Sensing the 'Health State' of our Society

Anmol Madan[1], Manuel Cebrian[1], Sai Moturu[1],
Katayoun Farrahi[2], and Alex 'Sandy' Pentland[1]

[1] MIT Media Laboratory, Cambridge, USA
[2] IDIAP and EPFL, Martigny and Lausanne, Switzerland

May 1, 2011

Abstract

Mobile phones are a pervasive platform for opportunistic sensing of behaviors and opinions. We show that location and communication sensors can be used to model individual symptoms, long-term health outcomes, and diffusion of opinions in society. For individuals, phone-based features can be used to predict changes in health, such as common colds, influenza, and stress, and automatically identify symptomatic days. For longer-term health outcomes such as obesity, we find that weight changes of participants are correlated with exposure to peers who gained weight in the same period, which is in direct contrast to currently accepted theories of social contagion. Finally, as a proxy for understanding health education we examine change in political opinions during the 2008 US presidential election campaign. We discover dynamic patterns of homophily and use topic models (Latent Dirchlet Allocation) to understand the link between specific behaviors and changes in political opinions.

1 Introduction

Opportunistic sensing of a variety of behaviors and opinions in natural settings can be achieved via the pervasive platform now provided by mobile phones. By harnessing this technology's potential, we can better understand and improve the functioning of our societies, as well as inform both individual and collective action. The 'health state' of our society functions at many different levels, starting from the physical health of citizens, moving to longer-term health norms and outcomes, as well as in the arena of positive behavior and opinion change.

In order to sustain a healthy society at the most fundamental level, it is important to maintain the physical health of its individual citizens, e.g., limit the spread of contagious diseases, stress, and mental health conditions. Our first study depicts the use of mobility and behavioral data to help understand the link between behaviors and symptoms at an individual level. We demonstrate that there are characteristic behavior changes for individuals suffering from common colds, influenza, and stress, as well as the early onset of mental health conditions. These behavior changes can be used to identify symptomatic individuals using their mobile phone data alone, without the need for additional health reports.

At the level of an entire society, we need to understand behaviors that lead to poor long-term outcomes, such as obesity. Our second study describes the use of mobile phones to model and understand the link between exposure and weight gain among students during the course of a semester. Mobile sensing and modeling tools demonstrate much potential to shed more light on the ongoing debate in public health and uncover information about the causal basis for contagion of obesity and other health-related behaviors.

At the highest level, our society requires better mechanisms to diffuse new opinions and ideas within the population to induce positive behavior change. In our third study, we use mobile phones to measure the spread of opinions subsequent behavior change, using the 2008 US presidential election campaign as a proxy for studying a health education campaign. We discover dynamic patterns of homophily within communities, results not seen previously in political science literature. We demonstrate the use of topic models (Latent Dirchlet Allocation) to understand the link between specific behaviors and changes in political opinions, for both Democrats and Republicans.

Mobile phones can be used for sensing rich social interactions unobtrusively, precisely because the underlying sensing technologies are now commonplace and readily available. This technology's other significant advantage is eliminating the dependence on self-reporting, which is prevalent in traditional social sciences. Bluetooth proximity sensing opens new doors as we can quantify the time spent in face-to-face interactions for individuals, as opposed to relying on binary social ties. Interactions from electronic exchanges (calling records, sms logs, email headers) and contextual data (location information) enable us to answer important research questions in a wide range of fields, from behavioral epidemiology to public health.

Mobile sensor data have been used to understand a broad spectrum of sensing and modeling questions. Some recent examples include automatically inferring co-location and conversational networks [21], linking social diversity and economic progress [10], automatic activity and event classification for mass market phones [17], identifying transportation modes [19], as well as feedback tools for improving health and fitness [8] and for modeling human mobility patterns [13].

These studies, along with our current work, paint a vision for how such opportunistic sensing can drive the future of computational social science. This leads to an essential question: How can these technologies serve as an agent for societal change in real world settings? To this effect, we explain our current efforts to scale this approach by orders of magnitude, enabling every consumer with an Android mobile phone to participate. Individuals, researchers, as well as our society overall stand to benefit from such opportunistic sensing and modeling.

2 Experimental Approach and Dataset

The dataset used for the following studies was generated by deploying a smartphone application within a tight-knit student community at an undergraduate university. The data were collected from 70 participants over an entire academic year from a university residence hall. The mobile phone dataset consists of 3.15 million scans of Bluetooth devices, 3.63 million scans of WLAN access-points, 61,100 call data records, and 47,700 logged SMS events, all dated between September 2008 and June 2010. Of these, 2.08 million scanned bluetooth devices belong to other experiment participants, and 11,289 calls and 9,533 SMS messages were exchanged with other experiment participants. In addition, participants provided dependent labels via monthly self-report surveys related to their health habits, diet and exercise, weight changes, and political opinions during the presidential election campaign. For three months during the peak influenza period (February to April 2009), participants also provided 2,994 daily symptom reports related to common colds, fever, influenza, and mental health.

An important concern with such long-term user data collection is informed consent and securing personal privacy for the participants. This study was approved by the Institutional Review Board (IRB). As financial compensation for completing monthly surveys and using data-collection devices as their primary phones, participants were allowed to keep the devices at the end of the study. The sensing scripts used in the platform captured only hashed identifiers, and the collected data were stored on a secure server and coded before aggregate analysis.

3 Predictive Modeling of Individual Symptoms

Mobile phones, with their increasing sensing capabilities, provide an unprecedented opportunity to monitor useful behavioral and contextual information about users. While these data are currently being used primarily as individual pieces of information for specific applications, they present a much greater benefit collectively. Understanding the link between behaviors and symptoms in an epidemiological context is one such novel predictive application.

Proximity interactions are the primary mechanism for propagation of airborne contagious disease [18]. An important question in behavioral epidemiology and public health is to understand how individual behavior patterns are affected by physical and mental health symptoms. Such research requires continuous, long-term experimental data about symptom reports, mobility patterns, and social interactions amongst individuals. Epidemiologists currently do not have access to sensing and modeling capabilities to quantitatively measure behavioral changes experienced by symptomatic individuals in real-world scenarios [9]. Quantitatively understanding how people behave when they are infected would be a fundamental contribution to epidemiology and public health, and could inform treatment and intervention strategies, as well as aid public policy decisions. Compartmental epidemiological models (e.g., the Susceptible, Infectious, Recovered or SIR model) commonly assume that movement and interaction patterns for individuals are stationary during infection, i.e., that individuals will continue their typical behavioral patterns when sick. More recent epidemiological models accommodate reduced mobility variations to fit epidemic curves, but do so in a heuristic way due to lack of data at the individual level [3, 7, 11], thus potentially limiting their prediction accuracy during real-world epidemics.

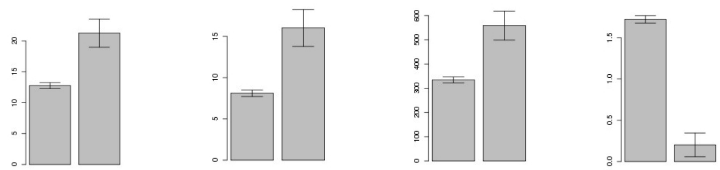
In our research, we use mobile phones as co-location and communication sensors to characterize the change in proximity, face-to-face interactions and individual trajectories in the contagion process. From mobile phone sensor data, we extract features that represent statistics of conversational partners and location, i.e., the total number of interactions, the diversity of interactions, and the diversity (entropy) of our behavior. These observational data have been shown to link important aspects of individual and collective behavior, such as friendships and individual job satisfaction. Entropy has been demonstrated to be a good indicator for the predictability of an individual's behavior.

The sensor features used are total communication (phone calls and SMS exchanged), late night and early morning communication (i.e., communication between 10pm and 9am on weekdays), communication diversity (i.e., unique individuals reflected in phone and SMS interactions), physical proximity entropy with other participants (i.e., the entropy of distribution of Bluetooth proximity with other participants), physical proximity entropy with other participants during late night and early morning, physical proximity entropy for bluetooth devices excluding experimental participants, and WLAN entropy based on university WLAN APs (i.e., the distribution of WLAN access points scanned within the given period and WLAN Entropy based on external WLAN APs).

The behavior changes reflected in these mobile features, corresponding to different conditions are shown in Figure 1. With common colds, we find that total communication and late-night, early-morning communication increases. For sore-throat and cough reported symptoms, we find that bluetooth entropy with respect to other dorm residents increases. For fever and CDC-defined influenza, we find that both WLAN entropy with respect to university access points and external access points show a dramatic decrease. For often-stressed and sad-lonely-depressed responses, participants show isolated behavior on symptomatic days, i.e., total communication, late-night communication, communication diversity and late-night bluetooth entropy decrease. All these examples work to illustrate the tremendous potential of mobile phones to monitor the health status of an individual in almost real-time.

In light of these characteristic behavioral changes associated with respiratory symptoms, fever, influenza, stress and depression, it is possible to devise a classification scheme from behavioral features alone that identifies when individuals are likely to be symptomatic. For example, assume that the user has a mobile sensing application installed on their personal phone. When this application detects uncharacteristic variations in behavior, it could predict the likelihood that the user is infected with a known symptom and potentially inform a nurse, family member, or healthcare professional. Such proactive healthcare is critical for conditions with high risk of patient under-reporting (e.g., mental health, elderly healthcare). Self-reported symptoms are correlated in our data (e.g., on days with a sore-throat-cough, a person is likely to also have a runny-nose). Due to unbalanced class sizes, we use a Bayesian-network classifier with MetaCost, a mechanism for making classifiers cost-sensitive to identify the symptomatic days. Structure learning for the network is performed using K2 hill climbing and the results are based on 4-fold cross-validation. Recall for different conditions ranges from 60% to 90% for the symptom class, and is considerably better than chance.

There is also extensive medical and health policy interest in understanding the temporal order between behavior change, stress and physical symptoms. We use the Phase Slope Index method to gain insight into the temporal relationships between signals of behavior features, stress and physical symptoms. This method is based on the result that the phase slope of the cross-spectrum of two signals can be used to estimate information flux between these signals in the time domain. Independent noise

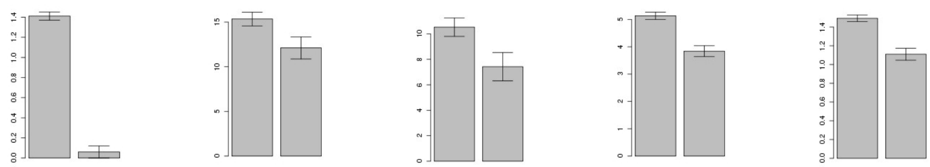


(a) Total communication increases for Runny Nose responses ***

(b) Late-night early morning communication increases Runny Nose responses **

(c) Bluetooth entropy with respect to other dorm residents increases for Sore Throat and Cough responses ***

(d) WLAN based entropy with respect to university WLAN APs decreases for Fever responses***



(e) WLAN Entropy with respect to external WLAN APs decreases for fever responses ***

(f) Total communication decreases for sad-lonely-depressed responses *

(g) Late-night early morning communication decreases sad-lonely-depressed responses *

(h) Communication diversity decreases for often-stressed responses **

(i) Late night early morning Bluetooth entropy with other experiment participants reduces for often-stressed responses **

Figure 1: Illustrative behavior changes with runny nose, cough and sore throat, fever and influenza, sad-lonely-depressed and often-stressed symptoms. For each graph, the left bar shows the normal, healthy behavior, behavior, and the right bar shows the behavior on symptomatic days. *: $p < 0.05$ **: $p < 0.01$ ***: $p < 0.001$

mixing does not affect the complex part of the coherency between multivariate spectra, and hence PSI is considered more noise immune than Granger and other less sophisticated analyses. The twelve largest PSI coefficients are illustrated in Figure 2. An example insight is that 'often-stressed' is useful in forecasting proximity, communication and WLAN behaviors, which suggests that an individuals cell phone recognizes and reports that the individual is stressed before it is reflected in its behavior. Another insight is that in two cases Bluetooth interaction features are used to forecast WLAN features, suggesting that a behavior change is reflected in personal interactions with others before it is reflected in the movement patterns of the individual.

While these results are specific to a few conditions, they clearly point toward a greater set of opportunities for managing an individual's health status. For example, how can such behavioral data be used to predict changes in mental health status, or detect when users are not getting adequate social support and introduce automatic interventions? How can data aggregated across many users be used for early detection of infectious disease outbreaks? Our findings lead us to believe that it be possible to answer such questions in the near future, and to begin planning how to influence the development of even greater health sensing capabilities in smart-phones.

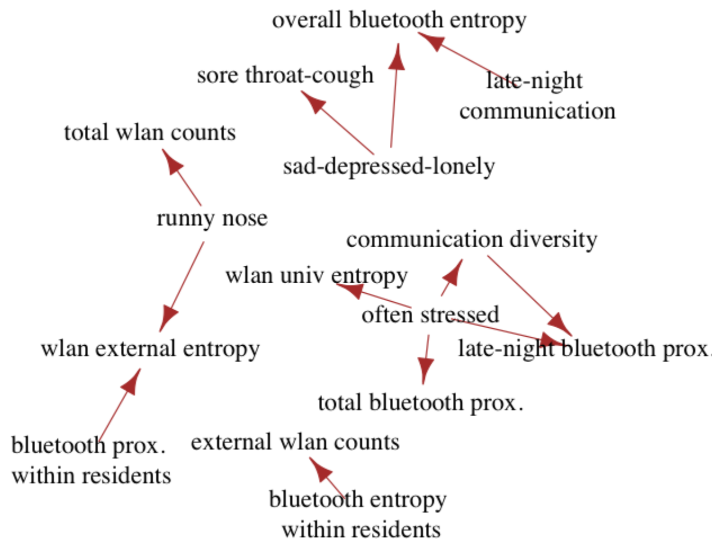


Figure 2: Highest-ranked PSI relationships across mobile behavioral features, physical symptoms and mental symptoms. Directed ties represent temporal flux.

4 Exposure and Long-term Health Outcomes

On the longer-term scale, it is important in our society to understand the social mechanisms leading to the adoption of unhealthy normative behaviors, such as obesity. The effects of social exposure on behavioral adoption, change, diffusion and outcomes has been the focus of several research studies in the past. Such work, however, was typically dependent on surveys and experience sampling to gather social interaction information. Mobile phones, through proximity and communication sensing, allow us to quantify social interactions with much higher resolution, accuracy and continuity than these traditional approaches. This allows us to better determine the important role that social exposure plays in our everyday life.

Recent work in public health suggests that obesity and other health-related behaviors spread through long-term social networks. This result, however, is widely debated, in part because it is challenging to disambiguate homophily from causal social contagion. We use pervasive sensing to study weight changes in a real-world social community, by comparing quantitative interactions among participants with their weight changes. We consider two interaction modalities: bluetooth proximity and phone communication. Bluetooth based features are calculated from co-location events, where at least one mobile device detects another in proximity. As long as the devices continue to detect each other, the two people are assumed to be in physical proximity. Call and SMS based features are based on call data records that include information about call duration. In our analysis, these mobile features are compared with self-reported close friend and social acquaintance ties in the community. We find that bluetooth based features outperform phone communication features and interaction in the context of explaining changes in health-related outcomes.

Interpersonal interactions can be broken down along both time of day and day of week. For example, the people that a person interacts with in the daytime (e.g., in the classroom, at lunch) are different from the people that he/she interacts with late at night or early in the morning (e.g., close friend, significant other). Similarly, activities and interactions on weekdays are different from those on the weekends (e.g., grocery shopping, entertainment) as compared to working weekdays. In the selection of features, we consider total interaction, weekday versus weekend interaction, and peak versus off-peak interaction (where 'off-peak' represents interactions between 9pm and 9am every day).

These different interaction types can be further conditioned upon the health-state and health behaviors of social contacts. We consider interactions with contacts who gained substantial weight in the same period, e.g., four pounds or more, and interactions with a broader set of participants, defined as those who did not gain weight, i.e., gained no more than one pound or lost weight. Interactions with others are expressed as number of unique contacts (also known as alters, in social network nomenclature), number of contact events, and time spent in seconds. Since participants reside in the same community, there is a very high probability that an individual has interacted with everyone else in the community on at least a few occasions. To differentiate between encounters that happen by chance as part of day-to-day living versus those that are intentional, a ranking function is used, wherein interactions that are greater than the 25^{th} percentile are weighted twice as high.

Body Mass Index (or BMI) is a common metric for healthy body weight and is given as the mass in kilograms divided by the square of an individual's height in meters. Individuals with a BMI of 30 or over are categorized as obese while those who have a BMI between 25 and 30 are considered overweight. We use an individual's change in BMI during the observation period as a dependent variable and estimate the influence of various exposure-based independent variables using linear regression.

We define three different metrics to measure users' exposure to different health-related behaviors– Alters, Events and Times. The Alters score is the number of alters who show the specific behavior or outcome. It is equal to the number of unique alters (with the specific health-related outcome or behavior) with whom the participant interacts, normalized by the total number of alters with whom the participant interacts. The Events score measures exposure to different health-related behaviors as total counts (e.g., total bluetooth detections) of interaction with others who have the specific opinion or outcome. The Times score measures exposure to different health-related opinions in terms of time spent (e.g., in seconds in bluetooth proximity), with others who have the specific opinion or outcome. Similar to Alters score, both Events and Times scores are calculated by normalizing the absolute value for a particular condition (e.g., 'time spent with people who eat unhealthy'), by total value for all conditions (e.g., 'time spent with all alters').

The Alters feature is the best-performer amongst these, explaining 38% of the variance in the dependent BMI variable. This result is somewhat surprising since it is also the least complex feature used in analysis. The Events and Times scores also show significant correlations, with the Times score being more predictive of the two (Adj. R^2 of 0.1 and 0.24 respectively)

We use the interaction-based features described above to model the changes in the participants' BMI. The Alters score, Events score and Times score features are broken down by temporal structure, e.g., weekend, off-peak, etc. and also by the relative size of weight gains. We consider two conditions: (i) exposure to peers who have not lost weight over the semester, and (ii) exposure to peers who have gained more than four pounds in weight over the semester.

Overall, significant correlations are observed with as much as nearly 38% of the variation in the dependent variable being explained by the bluetooth exposure features. This is an important result because it demonstrates that the change in an individual's BMI can be explained by face-to-face exposure to contacts who themselves gained weight. Another interesting result is that when exposure to peers who lost weight is considered, no significant correlations are observed. A possible explanation is that the mechanism for adoption of weight-gain related behaviors differs from the mechanism for adoption of weight-loss related behaviors.

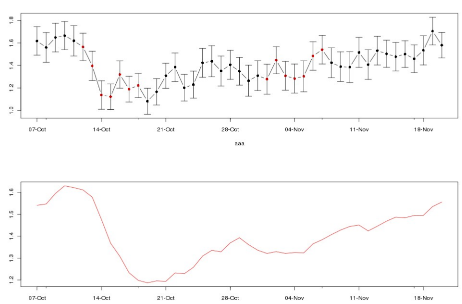
Furthermore, when exposure to self-identified close friends and social acquaintances is considered, no significant correlations are observed. This could be because face-to-face interactions play an important role, irrespective of whether the contacts are close friends or not. This is a significant departure from the findings using data from the Framingham Heart Study, which may be because of the substantial differences in both the length and population sampling characteristics of the two studies.

This study illustrates the role of social exposure in the context of longer-term health related behaviors such as obesity, and may extend to situations ranging from mood contagion, to influencing purchasing behaviors, to affecting sleep loss or job productivity. Insights gained from such studies could improve long-term behaviors and outcomes and optimize performance. Such insights can also be used to provide context-sensitive cues and nudges to help guide and persuade users in such tasks, thus completing a feedback loop that serves to empower users far beyond what is possible today.

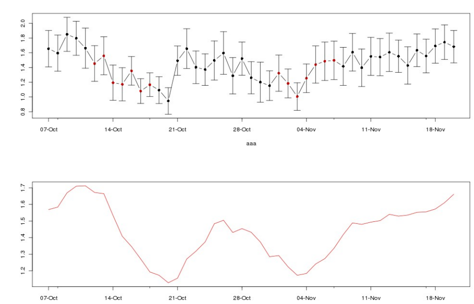
5 Modeling Opinion Change

A vital element in motivating collective action to improve the 'health state' of our society is diffusing positive opinion change through communities. This requires taking into account the rich identities and interactions among individuals, using both social and contextual information. We describe the use of mobile phone sensors to measure and model the face-to-face interactions and subsequent opinion changes amongst undergraduates during the 2008 US presidential election campaign, as a proxy for studying a large-scale health campaign. Mobile features can be used to estimate unique individual exposure to different opinions and we find patterns of dynamic homophily related to external political events, such as election debates and election day. To our knowledge, this is the first time such dynamic homophily effects have been reported.

Homophily, or the idea of "birds of a feather flock together," [14] is a fundamental and pervasive phenomenon in social networks and refers to the tendency of individuals to form relationships with others that have similar attributes. We propose a measure of dynamic homophily based on mobile exposure features that captures the difference between an individual's opinions and the opinions he/she is exposed to. Daily variations in this measure are due to changes in mobile phone interaction features that capture how participants interact with others. This daily measure captures dynamic homophily variations during the election period, an effect not captured using traditional homophily measures. For a few days around the election day and final debates, participants show a higher tendency overall to interact with like-minded individuals.



(a) Dynamic homophily of **political interest responses** (using bluetooth proximity) for all participants. Notice the decline, i.e. tendency to interact with others having similar opinions, lasting for a few days, around Oct 15th 2008, which was the last presidential debate.



(b) Dynamic homophily of **political interest responses** (using bluetooth proximity) only for **Freshmen** . There are two periods of decline, each lasting for a few days. The first is around Oct 15th (last presidential debate) and the second is around 4th Nov, Election Day.

Figure 3: Top: actual values of measured dynamic homophily with standard error bars. Bottom: Moving average.

Exposure based features described above can be used as a feature to train a linear predictor of future opinions. The coefficients used in a linear model of opinion change include normalized exposure during the period, the person's opinion at the start of the study (September 2008), and a constant term that represents a linearly increasing amount of media influence as we get closer to the election date (Nov. 2008). For the various political opinion questions, regression values are in the $R^2 = 0.7$ to $R^2 = 0.8, p < 0.01$ region. Using exposure based features explains an additional 15% - 30% variance across different political opinion questions. The exposure effects for freshmen are approximately twice as strong as compared to the entire population, similar to the variations in dynamic homophily related to external events.

It is important to understand what influences opinion change– is there an underlying mechanism resulting in the opinion change for some people? Can we measure this mechanism, and if so, can we predict future opinion changes from observed behavior? We propose a method for activity modeling based on the Latent Dirichlet Allocation (LDA) [2] topic model to contrast the activities of participants that change opinions with those that do not. In an unsupervised manner, we discover the dominating routines of people in the dataset, where routines are defined as the most frequently co-occurring political opinion exposure patterns (also referred to as topics).

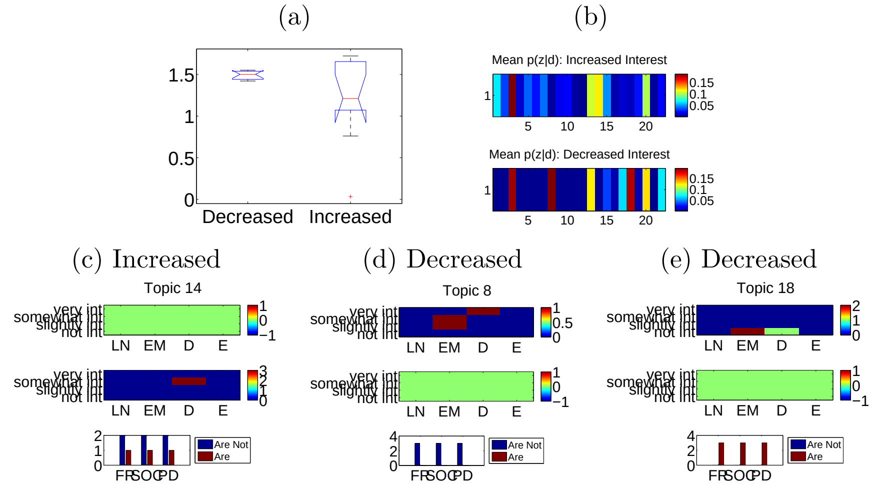


Figure 4: Routines of people who increased their interest in politics versus those that decreased their interest. (a)T-test results reveal the difference in the entropy of topic distributions for these groups is statistically significant. (b)Mean distribution of topics for users of both groups. (c)-(e)Topics which best characterized users' daily life patterns in both groups. People who increased their interest often communicated by phone (c) and those that decreased interest had many face-to-face interactions with people with little/no interest in politics (d-e). For a given topic ((c)-(e)), we display the 3 most probable words' (top) face-to-face interaction features (middle) phone interaction features and (bottom) relationship statistics, abbreviated by FR for friends, SOC for socialize and PD for political discussants.

This approach can be described with a specific example. We formulate a multimodal vector of exposure features (MME features) encompassing four components: (1) time; (2) political opinion; (3) type + amount of interaction; and (4) self-reported relationship. Overall, a MME feature captures the exposure to a particular political opinion, including details such as time and relationship with the people the person is exposed to. We consider four time components in these experiments as follows: 10pm-2am (late night = LN), 2-8am (early morning = EM), 8am-5pm (day = D), 5-10pm (evening = E). The political opinion component captures the political opinion survey questionnaire response of the user for a given opinion. For example, if a user is asked about their interest in politics, the possible feature could contain responses such as 'very interested.' Component (3) is the type and amount of interaction, where the type can be either face-to-face measured by Bluetooth proximity of phone communication measured by both incoming and outgoing call and SMS logs. Finally, the relationship metric is defined by $f \in$ [friend, not friend], $s \in$ [socialize, do not socialize], and $p_d \in$ [political discussants, not political discussants]. Topics, discovered by Latent Dirichlet Allocation [2], are essentially clusters of dominating 'opinion exposures' present over all individuals and days in the real-life data collection and described in terms of MME features. Topics z_k are characterized by $p(w|z_k)$ (probability distribution of MME features for the latent topic) and $p(z_k|d)$ (probability distribution of users for the latent topic). More details of the formulation can be found in [16].

The goal is to determine the difference between the interaction patterns of two groups, or in other words, to distinguish users that change opinion from users that do not. We do this by comparing the most probable topics, representing users' opinion exposure patterns, averaged over all the users of each group. In Figure 4 we present results for the 'interest in politics' opinion for 22 topics since the difference between the entropy of topic distributions is statistically significant as seen by Figure 4(a). Figure 4(b) is the mean probability distribution of topics given the users from the two groups. The mean topic distribution $p(z|d)$ is shown for (top) all users that increased their interest, and (bottom) all users that decreased their interest. Plots (c)-(e) show the most probable words for the dominating topics in both groups. Topic 14 (c) is highly probable for users that increased their interest. Topic 8 and 18 are highly probable for users that decreased their interest. Results reveal that people who increased interest were communicating often by phone during the day. People who decreased their interest had only face-to-face interactions (no phone communication) dominating their daily routines and it included interaction with people who showed little and no interest as seen by topics 8 and 18.

We repeat this approach for other political opinions, e.g., the 'preferred party' opinion reported by individuals in monthly survey responses. We observed the difference in the entropy of topic distributions of the groups 'people who changed preferred party' and 'people who did not' is statistically significant with $p = 0.01$. We find that users that changed preferred party to become Democrats predominantly had face-to-face interactions with political discussants that were non-friends and not people they socialize with (for clarification, 13 participants became more democratic, and 6 participants became more independent during this entire period). The preferred party of these political discussants was Democrat and this interaction occurred predominantly between 10pm-5pm. Furthermore, people who changed preferred party also showed heavy phone call and SMS activity with other Democrats and Independents, but not with Republicans.

These results demonstrate that we might indeed be able to detect a dynamic homophily effect around a specific context, political events in this case. More interestingly, such information could enable us to create a rich identity for each individual by mapping out a more dynamic and accurate social graph. Such information could be useful in creating better services and applications for users. Businesses may be able to use this information to reach and serve the needs of their customers better. It would greatly enhance the scope for social and behavioral research. Most importantly, such data could be used to help users be more socially engaged, leading to greater wellness and life satisfaction.

6 Discussion: Opportunistic Sensing for Every Mobile Phone

We have shown that mobile phone-based data for an undergraduate community provides insight into their behaviors and opinions. This opens new possibilities, specifically, questions about how we can scale this approach and deploy 'health state' tools more broadly. From the human-computer interaction perspective, it is important to ask whether such aggregate analysis brings value to the individual user and if a behavior feedback loop leads to positive behavior change.

To answer these scalability questions, we have devised a new data collection and feedback platform for commodity Android mobile phones. This platform has been designed with two goals in mind. In the vein of the quantified-self community, this tool provides users objective data and insights about their behaviors and life. Secondly, it provides researchers with behavioral data from a much-larger pool of participants, symptoms and health conditions. In addition to reporting detailed movement and communication statistics, we also provide users with an easy-to-parse summary via an avatar, predictions of daily stress levels, annotation and correlation of self-reported symptom and treatment events to activity and communication features, as well as aggregate comparisons relative to their peers. Apart from being a broader tool for the quantified-self community, this platform is being adopted by patients managing chronic conditions such as chronic IBS, bi-polar disorder, cystic fibrosis, and congestive heart disease.

Personal privacy becomes even more important for such a widely-distributed mobile application. We propose a two-step stance on end-user privacy. Firstly, users own their personal data and may elect to share it or delete it permanently. Secondly, if users wish to compare their behavior to the aggregate, then they must be willing allow de-identified statistics of their behavior into the aggregate pool.

There are several future extensions of our work. The prediction models used could be improved by increasing their complexity and taking into account the Markov properties of human behavior. The statistical tests used to identify behavior changes for symptomatic individuals could be improved using a repeated-measures approach, along with accounting for confounding behavior changes due to external events, e.g., exams or the weather. Since all users do not carry mobile phones on them all the time, mobile phones are not perfect sensors of human behavior. Nevertheless, the opportunities for sensing and real-world data collection provided by these devices allows for the measurement and understanding of human phenomena that would not be possible otherwise.

Our understanding of aggregate behavior will benefit from the continuous evolution of mobile sensing technologies. For example, our mobile research platform did not support Bluetooth signal strength, which could provide a better measure of physical proximity. In our Android platform, WLAN-based location sensing has been replaced with more accurate GPS-based location sensing. With such improvements in sensing technology accompanied by the escalating use of advanced mobile phones, we hope that our work will guide the next generation of mobile phone-based studies and real-world applications.

7 Acknowledgements

We would like to thank David Lazer for help designing the political questionnaires, Devon Brewer for helping design the daily health questionnaires, Daniel Gatica-Perez for his role as advisor to Ms. Farrahi, Iolanthe Chronis and various undergraduate contributors for their help with the experimental deployment and analysis. Anmol Madan and Alex Pentland were partially sponsored by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053, and by AFOSR under Award Number FA9550-10-1-0122. Views and conclusions in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ARL, AFOSR, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation. Katayoun Farrahi was supported by the Swiss National Science Foundation through the MULTI project. Sai Moturu was partially sponsored by an AFOSR award to Prof. Huan Liu at Arizona State University under Award Number FA9550-08-1-0132.

References

[1] S. Aral, L. Muchnik, and A. Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544, 2009.

[2] D. M. Blei, A. Y. Ng, M. I. Jordan, and J. Lafferty. Latent dirichlet allocation. *JMLR*, 3:2003, 2003.

[3] D. Brockmann. Human mobility and spatial disease dynamics. *Review of Nonlinear Dynamics and Complexity*, 2009.

[4] R. Burt. Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology*, 1987.

[5] N. Christakis and J. Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370, 2007.

[6] E. Cohen-Cole and J. Fletcher. Detecting implausible social network effects in acne, height, and headaches: longitudinal analysis. *British Medical Journal*, 337:a2533, 2008.

[7] V. Colizza, A. Barrat, M. Barthelemy, A. Valleron, and A. Vespignani. Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions. *PLoS Medicine*, 4(1):95, 2007.

[8] S. Consolvo, D. McDonald, T. Toscos, M. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, et al. Activity sensing in the wild: a field trial of ubifit garden. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1797–1806. ACM, 2008.

[9] D. M. Musher. How Contagious are Common Respiratory Track Infections. *New England Journal of Medicine*, 2003.

[10] N. Eagle, M. Macy, and R. Claxton. Network diversity and economic development. *Science*, 328(5981):1029, 2010.

[11] J. Epstein, D. Goedecke, F. Yu, D. Morris, R.J.and Wagener, and G. Bobashev. Controlling pandemic flu: the value of international airtravel restrictions. *PLoS One*, 2(5), 2007.

[12] Friedkin N.E. *A Structural Theory of Social Influence*. Cambridge University Press, 1998.

[13] M. Gonzalez, C. Hidalgo, and A.-L. Barabasi. Understanding Individual Human Mobility Patterns. *Nature*, 453:779–782, 2008.

[14] P. Lazarsfeld and R. K. Merton. Friendship as a Social Process: A Substantive and Methodological Analysis. *Freedom and Control in Modern Society*, 1954.

[15] R. Lyons. The Spread of Evidence-Poor Medicine via Flawed Social-Network Analysis. *Arxiv*, page arXiv:1007.2876, 2010.

[16] A. Madan, K. Farrahi, D. Gatica-Perez, and A. Pentland. Pervasive sensing to model political opinions in face-to-face networks. In *Pervasive*, San Francisco, USA, 2011.

[17] E. Miluzzo, N. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. Eisenman, X. Zheng, and A. Campbell. Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 337–350. ACM, 2008.

[18] P. Elliott, et al. *Spatial Epidimiology*. Oxford University Press, 2000.

[19] 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 (TOSN)*, 6(2):1–27, 2010.

[20] World Health Organization Global strategy Report on Diet, Physical Activity and Health. http://www.who.int/dietphysicalactivity/publications/.

[21] D. Wyatt, T. Choudhury, J. Bilmes, and J. Kitts. Inferring colocation and conversation networks from privacy-sensitive audio with implications for computational social science. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(1):7, 2011.