes that a participant visits. Run-time traces of time-stamped locations annotated with transportation modality [17] and other context, captured by a background sensing service running on a participant's mobile phone, can be checked for consistency against the participant's mobility profile, and the profile updated in case of a mismatch.

(*c*) *Data Quality and Participant Performance Challenge*

There may be significant differences between the data collection performance of different participants, and the performance of a participant may vary over time across different campaigns or even during the course of the same campaign. Desirable for this would be a metric that may be used to measure and track participation and expertise of individuals for adapting the set of participants and adjust feedback and incentives [22]. One way recent literature proposed to realize this is using "track record" or a reputation system, similar to those used in e-commerce (e.g., e-Bay). Its use has also been proposed in traditional sensor networks to monitor sensor quality. Such an approach, for example, was used in [20] where a watchdog entity observes the quantity, quality, and utility of participants' sensor data contributions. The watchdog compares against known ground truth, predictions from models, and with measurements from other participants. For complex sensing modalities, such as imagery, watchdog's task may be algorithmically quite complex, and may require human assist. The watchdog can consider factors such as data quality, relevance, and timeliness along with participant follow-through and responsiveness. The reputation metric is maintained as a Beta distribution for each participant, and consists of two parameters *alpha* and *beta*, where the former captures the number of good quality data contributions made by that participant as observed by the watchdog and the latter captures the number of poor quality ones. Reputation metrics based on Beta distribution compactly capture the stochastic uncertainty, and also allow for aging (i.e., give heavier weight to more recent performance of a participant). The system can set initial values of the reputation metric by assessing participants' performance on calibration tasks. Reputation-based approaches however do not offer a complete solution, and fail to detect more complex forms of participant misbehaviors such as collusions and sudden changes.

(*d*) *Sparse Sampling and Generalization Challenge*

Once data is collected, a data consumer may want to compute useful local features or statistics from measurements. For example, traffic patterns monitored in a city may produce statistics that help local drivers avoid congestion areas [2]. Bike route data collected by biking enthusiasts may help them understand conditions on these routes [3]. Hiker encounters may be recorded on mountain trails to help locate missing hikers on those trails [23].

More complex applications may aim at creating viable *generalizations* from data, where data collected in some locations may help create prediction models that are generalizable and can affect human decision-making elsewhere. For example, sharing data collected by smart energy meters installed in some households, together with relevant context, can lead to a better understanding of energy consumption in contemporary homes and best practices that increase energy efficiency elsewhere around the nation. Similarly, sharing data collected by activity sensors among fitness enthusiasts can lead to lifestyle recipes that promote healthier behaviors for multitudes of others. Also, sharing data on fuel consumption of individual vehicles on different road types in different conditions can help build generalizable models that predict fuel consumption of other vehicles on other roads. A common feature across these applications is the existence of a generaliazable model of the studied system or phenomenon, that can be infered or “trained” using limited available data, but that can ultimately be used to predict outcomes in a much broader context [24]. Recent work has addressed the challenge of building good models and statistics from sparse human measurements [25, 26, 27].

Consider applications that attempt to learn from collected observations and generalize by building models of system behavior where some components, interactions, processes, or constraints are not well-understood. For example, predicting the fuel consumption of a vehicle depends not only on fixed factors such as weight, frontal area, and engine type, but also on variables such as vehicle speed, acceleration, congestion patterns, and idle time, which are hard to predict accurately in advance. Building first principle models from scratch is not always practical, as too many parameters are involved. In contrast, using regression to estimate model coefficients is challenging because reliable estimation suffers the curse of dimensionality. The state space grows exponentially in the number of parameters, making sampling of that space sparse. As the number of parameters increases, estimated models become less reliable.

Recent work to address the above dilemma focused on modeling techniques that combine estimation-theory and data mining techniques to enable modeling complex socio-physical systems reliably at multiple degrees of abstraction. A reliable model is the one that remains sufficiently accurate over the whole input range. The general idea [28] is to jointly (i) partition sparse, high-dimensional data into subspaces within which reliable linear models apply and (ii) determine the best model for each partition using standard regression tools. Importantly, the modeling technique must uncover the inherent generalization hierarchy across such subspaces. For instance, in the example of predicting fuel efficiency of cars on different roads (as a function of car and road parameters), it should tell how best to categorize cars for purposes of building fuel prediction models in each category. Categorization could be by car class, make, model, manufacturer, year, or other attributes. These categories have a hierarchical structure. For example, one may build prediction models for cars by make, model and year (e.g., Ford Taurus 2005, Toyota Celica 2000). One may also aggregate these over years or over car models to generate prediction models for larger categories (e.g., all Ford Taurus cars, or all Toyotas of 2000). Such generalizations help when there is not enough data on each type of car to build a reliable model for that type alone. They are also good for predicting performance of a car from performance of others (in the same generalized category). Hence, finding accurate generalizations is an interesting problem in human-centric sensing systems where sampling is sparse and the number of parameters is large.

An interesting question is to analyze the trade-off between modeling accuracy and cost of data collection. Normally, more accurate models involve more input parameters, which makes them more expensive. By judiciously replacing a complex general model with a tree of simpler specialized models for different sub-cases (branches), one can do better both in terms of accuracy and cost; specialization may increase accuracy, while at the same time reducing the number of model paramaters needed in each special case at hand, hence reducing cost. The main challenge in achieving such an improved trade-off lies in appropriately defining the special cases and the simpler models that apply in each case.

(*e*) *The Privacy Challenge*

Privacy becomes a broad concern when a sensing system has multiple users and when data can be exposed to unauthorized parties. For example, vehicular and smart highway applications that collect and utilize traffic measurements from a variety of distributed vehicle-mounted sensors could compute real-time traffic conditions, but only if individuals shared private data on their speed and location. Adequate architectural and algorithmic privacy solutions should be designed into such data collection systems from the start. These solutions should encourage privacy-preserving information sharing. Several special characteristics of human-centric sensing systems make general (encryption-based) privacy solutions inadequate, as described below.

1. Spatio-temporal affinity: Data in human-centric sensing systems originate at a physical time and location. Privacy mechanisms must protect the data since its origination time. These mechanisms must therefore be integrated with the physical context in which the data originate and address physical attacks in that context. For example, wireless fingerprinting techniques can disclose certain information (e.g., location) about a transmitter even without decoding the content of messages [29]. In turn, the pattern of transmissions from identified sensors may disclose something about the activity monitored. For example, in a smart appartment complex, identifying the electronic footprint of a bathroom light sensor can allow others to detect a neighbor's washing habits. There is a tradeoff between effort spent on hiding this pattern and the privacy achieved. In a recent study that explored privacy mechanisms for data with spatio-temporal affinity, it was shown that an outdoors eavesdropper could fingerprint different wireless devices inside a smart home and identify distinct transmitters. It could then correlate their transmission patterns to general knowledge of human behavior (e.g., typical patterns of use of different rooms and household items), and identify semantics of different tranmitters. Finally, it could use the pattern of tranmissions together with infered semantics to recognize activities of daily living, all without access to the actual data transmitted. It was shown that the eavesdropping attack was surprizingly robust unless significant resources were invested into privacy mechanisms (e.g., a very large number of spurious messages were sent). Efficient combinations of different privacy mechanisms that achieve a better cost-privacy trade-off remain a subject of current research.

2. Streams and time-series: Data in sensor networks are usually stream oriented. The main data type is a correlated time-series (of measurements). Correlations in time-series data offer additional challenges in protecting privacy, as partial disclosure of correlated data can divulge information about other data not shared. A significant amount of work was recently dedicated to mechanisms and theory for protecting privacy of correlated data streams. Of particular interest are mechanisms that work in the absence of a trusted entity responsible for data cleaning. Anonymity alone does not always work because the shared data often has enough information to infer the anonymized source. For example, sharing GPS traces that end up repeatedly at the same private residence may suggest the identity of the data owner. In such cases, data perturbation could ensure that (i) privacy of original sensor data is preserved, yet (ii) computation of accurate community statistics remains possible from the perturned data. Recent work addressed the problem of generating the optimal noise time series for a given data stream such that minimum information is leaked (in an information-theoretic sense) about the original stream data. Yet, when many such streams are shared, the accuracy of data statistics computed over the community approach that of unperturbed data. For example, it may be possible to compute the average percentage weight gain of a community of individuals (as a function of time), while obfuscating each individual stream in ways that mask both the absolute values and trends [25]. Another set of randomization techniques preserve *differential privacy* using randomized aggregation functions [30, 31, 32]. When an aggregate value is derived by a trustworthy entity or the client, differential privacy is preserved if adding or removing a data item does not significantly change the output (aggregate) probability distribution. Like data point perturbation, differential privacy methods rely on randomization that introduces noise to the regression model.

3. Data Fusion: Sensor fusion is a common operation. When data is collected from multiple users, a trade-off exists between individual privacy (that favors nondisclosure) and community-wide utility from aggregate information (that requires data to be disclosed in some form). Recent work has shown that this trade-off can be significantly improved by exploiting the specifics of the data fusion algorithms themselves. It makes the observation that, for instance, an algorithm that computes a public regression model (from community data) does not actually need the individual training samples. Sharing properly computed *aggegates* of such samples from each user can result in the same community model as that computed from raw user data. Computing such aggregates, called *neutral features* [27], must meet two constraints; (i)*perfect modeling*, which means that construction of models from shared neutral features should produce exactly the same model as if the original private data traces were used, and (ii) *perfect neutrality*, which requires that reconstruction of private user data from shared features yield the same error as if no additional information was available to the outside world besides the computed community model. Neutral features were recently described that provide perfect modeling and approach perfect nuetrality [27].

5. Humans as Data Sources: Sensing in Social Networks

In the previous section, we discussed how humans can assist in coordinated sensing campaigns that exploit sensors they operate. Coordination may have been explicit (by recruitment) or implicit (by downloading a shared application, designed for a particular sensing purpose). Here, in contrast, we focus on the emergent behavior of humans, interconnected by communication media, when they naturally act as data sources by volunteering information they care about. While individuals may be acting without any prior coordination and without specialized sensing application download, the very means of information sharing the use, such as Twitter, Facebook, or YouTube, eventually produce emergent community sensing behavior. In these scenarios, information is generated, observed, and collected by humans, and then propagated through a social network. Information disseminated in the network can be treated as the output of a very sophisticated sensor [33], a human with its high capacity to process and filter collected observations. With a variety of software platforms, such as browsers and social network interfaces, executed on personal hardware platforms, such as smart phones, PDAs, or plain cell phones with cameras, humans can monitor and collect streaming data or generate status updates to their informal social networks, businesses, and media outlets all over the world. The aggregation of such updates creates a live, vivid model of the social and physical environment from which those data are collected. These technologies in the hands of willing humans can create a very broad, distributed system for collaborative sensing, processing and communication of information, based on the most versatile mobile platform: the human source. The system puts to use the human capabilities that uniquely distiguish them from electronic sensors. It also comes with the fallacies of willful falcification, human deceipt and data manipulation, that need to be addressed.

Consider a scenario in which human-centric monitoring is triggered by human reactions to unfolding events or observations spontaneously, unilaterally, and without prior coordination. For example, a person might report local conditions to further the cause of their favorite political belief, warn of abuses or law violations, or just capture an event of general or humanitarian interest, such as an accident (especially involving people who require assistance), an uncommon occurrence or an unusual observation. In such cases, the human ability to judge (most of the time correctly) the need for reporting the event makes social sensing a very powerful tool, difficult to match even by the most sophisticated software applications deployed in a sensor or a robot. Often the "sensing" is first done visually by a human, and then some corroborating evidence (e.g., a sound clip or a video) is collected by a human-operated device. The net effect of humans engaging in such sensing behavior creates an emergent ad-hoc, self-organizing structure to report socially important events. Examples include reporting illegal arrests, brutal police interventions, and civil rights abuses.

What makes such reporting especially powerful is the ubiquity of social networks such as Twitter, Facebook or YouTube that offer a forum for global dissemination of the reported data. These networks provide three technical capabilities that empower the human user; namely, (i) a publishing capability, (ii) a global search capability, and, perhaps most importantly (iii) a capability to subscribe to data feeds from trusted sources. Both the theoretical analysis [34] and current and historical evidence indicate that the reporting structure that emerges from uncoordinated human sensing has a behavioral phase transition. Below a critical threshold on the percentage of involved population, referred to as the *tipping point*, the structure fails at significant information dissemination. Above the tipping point, however, successful dissemination reaches the broader community.

It is both challenging and interesting to fully understand the process of information propagation through the aforementioned structure and the influence of its underlying social network. Such very informal social sensing has already increased accountability of law enforcement and became the basis for supporting democratization of governments in many countries. Analysis of its impact on future society, on human interactions, on trust in media and governments, as well within formal and informal social structures are challenges of enormous complexity but also of fundamental importance.

(*a*) *Information Propagation Challenges in Social Networks*

To address the challenge of understanding the emergent behavior of human-centric sensing systems described above, such as the dissemination of opinions and their supporting evidence, it is first necessary to understand diffusion of information in social networks and the conditions under which the information reaches the majority of the society involved. The current research identified three factors fundamental to predicting whether such information spread will be successful. They are (i) trust among members of the social network that collects the data and distributes opinions and their corroborating evidence, (ii) commitment of members of the social network to the distributed opinion, and (iii) the size of the committed community. It appears, that there is a tipping point of the fraction of the committed community in the society above which the spread of minority opinion is rapid (logarithmic in the size of the network, but below it is excruciatingly slow (exponential in the size of the networks) [35]. We discuss each of these three factors below.

In general, the structure of the communication layer of a social network is one of the primary factors in defining the degree of information spread in the negative sense, insufficient communication layer suffocates spreads, but once it exceeds certain threshold, other factors become dominant in defining the information spread. One of them is the social network structure. Usually, scale-free networks and small-world networks are used in studying social networks because they seem to match well characteristics of such networks. In scale-free networks, node degrees are distributed according to the power law, resulting in a small subset of nodes having connections of high degrees, while most of the nodes have low connectivity. In such networks, dynamics of many processes are independent of the number of nodes in the network. In small-world networks, nodes are highly clustered, with short path lengths between nodes. They are commonly found in biological, social, and synthetic systems [36, 37], and were also identified in patterns of co-authorship of scientific publications [38] as well as in involvement of actors in the same movies [39]. The information spread has been found to be especially fast in small-world networks [40]. Dynamic networks, in which nodes and edges may appear and disappear with time, are increasingly popular in the recent studies [41, 42, 43] as they provide more realistic model of evolving communication graphs connecting a relatively static set of social network members. The dynamics of connectivity of social network members define interacting pairs of members at each time step (e.g., see [44]) and therefore determines the range of the information spread. Moreover, the information itself may change the opinion of the recipient, motivating him or her to join the social network that originated the information spread.

In [45], the authors present a general model of information diffusion in dynamic social networks to examine how network structure, seeding strategy, and trust among interacting participants affect the diffusion process. The model is based on four axioms: (1) Information Loss Axiom, (2) Source Union Axiom, (3) Information Fusion Axiom, and (4) Threshold Utility Axiom. These axioms define the diffusion process by specifying what happens to the message as it is propagated, how the nodes handle information they receive, and how nodes update their properties based on their interactions and the information they receive. The results of diffusion following these for axioms demonstrate that that trust between communicating individuals strongly affects the reach and impact of the diffused information. Assessing trust between communicating members of a social network is therefore essential to understanding information diffusion within such network. Assessment of trust is a challenging problem, even in a small community whose members can be provide their self-assessment of trust. The recent rise of large social networks has created a bigger challenge: how to measure trust between members of a large social community whose many members never meet each other in person?

(*b*) *The Trust Assessment Challenge*

In general, trust is complex and little understood dyadic relation among members of a social network. Yet, trust is often fundamental in the very formation of social networks. It defines each members assessment of the quality and credibility of information received and determines the range and impact of information spread. Recent measurement of social interactions [46] show that such assessments are greatly influence by often unconscious, social signals easily perceived in the direct interactions, and practically totally lost in communication via email. Once gained, trust can serve to identify influential nodes in a network and to determine whether other nodes will believe information that they receive and whether they will transmit it to other nodes or act on it themselves.

It has been also established that receiving information that is believed to be true enhances the trust of the receiver in the sender. Consequently, continued information exchange between members of a community can enhance trust in their relationships. In [47], the authors present algorithmically quantifiable measures of trust based on communication behavior of the members of a social network with sparse direct contacts. The basis of this approach is an assumption that trust results in communication behavior patterns that are statistically different from communication between random members of a network. The proposed quantitative measure of who-trusts-whom relation in the network relies on detecting statistically significant patterns of the trust-like behavior. This measure is based on quantifiable behavior of participants (not the textual content, as many others do) and, thus, it is referred to as *behavioral trust*. The authors developed algorithms to efficiently compute behavioral trust and validated these measures on a Twitter network data. They also demonstrated that this new set of behavioral measures can be used to assess the existence, emergence or dissolution of trusting relationships in large social networks.

Two types of behavioral trust were identified in a study involving Twitter data. In a *conversational trust*, the basis for measuring trust is the length and balance of conversations between two nodes. In *propagation trust*, the metric is based on the percentage of tweets sent by node A that node B retweets. The conversational trust is symmetric, but the propagation trust is not, because node A may not trust node B, even if B retweets all tweets of A. The authors conjectured that trust is the foundation of communities, and that it should be possible to discover communities in the Twitter network by identifying clusters whose members trust each other. To test this conjecture, they analyzed the tweets of 2 million nodes Twitter network and created communities based on conversational and propagation trust. The resulting two trust-graphs have similar structure, having roughly the same number of communities, as well as a very similar average community size. The trust-based communities created from conversational and propagation trust have a similarity higher modularity [48] than could be expected for random graphs of the same size and node degree distribution. This result confirms that the trust-based communities capture similar relationships.

(*c*) *Formation of Opinions and the Tipping Point Quantification Challenge*

Opinions that dictate human attitudes and behavior arise dynamically via interactions among individuals within their social networks. In the past, personal interactions within social networks have been the major force in moving societies towards consensus in the adoption of ideologies, traditions and attitudes [46, 49, 50]. Today, as a result of proliferation of online social networks and ubiquity of wireless communication, the dynamics of social influence have been made much more complex by addition of interactions enabled by the technology. These interactions are strongly enriched by the ability of advocates of an opinion to provide direct evidence of their claims, such as examples of abuse of power, or demonstration of strength of the opposition rallies. While, in the past, such evidence would not have been available or would have been easily suppressed, today it can spread widely and quickly via the internet, tweets or social networks. The effects of this change have been extensively studied in prior literature [51, 52, 53]. A question of great interest in emergent human-centric sensing is how well-supported an opinion has to be for it to spread to the entire population. Such processes are well recognized in sociological literature and referred to as minority influence [54].

This question is studied in [34, 35] under a model of the dynamics of competing opinions that assumes that switching an individual's state has little overhead (hence, a very democratic society is assumed, in which joining the opposition carries no penalty). The authors adopt a two-opinion variant [55] of the Naming Game (NG) [56, 57, 58]. The evolution of the system in this model takes place through the usual NG dynamics, which can be summarized as follows. At each simulation step, a randomly chosen *speaker* voices a random opinion from his list of opinions to a randomly chosen neighbor, designated as a *listener*. If the listener possesses the spoken opinion in his list, both speaker and listener retain only that opinion; otherwise the listener adds the spoken opinion to her list. The order of selecting speakers and listeners is known to influence the dynamics [58]. In [34], authors choose the order in which the speaker selects the listener. The NG has also been investigated on spatially-embedded sparse random networks [59]. Unlike frequently used opinion dynamics models, in NG, a node may possess many opinions simultaneously. This is significant because it impacts the times that the network needs to reach a consensus when starting from a uniform initial condition.

In [34], the authors study the evolution of this model starting from an initial state where all nodes adopt a given opinion B, except for a finite fraction of nodes, called the *committed nodes*, that are in state A. Such committed nodes, introduced in [60], behave differently than normal nodes. If chosen as speakers, they speak their committed opinion, but when acting as listeners, they ignore their input. Thus the committed nodes never change their opinion. When the committed nodes of only one opinion are present, their opinion is the only absorbing fixed point of the system, and the consensus state in which all nodes eventually end up is the one in which all nodes share the committed opinion.

The authors show that under a simple condition, the majority opinion in a population can be rapidly reversed. This condition is related to the existence of a tipping point in terms of the fraction of nodes in the network that are committed. When this fraction reaches a critical value (which is about 0.0979 for the complete graphs), there is a dramatic change in the time needed for the entire population to adopt the committed opinion. Converting the entire community to the committed opinion of a minority is exponential in the size of the entire population when the fraction of the committed nodes is below the tipping point. However, above the tipping point, the conversion time is logarithmic in the population size. Simulation results for Erdos-Renyi random graphs showed qualitatively similar behavior. The authors also pointed to some historical examples of such conversions that included the suffragette movement in the early 20th century and the success of the American civil rights movement that followed quickly the time at which the size of the African-American adult population crossed the 10% mark.

In conclusion, the sensing in social networks is limited by the trust that the senders of the evidence (or the information about the evidence) enjoy in their community. Yet, the challenge is how to measure trust in large social networks, with members interacting via communication links and not directly. How widely such information diffuses is also strongly affected by the social community structure. The challenge in discovery of a community structure is how to distinguish between casual interactions and the social relations of two members of a social community. With strongly committed nodes, once the size of the community exceeds the tipping point of about 10%, the opinion supported by the collected evidence and associated opinions can rapidly propagate throughout the entire community. However, we do not know today how to make the node committed at the model and in the real-life why some people stick to their opinion, while others are willing to adjust them based on opinions of their neighbors. Another interesting and not fully satisfactorily answered question is what impact media have on the opinions of the community. Currently researched issue is the dynamics of the model when several opinions have committed members adopting them. Close collaboration of social and political scientist with computer and network scientist is needed to address those questions and challenges.

6. Conclusions

In this paper, we reviewed the technological trends that enable a new brand of sensing applications, where humans are more intimately involved in the sensing and data processing loop. Proliferation of sensors in common use, wide-spread deployment of communication capabilities, and the advent of social networks that enable broad information dissemination make up the technological foundations of human-centric sensing. Several sensing scenarios were discussed ranging from those where humans collect sensor data for personal use to those where globally directed, self-organized, or emergent data collection and sharing takes place in a community of interest. Several challenges remain topics of current research. Those include front-end challenges (e.g., energy consumption), coordination challenges (e.g., campaign recruitment), back-end challenges (e.g., modeling and prediction), and challenges in the overall understanding of the emergent behavior of social sensing systems as large. While a significant amount of research has already been undertaking along those fronts, much remains unsolved. New interdisciplinary research is needed to bring about better mechanisms and a better theoretical understanding of emerging human-centric sensing systems in a future sensor- and media-rich world.

Acknowledgements

This material is based in part upon work supported by NSF under awards CNS-0910706, CNS-1040380, and CNS-0627084, by the NSF Center for Embedded Networked Sensing at UCLA, and by the Army Research Laboratory, and accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] J. Burke *et al.*, "Participatory sensing," Workshop on World-Sensor-Web, co-located with ACM SenSys, 2006. [Online]. Available: http://www.sensorplanet.org/wsw2006/

[2] B. Hull *et al.*, "Cartel: a distributed mobile sensor computing system," in *Proc. of SenSys*, 2006, pp. 125–138.

[3] S. B. Eisenman *et al.*, "The bikenet mobile sensing system for cyclist experience mapping," in *Proc. of SenSys*, November 2007.

[4] M. Davis *et al.*, "Mmm2: Mobile media metadata for media sharing," in *CHI Extended Abstracts on Human Factors in Computing Systems*, 2005, pp. 1335–1338.

[5] S. Reddy *et al.*, "Image browsing, processing, and clustering for participatory sensing: Lessons from a dietsense prototype," in *Proceedings of Embedded Networked Sensors, EmNets '07*, 2007, pp. 13–17.

[6] A. PENTLAND, "To signal is human," *American scientist*, vol. 98, no. 3, pp. 204–211, 2010.

[7] T. Yan, V. Kumar, and D. Ganesan, "Crowdsearch: Exploiting crowds for accurate real-time image search on mobile phones," in *8th Annual International Conference on Mobile Systems, Applications and Services (MobiSys 2010)*, 2010.

[8] J. Gemmell, G. Bell, R. Lueder, S. Drucker, and C. Wong, "Mylifebits: Fulfilling the memex vision," in *ACM Multimedia*, December 2002.

[9] M. Mun, S. Reddy, K. Shilton, N. Yau, P. Boda, J. Burke, D. Estrin, M. Hansen, E. Howard, and R. West, "Peir, the personal environmental impact report, as a platform for participatory sensing systems research," in *7th Annual International Conference on Mobile Systems, Applications and Services (Mobisys)*, June 2009.

[10] W. Ballantyne, G. Turetzky, G. Slimak, and J. Shewfelt, "Powerdown: Achieving low energy-per-fix in cell phones," *GPS World*, vol. 17, no. 7, pp. 24–32, 2006.

[11] D. Kim, Y. Kim, D. Estrin, and M. B. Srivastava, "Sensloc: Sensing everyday places and paths using less energy," in *ACM Sensys*, November 2010.

[12] I. Constandache, R. Choudhury, and I. Rhee, "Towards mobile phone localization without war-driving," in *INFOCOM, 2010 Proceedings IEEE.* IEEE, 2010, pp. 1–9.

[13] B. Priyantha, D. Lymberopoulos, and J. Liu, "Enabling energy efficient continuous sensing on mobile phones with littlerock," in *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks.* ACM, 2010, pp. 420–421.

[14] J. Paek, J. Kim, and R. Govindan, "Energy-efficient rate-adaptive gps-based positioning for smartphones," in *Proceedings of the 8th international conference on Mobile systems, applications, and services*, ser. MobiSys '10, 2010, pp. 299–314.

[15] Z. Zhuang, K.-H. Kim, and J. P. Singh, "Improving energy efficiency of location sensing on smartphones," in *Proceedings of the 8th international conference on Mobile systems, applications, and services*, ser. MobiSys '10, 2010, pp. 315–330.

[16] M.-R. Ra, J. Paek, A. B. Sharma, R. Govindan, M. H. Krieger, and M. J. Neely, "Energy-delay tradeoffs in smartphone applications," in *Proceedings of the 8th international conference on Mobile systems, applications, and services*, ser. MobiSys '10, 2010.

[17] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, ""using mobile phones to determine transportation modes," *ACM Transactions on Sensor Networks*, vol. 6, no. 2, pp. 1–27, 2010.

[18] B. Longstaff, S. Reddy, and D. Estrin, "Improving activity classification for health applications on mobile devices using active and semi-supervised learning," in *ICST Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health)*, March 2010.

[19] S. Reddy, K. Shilton, G. Denisov, C. Cenizal, D. Estrin, and M. Srivastava, "Biketastic: Sensing and mapping for better biking," in *ACM Conference on Human Factors in Computing Systems (CHI)*, April 2010.

[20] S. Reddy, D. Estrin, and M. Srivastava, "Recruitment framework for participatory sensing data collections," in *International Conference on Pervasive Computing (Pervasive)*, May 2010.

[21] S. Reddy, K. Shilton, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Using context annotated mobility profiles to recruit data collectors in participatory sensing," in *4th International Symposium on Location and Context Awareness*, May 2009.

[22] S. Reddy, D. Estrin, M. Hansen, and M. Srivastava, "Examining micro-payments for participatory sensing data collections," in *International Conference on Ubiquitous Computing (Ubicomp)*, September 2010.

[23] J.-H. Huang, S. Amjad, and S. Mishra, "Cenwits: a sensor-based loosely coupled search and rescue system using witnesses," in *Proc. of SenSys*, 2005, pp. 180–191.

[24] R. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. Abdelzaher, "Greengps: A participatory sensing fuel-efficient maps application," in *Mobisys*, 2010.

[25] R. Ganti, N. Pham, Y.-E. Tsai, and T. Abdelzaher, "Poolview: Stream privacy for grassroots participatory sensing," in *ACM Sensys*, Raleigh, NC, November 2008.

[26] N. Pham, R. Ganti, M. Y. Uddin, S. Nath, and T. Abdelzaher, "Privacy-preserving reconstruction of multidimensional data maps in vehicular participatory sensing," in *EWSN*, Coimbra, Portugal, February 2010.

[27] H. Ahmadi, N. Pham, R. Ganti, T. Abdelzaher, S. Nath, and J. Han, "Privacy-aware regression modeling of participatory sensing data," in *ACM Sensys*, Zurich, Switzerland, November 2010.

[28] H. Ahmadi, T. Abdelzaher, J. Han, R. Ganti, and N. Pham, "On reliable modeling of open cyber-physical systems and its application to green transportation," in *ICCPS*, Chicago, IL, April 2011.

[29] V. Srinivasan, J. Stankovic, and K. Whitehouse, "Protecting your daily in-home activity information from a wireless snooping attack," in *Ubicomp*, April 2008.

[30] C. Dwork, "Differential privacy: a survey of results," in *Proc. of 5th international conference on Theory and applications of models of computation (TAMC'08)*. Xi'an, China: Springer-Verlag, 2008, pp. 1–19.

[31] B.-C. Chen, D. Kifer, K. LeFevre, and A. Machanavajjhala, "Privacy-preserving data publishing," *Foundations and Trends in Databases*, vol. 2, pp. 1–167, 2009.

[32] V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in *SIGMOD'10*, 2010.

[33] A. Khrabrov, G. Stocco, and G. Cybenko, "Exploratory community sensing in social networks," in *Proceedings of SPIE*, vol. 7693, 2010, p. 76931G.

[34] J. Xie, S. Sreenivasan, G. Korniss, W. Zhang, C. Lim, and B. Szymanski, "Social consensus through the influence of committed minorities," *Physical Review E*, vol. 83, no. 7, 2011.

[35] W. Zhang, C. Lim, S. Sreenivasan, J. Xie, B. Szymanski, and G. Korniss, "Social Influencing and Associated Random Walk Models: Asymptotic Consensus Times on the Complete Graph," *Chaos*, vol. 44, no. 6, 2011.

[36] M. Newman, "Models of the small world," *Journal of Statistical Physics*, vol. 101, no. 3, pp. 819–841, 2000.

[37] D. Watts, "Networks, dynamics, and the small-world phenomenon," *American Journal of Sociology*, vol. 105, no. 2, pp. 493–527, 1999.

[38] M. Newman, "The structure of scientific collaboration networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 2, p. 404, 2001.

[39] R. Albert and A. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, pp. 47–97, 2002.

[40] S. Delre, W. Jager, and M. Janssen, "Diffusion dynamics in small-world networks with heterogeneous consumers," *Computational & Mathematical Organization Theory*, vol. 13, no. 2, pp. 185–202, 2007.

[41] T. Berger-Wolf and J. Saia, "A framework for analysis of dynamic social networks," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 2006, pp. 523–528.

[42] C. Cortes, D. Pregibon, and C. Volinsky, "Computational methods for dynamic graphs," *Journal of Computational and Graphical Statistics*, vol. 12, no. 4, pp. 950–970, 2003.

[43] M. Lahiri, A. Maiya, R. Sulo, and T. Wolf, "The Impact of Structural Changes on Predictions of Diffusion in Networks," in *Data Mining Workshops, 2008. ICDMW'08. IEEE International Conference on.* IEEE, 2008, pp. 939–948.

[44] M. Goldberg, S. Kelley, M. Magdon-Ismail, K. Mertsalov, and W. Wallace, "Communication dynamics of blog networks," *Advances in Social Network Mining and Analysis*, pp. 36–54, 2010.

[45] C. Hui, M. Goldberg, M. Magdon-Ismail, and W. Wallace, "Simulating the Diffusion of Information: An Agent-Based Modeling Approach," *International Journal of Agent Technologies and Systems (IJATS)*, vol. 2, no. 3, pp. 31–46, 2010.

[46] A. Pentland, "To Signal Is Human," *American Scientist*, vol. 98, no. 3, pp. 204–210, 2010.

[47] S. Adali, R. Escriva, M. Goldberg, M. Hayvanovych, M. Magdon-Ismail, B. Szymanski, W. W.A., and G. Williams, "Measuring behavioral trust in social networks," in *Intelligence and Security Informatics (ISI), 2010 IEEE International Conference on.* IEEE, 2010, pp. 150–152.

[48] M. Newman, "Modularity, Community Structure, and Spectral Properties of Networks," *Physical Review E*, vol. 67, p. 026126, 2003.

[49] F. Harary, "A criterion for unanimity in French's theory of social power," *Studies in social power*, pp. 168–182, 1959.

[50] N. Friedkin and E. Johnsen, "Social influence and opinions," *The Journal of Mathematical Sociology*, vol. 15, no. 3, pp. 193–206, 1990.

[51] T. Schelling, "Thermostats, Lemons, and Other Families of Models," *Micromotives and Macrobehavior*, pp. 81–133, 1978.

[52] C. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics," *Reviews of modern physics*, vol. 81, no. 2, pp. 591–646, 2009.

[53] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 2003, pp. 137–146.

[54] S. Moscovici, E. Lage, and M. Naffrechoux, "Influence of a consistent minority on the responses of a majority in a color perception task," *Sociometry*, vol. 32, no. 4, pp. 365–380, 1969.

[55] X. Castelló, A. Baronchelli, and V. Loreto, "Consensus and ordering in language dynamics," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 71, no. 4, pp. 557–564, 2009.

[56] L. Steels, "A self-organizing spatial vocabulary," *Artificial life*, vol. 2, no. 3, pp. 319–332, 1995.

[57] A. Baronchelli, M. Felici, V. Loreto, E. Caglioti, and L. Steels, "Sharp transition towards shared vocabularies in multi-agent systems," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2006, p. P06014, 2006.

[58] L. Dall'Asta, A. Baronchelli, A. Barrat, and V. Loreto, "Nonequilibrium dynamics of language games on complex networks," *Physical Review E*, vol. 74, no. 3, p. 36105, 2006.

[59] Q. Lu, G. Korniss, , and B. Szymanski, "Naming Games in Two-Dimensional and Small-World-Connected Random Geometric Networks," *Physical Review E*, vol. 77, no. 1, p. 016111, 2008.

[60] Q. Lu, G. Korniss, and B. Szymanski, "The Naming Game in social networks: community formation and consensus engineering," *Journal of Economic Interaction and Coordination*, vol. 4, no. 2, pp. 221–235, 2009.