Social Sensing for Epidemiological Behavior Change

Anmol Madan, Manuel Cebrian, David Lazer[†] and Alex Pentland
MIT Media Lab and Harvard University[†]
Cambridge MA
anmol, cebrian, pentland@media.mit.edu; david_lazer@harvard.edu

ABSTRACT
An important question in behavioral epidemiology and public health is to understand how individual behavior is affected by illness and stress. Although changes in individual behavior are intertwined with contagion, epidemiologists today do not have sensing or modeling tools to quantitatively measure its effects in real-world conditions. In this paper, we propose a novel application of ubiquitous computing. We use mobile phone based co-location and communication sensing to measure characteristic behavior changes in symptomatic individuals, reflected in their total communication, interactions with respect to time of day (e.g., late night, early morning), diversity and entropy of face-to-face interactions and movement. Using these extracted mobile features, it is possible to predict the health status of an individual, without having actual health measurements from the subject. Finally, we estimate the temporal information flux and implied causality between physical symptoms, behavior and mental health.

Author Keywords
Social computing, Spatial Epidemiology, Mobile Sensing

ACM Classification Keywords
I.5.4 Pattern Recognition: Applications; H.4.m Information Systems: Miscellaneous

General Terms
Algorithms, Experimentation, Measurement.

INTRODUCTION
Face-to-face interactions are the primary mechanism for propagation of airborne contagious disease [28]. An important question in behavioral epidemiology and public health is to understand how individual behavior patterns are affected by physical and mental health symptoms. Epidemiologists currently do not have access to sensing and modeling capabilities to quantitatively measure behavioral changes experienced by symptomatic individuals in real-world scenarios [11]. Such research requires continious, long-term data about symptom reports, mobility patterns and social interactions amongst individuals. In this paper, we propose a novel application of ubiquitous computing, to better understand the link between physical respiratory symptoms, influenza, stress, mild depression and automatically captured behavioral features. This is an important problem for several reasons.

Quantitatively understanding how people behave when they are infected would be a fundamental contribution to epidemiology and public health, and can inform treatment and intervention strategies, as well as influence public policy decisions. On one hand, clinical epidemiology has accurate information on the evolution of the health of individuals over time but lacks realistic social interaction as well as spatio-temporal data [15]. On the other hand, current research efforts in theoretical epidemiology model the rate of infection in a population whose behavior is stationary over time and do not account for individual changes [26]. For instance, if a person infected with influenza continues his habitual lifestyle instead of isolating himself, he could pose a bigger risk to others in proximity. Based on our analysis and results, policymakers can recommend social interventions that minimize such risk.

On the modeling front, 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 to fit epidemic curves, but in a heuristic way due to lack of data at the individual level [4, 9, 14], which possibly limited their prediction accuracy during the 2009 H1N1 influenza epidemic [22]. To our knowledge, we provide the first quantitative results on this important measurement based on mobile sensing. Our results, described below, can be plugged into the SIR model by specifying the number and frequency of contacts that individuals will likely have when going from the S(usceptible) to the I(infected) state, and therefore improving prediction accuracy. Furthermore, predicting likelihood of symptoms from behavior could lead to a possible early-warning system and intervention by medical experts, with the associated savings in economic and human resources [7].

In this paper, we describe experimental work that illustrates the use of co-location and communication sensors in mobile phones to characterize the change in face-to-face inter-

UbiComp '10, Sep 26-Sep 29, 2010, Copenhagen, Denmark.

actions and individual trajectories in the contagion process. The experimental context consists of residents of an undergraduate residence hall for two months, from February to April 2009. Individuals were surveyed daily for symptoms of contagious diseases like common colds, influenza and gastroenteritis. We find that there are characteristic changes in behavior when individuals are sick, reflected in automatically captured features like their total communication, communication patterns with respect to time of day (e.g., late night, early morning), diversity of their network and entropy of movement within and outside the university, and these variations can be used to identify symptomatic days for individuals. Finally, we use a recently developed signal processing approach [17] to shed light on the temporal information flux between physical symptoms, behavior changes and mental health symptoms. This is, to our knowledge, the first comprehensive empirical study in this direction.

RELATED WORK

Mobile Phones as Social Sensors

The four billion mobile phones worldwide are ubiquitous social sensors of location, proximity and communication. Eagle and Pentland [13] coined the term Reality Mining, and used mobile phone Bluetooth proximity, call data records and cellular-tower identifiers to detect the social network structure and recognize regular patterns in daily user activity. For human location traces, Gonzalez et. al [19] showed that call detail records can be used to characterize temporal and spatial regularity in human mobility patterns better than random walk or Levy flight simulations. Similarly, electronic sensor badges like the Sociometric badge [25] have been used to identify human activity patterns and analyze conversational prosody features. Other examples of the use of mobile phones to map human interaction networks include the CENS and mHealth projects [1, 2].

In addition to mobile phones, web based and survey based data sources have also been used as social sensors in the healthcare space. Google Flu Trends uses aggregated Google search data to estimate current flu activity around the world in near real-time [18]. In [3], Barabasi reported how social networks impact the spread of obesity and pathogens including influenza, severe acute respiratory syndrome and human immunodeficiency virus. Finally, and more related to our project, Christakis and Fowler used archival data from the Framingham Heart Study to study how social interactions affect the spread of obesity, smoking behavior and mental health [5, 6, 16]. In summary, the emerging field of Computational Social Science [10] leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors, with an enormous potential for social health applications.

Link Between Physical Symptoms, Behavior Changes and Stress

Ubiquitous computing approaches may allow us to make a fundamental contribution to epidemiology, in understanding the link between physical symptoms, mental health symptoms, and their expression in social interactions and behavior. The intertwined relationships between these is not well understood, due to limitations of existing clinical diagnosis and public health tools, but plays an important role in medical detection, treatment and management of conditions.

In medical literature, substantial evidence has been found for an association between stress and increased illness behavior, and less convincing but provocative evidence was found for a similar association between stress and infectious pathology: Introverts, isolates, and persons lacking social skills may also be at increased risk for both illness behaviors and pathology [8]. Various medical conditions that involve activation of the immune system are associated with psychological and neuro-endocrine changes that resemble the characteristics of depression. Recent studies have presented empirical evidence on the relationship between the behavioral effects of immune activation and depressive symptomatology, characterized by reduced locomotor, exploratory, and social behavior [30].

The association between psychosocial stress and susceptibility to upper respiratory tract infection has also been investigated in people with a history of recurrent common colds and flu. Several dimensions of psychosocial stress, including exposure to stressful experiences, stress-prone personality traits, and signs of emotional disturbance have been investigated in people with a history of recurrent common colds and flu. Experts conjecture that stress depletes local immune protection, increasing susceptibility to colds and flu. Alternatively, psychological disturbances could develop in response to frequent illness [12].

METHODOLOGY

Several projects have used existing call data records and mobile operator location information to model movement patterns and social ties. Our approach is to build a mobile phone software platform for primary personal use by participants, as a tool for long-term data collection.

The dataset described below was collected as part of a longitudinal study with seventy residents of an undergraduate residence hall (referred to as an undergraduate dormitory in North America), that serves as the primary residential, cooking, social activity and sleeping quarters for the residents. This residence hall was the smallest undergraduate dormitory at the university. The participants in the study represent eighty-percent of the total population of this hall, and most of the remaining twenty-percent are spatially isolated. The dormitory is known within the university for its pro-technology orientation and the decision of students to reside within the dorm is determined by self-selection by both students and the existing residents. The students were distributed roughly equally across all four academic years (freshmen, sophomores, juniors, seniors), about 54% of the students were male, and predominantly engineering, mathematics and science majors. The study participants also included four graduate resident tutors that supervised each floor. The participants used these data collection Windows Mobile devices as their primary phones, with their existing voice plans. Students had data access on these phones due to pervasive wifi on the university campus and in the metropolitan

Table 1. Symptom Survey Questionnaire. All questions were Yes/No responses

Survey Question (as shown on mobile phone)
Do you have a sore throat or cough?
Do you have a runny nose, congestion or sneezing?
Do you have a fever?
Have you had any vomiting, nausea or diarrhea?
Have you been feeling sad, lonely or depressed lately?
Have you been feeling stressed out lately?

area. As compensation for their participation for the entire academic year, they were allowed to keep the smart phones at the end of the experiment. In this paper however, we describe the mobile platform, dataset and analysis related to measuring the spread of influenza, common colds and stress in this community over two months. Additional information about the broader experiment is available here [21].

The dataset used in this analysis was collected from two different sources. Social interaction data from mobile phones was collected via call data records, SMS logs, Bluetooth co-location sensing and WLAN-based location sensing. Intuitively, these sensors reflect pairwise communication and face-to-face proximity, intensity and nature of social ties, the homogeneity of behaviors across individuals, and the dynamics of network structure. The use of these particular sensors to understand human behavior is not novel in itself– however to our knowledge this is the first time they have been used to understand epidemiological behavior change.

Symptom data was collected using a daily-self report survey instrument, designed by an experienced epidemiologist. The survey instrument consisted of six questions with yes/no responses, which are listed in Table 1. The working of the survey launcher mobile application is described in more detail in the next section. In addition, a baseline survey instrument asked participants whether they had been immunized with flu vaccinations (flu-shots or flu-mist sprays), as well as their international travel before the start of the experiment. Of a total of 69 participants that completed the baseline survey, 20 reported that they had been immunized for influenza, via either a flu-shot or flu-mist spray.

User Privacy Considerations

A key concern with such long-term user data collection approaches is securing personal privacy for 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 capture only hashed identifiers, and collected data is secured and anonymized before being used for aggregate analysis. To minimize missing data from daily symptom reports, participants were compensated $1 per day that they completed the on-device symptom survey.

MOBILE SENSING PLATFORM

With the above goals in consideration, the mobile phone based platform for data-collection was designed with the following long-term continuous sensing capabilities, based on Windows Mobile 6.x devices. Daily captured mobile sensing data was stored on-device on read/write SD Card memory. On the server side, these logs files were merged, parsed and synced by an extensive Python post-processing infrastructure, and stored in MySQL for analysis. This sensing software platform for Windows Mobile 6.x has been released under the LGPLv3 open source license for public use [24].

Proximity Detection (Bluetooth)

The software scanned for Bluetooth wireless devices in proximity every 6 minutes (a compromise between sensing short-term social interactions and battery life, [13]). The Windows Mobile phones used in our experiment were equipped with class 2 Bluetooth radio transceivers, with practical indoor sensing range of approximately 10 feet. Scan results for two devices in proximity have a high likelihood of being asymmetric, which is accounted for in our analysis. Due to API limitations of Windows Mobile 6.x, signal strength was not available during scans.

Approximate Location (802.11 WLAN)

The software scanned for wireless WLAN 802.11 Access Point identifiers (hereafter referred to as WLAN APs) every 6 minutes. WLAN APs have an indoor range of approximately 125 feet and the university campus had almost complete wireless coverage. Across various locations within the undergraduate residence, over 55 different WLAN APs with varying signal strengths can be detected.

Communication (Call and SMS Records)

The software logged Call and SMS details on the device every 20 minutes, including information about missed calls and calls not completed.

Daily Survey Launcher

For collection of self-report data, the sensing platform included a daily survey launcher. The application launches a foreground survey dialog at 6am everyday that asked the user to respond to symptom-related questions. After three reminders, the smartphone was unusable until the user completed the survey. In the experiment deployment, users were paid $1 USD for every completed daily survey as participation incentive. The survey launcher invoked the daily symptom survey instrument described in the previous section. The design of the survey questionnaire and subsequent labeling of self-report responses was supervised by a trained epidemiologist.

Battery Impact

The battery life impact of periodic scanning has been previously discussed [13]. In this study, periodic scanning of Bluetooth and WLAN APs reduced operational battery life by 10-15%, with average usable life between 14-24 hours (varying with handset models and individual usage). Windows Mobile 6.x devices have relatively poorer battery performance than other smartphones, and WLAN usage (web browsing by user) had a bigger impact on battery life than periodic scanning.

DATASET CHARACTERISTICS

The dataset described here corresponds to the date range from 1st February to 15th April 2009, the peak influenza months in New England. The phone sensor data during this period consists of 1.4 million scanned Bluetooth devices, 201,000 scanned WLAN APs, 15,700 call data records and 11,269 SMS records. In order to perform meaningful analysis, it is important to separate the effects due to immunization prior to the experiment. In a pre-study baseline survey completed few days before start of the study, 20 participants reported received influenza immunization via a flu-shot or flu-mist spray. These participants are not considered in the analysis in the next section.

Daily Symptom Surveys

A total of 2994 survey responses were generated using the smartphone-based survey launcher described in the previous section within the relevant date range, of which 2099 responses were from individuals who had not been immunized, with an approximate survey completion rate of 63% overall during the study. These responses were converted into 48-hour windows, since individuals take up to one day to report a symptom. This approach also reduces the impact of uncompleted survey responses, since there is a higher likelihood that a participant will have completed at least one survey during a 48-hour period. A few samples were dropped because of missing sensor data during the period, e.g., due to lost or broken mobile phones. With these steps, a total of 2283 sets of good quality mobile behavior sensor data and dependent variable reports were obtained.

For analysis, mobile symptom reports have to be converted into syndrome conditions. Specifically, it is important to distinguish between symptoms that represent common colds and allergies versus CDC-defined influenza [20] which has a characteristic signature reflected in runny nose, sore throat and fever symptoms. Due to our limited expertise in this area, selected combinations of self-reported symptoms were labeled as CDC-defined influenza by a medically trained epidemiologist. In our dataset, twelve such cases of influenza, on average lasting 5-7 days and each affecting a distinct individual were observed. The respiratory symptoms not identified as influenza cases by our expert, are considered common colds or seasonal allergies.

ANALYSIS

Mobile Behavioral Features

The following features were extracted from mobile phone sensor data over 48-hour window sizes, with 50 percent overlapping windows. The window size was chosen for epidemiological reasons, as individuals take up to one day to report a symptom. The features chosen represent statistics of whom we talk to and where we are, i.e. the total number of interactions, the diversity of interactions, and the entropy of our behaviors. Such observational data has been shown to reflect important aspects of individual and collective behavior, like friendships and individual job satisfaction [13, 25]. For days with reduced activity, the entropy features capture the higher predictability of an individual's behavior, and was a better feature than total number of bluetooth devices or WLAN APs observed.

Total Communication

This is the total number of phone calls and SMS exchanged, both with other participants as well as third parties. This measure includes incoming and outgoing communication.

Late night and Early Morning Communication

Call and SMS communication between 10pm and 9am on weekdays, with both other participants and non-participants.

Communication Diversity

The number of unique individuals reflected in phone and SMS communication within the 48-hour period.

Physical Proximity Entropy with Other Participants

This is the entropy of distribution of Bluetooth proximity with other participants.

$$H_p = -\sum_{i=1}^{n} p(x_i) log_2 p(x_i)$$

where $p(x_i)$ is the empirical probability of Bluetooth proximity with the remote device x_i belonging to another participant, within the particular time-window, i.e., $p(x_i)$ is the ratio of the number of times the remote device x_i was scanned divided by total count of scanned devices in the 48-hour period.

Physical Proximity Entropy with Other Participants Late Night and Early Morning

Similarly, this is the entropy of the distribution of Bluetooth proximity with other participants in the study, but only during late-night and early morning periods.

Physical Proximity Entropy for Bluetooth Devices Excluding Experimental Participants

Similarly, this is the entropy of distribution of Bluetooth proximity. However, all Bluetooth devices in discoverable mode scanned on the phone are considered in this case. This feature reflects variations in interactions with 'familiar strangers', i.e., bluetooth beacons that the user is often in proximity to, say at the bus-stop or in the classroom [23].

WLAN Entropy based on University WLAN APs

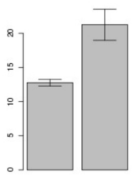
This is entropy for the distribution of WLAN access points scanned within the given period. Only WLAN APs belonging to the university are considered.

$$H_w = -\sum_{i=1}^{n} p(x_i) log_2 p(x_i)$$

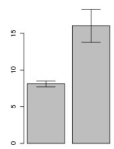
where $p(x_i)$ is the empirical probability of scanning a WLAN AP x_i within the particular time-window. Similar to Bluetooth physical proximity above, $p(x_i)$ is the ratio of the number of times the WLAN AP x_i was scanned divided by the total count of scanned WLAN APs in the 48-hour period.

WLAN Entropy based on external WLAN APs
Similarly, this is entropy for the distribution of WLAN access points scanned within the given period. Only WLAN APs external to the university are considered.

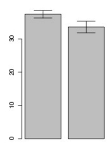
Behavioral Effects of Low Intensity Symptoms (Runny Nose, Sore Throat and Cough)



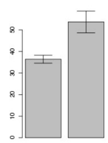
(a) Total communication increases ***



(b) Late-night early morning communication increases **



(c) Overall Bluetooth entropy decreases *



(d) Total WLAN APs detected increase **

Figure 1. Behavior effects of runny nose, congestion, sneezing symptom, n=587/2283, *: $p < 0.05$ **: $p < 0.01$ ***: $p < 0.001$

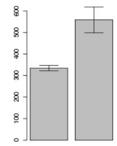
A sore throat or runny nose report may either be a symptom of CDC-defined influenza or simply an independent respiratory condition due to common colds or allergies.

For the runny nose condition (n=587/2283), participants show increased total communication as well as increased late night early-morning communication. Additionally, total counts of Bluetooth proximity and measures of WLAN entropy increases, which is perhaps counter-intuitive. P-values are generated using unbalanced t-tests assuming unequal variance.

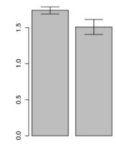
For sore-throat reports, Bluetooth-based entropy with respect to other residents in the study dormitory increases. This again, is slightly counter-intuitive, but may be explained if participants are spending more time indoors and hence have a higher likelihood of interacting with other participants, than they would if they were spending time in classes and activities. It is also found that WLAN based entropy measures, both with respect to university WLAN APs and external WLAN APs decrease with sore-throat reports, indicating more predictable movement patterns for the individual.

Behavior Effects of Higher-Intensity Symptoms (Fever and Influenza)

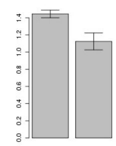
For more intense conditions like a fever or CDC-defined influenza, participants have lower activity and entropy levels,



(a) Bluetooth entropy with respect to other dorm residents increases ***



(b) WLAN entropy with respect to university WLAN APs reduces *



(c) WLAN entropy with respect to external WLAN APs reduces **

Figure 2. Behavior effects of sore throat and cough symptom, n=393/2283, *: $p < 0.05$ **: $p < 0.01$ ***: $p < 0.001$

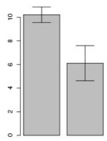
and this is captured using mobile sensors. Due to the severity of these symptoms, the number of reported cases in our dataset is lower than that of low intensity symptoms (runny nose, sore throat/cough). The number of rate of infection amongst participants and study cohort sizes, however, are comparable to Phase I clinical trials [29].

For fever, variations are observed in the late night early morning behavior. Phone communication, Bluetooth proximity counts, and Bluetooth entropy all show a decrease for the late night early morning window. WLAN-based entropy measures with respect to the university WLAN APs as well as external WLAN APs both reduce dramatically.

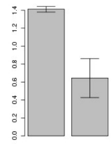
Similar effects are seen for days labeled as CDC-defined Influenza, as overall Bluetooth entropy, Bluetooth entropy with regard to other dorm residents and WLAN based entropy features decrease. This is also expected because fever is a known influenza symptom.

Behavioral Effects of Stress and Mental Health Symptoms

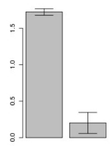
In addition to the physical symptoms described in the above section, the on-device mobile questionnaire also includes two daily questions related to stress levels and sadness, loneliness or depression. As discussed in the previous section, the link between behavior change, physical symptoms and stress is not very well understood. Measuring these self-report variables alongside symptom data allows modeling the covariance and potentially causation across the three sets of variables. With both often-stressed and sad-depressed-lonely responses in our dataset, participants show a consis-



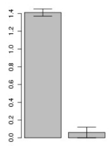
(a) Late night early morning calls and SMS decrease **



(b) Late night morning Bluetooth counts and entropy decrease*

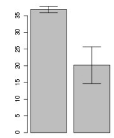


(c) WLAN based entropy with respect to university WLAN APs decreases ***

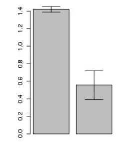


(d) WLAN Entropy with respect to external WLAN APs decreases ***

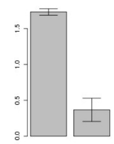
Figure 3. Behavior effects of fever, n=36/2283, *: $p < 0.05$ **: $p < 0.01$ ***: $p < 0.001$



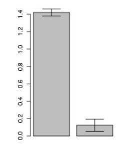
(a) Total Bluetooth interactions and entropy decrease **



(b) Late night early morning Bluetooth entropy with respect to other participants decreases **



(c) WLAN based entropy with respect to university WLAN APs decreases ***



(d) WLAN Entropy with respect to external WLAN APs decreases ***

Figure 4. Behavior effects of CDC-defined influenza, n=54/2283, *: $p < 0.05$ **: $p < 0.01$ ***: $p < 0.001$

tent tendency to isolate themselves, reflected in various sensor modalities.

For the often-stressed response, participants communication diversity decreases, both overall Bluetooth based entropy and Bluetooth entropy with respect to other residents during late-night early morning hours decreases, and WLAN based entropy decreases both with respect to university WLAN APs and external WLAN APs.

For the sad-lonely-depressed responses, a similar tendency to isolate themselves is observed. Total communication decreases and communication during late-night early morning decrease, overall Bluetooth entropy and Bluetooth entropy with respect to other residents decreases.

Symptom Classification using Behavioral Features

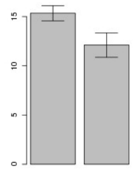
It is evident that there are characteristic behavioral changes associated with respiratory symptoms, fever, influenza, stress and depression. With this is in mind, is it useful to train a classification scheme that identifies when individuals are likely to be symptomatic from behavioral features alone. There are two key considerations with regard to designing such a classification scheme.

First, consider how such a classification system would be used in a scenario where the user has the 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 especially useful for conditions with risk of underreporting by patients (e.g., mental health, elderly healthcare). With this goal in mind, the classification model should have asymmetric misclassification penalties.

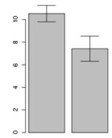
A second consideration is due to correlations amongst dependent symptoms. While behavior variations with respect to symptoms are reported individually in the previous section, in reality, self-reported symptoms are correlated. Figure 7(a) shows the correlations between these variables, reordered using K-nearest-neighbor clustering based on effect size. Four main clusters that emerge are: stress + depression; runny nose + sore throat; fever + influenza; and runny nose + sore throat + fever + influenza.

Given these considerations and unbalanced class sizes, classification is done using a Bayesian-network classifier with MetaCost, a mechanism for making classifiers cost-sensitive [27]. Structure learning for the network is performed using K2 hill climbing and the results are based on 4-fold cross-validation.

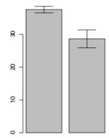
Recall, Precision and F-measure for the symptoms class as a function of increasing misclassification penalty for the symptoms class are plotted in Fig 7(b) - 7(f), for different symptom clusters. Recall from the trained classifier is also compared with random assignment of priors averaged over 1000 simulated runs, to demonstrate improvement over 'chance'. The X-Axis represents increasing MetaCost misclassifica-



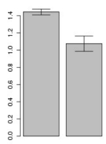
(a) Total communication decreases *



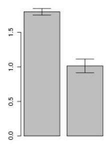
(b) Late-night early morning communication decreases *



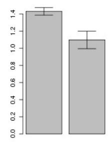
(c) Overall Bluetooth entropy decreases *



(d) Late night early morning Bluetooth entropy with respect to other experiment participants reduces **



(e) WLAN based entropy with respect to university WLAN APs decreases ***



(f) WLAN Entropy with respect to external WLAN APs decreases ***

Figure 5. Behavior Changes with self-reported sad-lonely-depressed responses n=282/2283, *: $p < 0.05$ **: $p < 0.01$ ***: $p < 0.001$

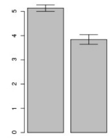
tion penalties, and the Y-Axis shows Recall, Precision, F-Measure (for symptom class) and Recall based on chance. As seen, Recall of the symptom class improves substantially for the sore throat + runny nose + congestion cluster, and also for influenza and fever clusters. Overall prediction accuracy (not shown) is not a useful quality metric due to unbalanced classes, and ranges between 60-80%.

Temporal Flux Between Behavior, Stress and Physical Symptoms

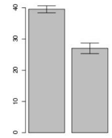
There is extensive medical and health policy interest in understanding the temporal link between behavior change, stress and physical symptoms. In statistics, the Granger causality test is a technique for determining whether one time series is useful in forecasting another– A time series X is said to Granger-cause Y if it can be shown, usually through a series of F-tests on lagged values of X (and with lagged values of Y also known), that those X values provide statistically significant information about future values of Y. Unfortunately, Granger causality tests have been shown to have poor noise immunity. In this section, we use a recently proposed spectral method, Phase Slope Index to gain insight into the temporal relationships between signals of behavior features, stress and physical symptoms.

The Phase Slope Index (PSI) Method

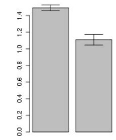
PSI [17] is a recently proposed spectral estimation method designed to measure temporal information flux between time-series signals. The method is based on the knowledge that



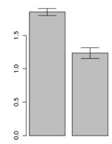
(a) Communication diversity decreases **



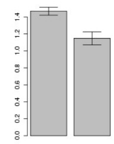
(b) Overall Bluetooth entropy decreases **



(c) Late night early morning Bluetooth entropy with other experiment participants reduces **



(d) WLAN based entropy with university WLAN APs decreases ***



(e) WLAN Entropy with external WLAN APs decreases ***

Figure 6. Behavior Changes with self-reported often-stressed responses n=559/2283, *: $p < 0.05$ **: $p < 0.01$ ***: $p < 0.001$

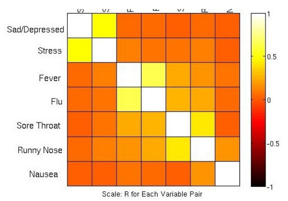
the phase slope of the cross-spectrum of two signals can be used to estimate information flux between these signals in the time domain. Independent noise mixing does not affect the complex part of the coherency between multivariate spectra, and hence PSI is considered more noise immune than Granger analysis. PSI has been used to make causal inferences for brain cell activation and other domains, and is calculated as,

$$\Psi_{ij} = \Upsilon\left(\sum_{f \in F} C^{*}_{ij}(f) C_{ij}(f + \delta f)\right)$$

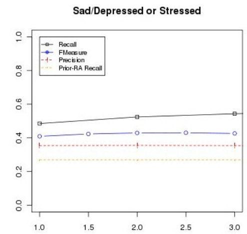
where C_{ij} is the complex coherency. When the input signals are distributed across multiple epochs, then this estimate is normalized by its standard deviation, calculated using the Jackknife method [17].

Results

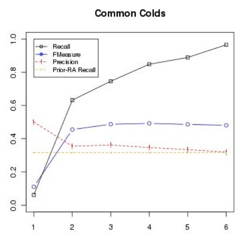
Our approach to using PSI for measuring information flux is based on validating causal links consistently across multiple participants in our dataset. This approach is first validated on two simulated time series of varying sequence lengths (n, representing number of continuous samples available per user) that have a partial-causal relationship between them, and additive noise. The leading time series has x symptom days. The follower time series has y lagged symptom days and z days of additive uniform noise, where the lag between the two series for symptom days is 1 or more days. The scatter-plot in Figure 8 shows the ability of PSI to recover causal structure (normalized PSI coefficient > 0) across different ranges of parameters for the simulated signals. The X



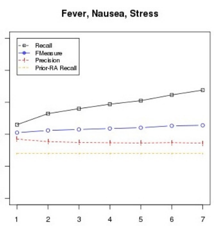
(a) KNN reordered correlations between dependent symptom variables



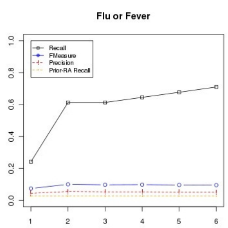
(b) Sad-Depressed-Stressed Symptoms



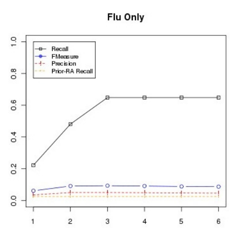
(c) Sore-Throat, Cough, Runny Nose, Congestion, Sneezing Symptoms



(d) Fever, Nausea, Stress Symptoms



(e) Flu and Fever Symptoms



(f) Flu only (as per CDC definition)

Figure 7. Classification results, recall for different symptoms ranges from 0.6 to 0.9 for the symptom class. Y-Axis shows Recall, Precision, F-Measure (for symptom class) and Recall based on chance. X-Axis is the MetaCost misclassification penalty used with the Bayesian Network classifier.

and Y-axes represent n and x, and each point is averaged over 1000 runs with y=x/3 and z=x/3, (these values are such that would be intuitively expected for symptoms in our dataset), and all points above the plane of Z = 0 represent correctly-estimated PSI values. It is important to note that the method recovers the correct direction of information flux for 97.6% of the samples over the surface of the simulated signals.

In order to apply PSI to our dataset, the subset of participants that show both physical symptoms and stress and depression related responses are considered. There is however, a trade off to be made between using data from fewer participants with longer sequences and hence more reliable estimates, versus using data from more participants with shorter sequences, and better validation across participants. Hence, PSI was estimated for two sets of data– sequences of minimum length 40 days and minimum length 60 days, shown in Figure 9.

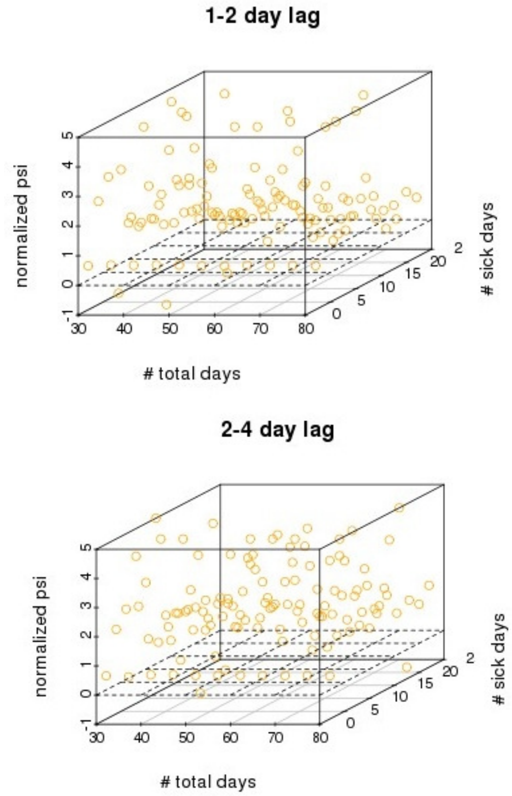


Figure 8. PSI evaluation on simulated data. Z-axis is the estimated PSI value, across a wide range of total days (n) and sick days(x), with additive noise. Points above the Z=0 plane (97.6%) represent correctly estimated direction of information flux.

Each approach generates slightly different directed links and normalized coefficients. The twelve largest PSI coefficients across both methods on the basis of a combined ranking score are listed in descending order in Table 2 and illustrated in Figure 10. An example insight is that 'often-stressed' is useful in forecasting proximity, communication and WLAN behaviors, which suggests that individuals realize and report that they are stressed before it is reflected in their behavior. Another insight is that in two cases Bluetooth interaction features are used to forecast WLAN features– this suggests that a behavior change is reflected in face-to-face interactions with others before it is reflected in the movement patterns of the individual.

CONCLUSION

In this paper, we describe a novel application of ubiquitous computing. We use mobile phones as an active sensing and prediction platform to identify behavior changes reflected in mobile phone sensors, when individuals suffer from common colds, influenza, fever, stress and mild depression. We show that it is possible to determine the health status of individuals using information gathered by mobile phones alone, without having actual health measurements about the subject. Given the pervasiveness of the mobile phone, this opens an avenue for modeling of epidemiological contagion in social networks without medical health reports. We hope our findings apprise spatial and behavioral epidemiology.

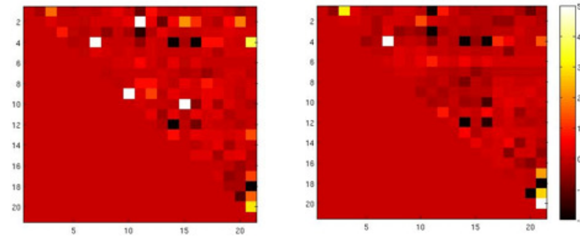
Table 2. PSI Results ordered by combined scores

Source	Follower
Runny nose	WLAN entropy with external APs
Sad-depressed-lonely	Sore throat-cough
Often stressed	Total Bluetooth proximity counts
Communication diversity	Late-night early morning Bluetooth proximity counts
Often stressed	Communication diversity
Often stressed	Late-night early morning Bluetooth proximity counts
Bluetooth entropy with other residents	External WLAN entropy
Runny nose	Total WLAN counts
Often stressed	WLAN entropy with university APs
Bluetooth proximity counts with other residents	External WLAN entropy
Late-night early morning communication	Overall Bluetooth entropy
Sad depressed lonely	Bluetooth entropy

There are nonetheless, several limitations of this work, that we are trying to overcome. The statistical tests in the analysis section assume that the samples are independent, which is not an entirely correct assumption, since we have repeated measures from a limited set of individuals. This would be improved with a repeated-measures approach, that will also allow us to understand how much of the observed variation in symptomatic behavior is individual dependent. Our work does not account for confounding behavior changes due to external events, e.g., exams or the end of semester. On the modeling front, the Bayesian classifier used does not incorporate stochastic information about symptoms or behaviors from previous days– this temporal structure could be better modeled using the latent Markov family of models. As mentioned earlier, there is also extensive interest in augmenting compartmental epidemiological models with parameters that represent empirical behavior changes for symptomatic individuals. In addition to classifying observed symptoms from behavior, we are also interested in forecasting when individuals are likely to be sick in future analysis.

The understanding of behavior change in social health will also benefit from the continuous evolution of mobile sensing technologies. Our mobile platform did not support Bluetooth signal strength, which could provide a better measure of physical proximity. WLAN-based location sensing could be replaced with GPS and other location technologies in the future.

Overall, we believe this work opens a valuable new area for the ubiquitous computing community in social health and predictive healthcare. The development of predictive health



(a) Min Length 40 (b) Min Length 60

Figure 9. PSI co-efficients for two sets of sequences based on participant data. List of features: 1=sad-depressed-lonely 2=often-stressed 3=sore-throat 4=runny-nose 5=fever 6=nausea 7= influenza 8=total communication 9=latenight/early morn comm. 10= communication diversity 11=total Bluetooth proximity 12=overall Bluetooth entropy 13=Bluetooth proximity with other residents 14=Bluetooth entropy with other residents 15=late-night/early morn Bluetooth proximity with other residents 16=late-night/early morn Bluetooth entropy with other residents 17=WLAN counts 18=external WLAN counts 19=overall WLAN entropy 20=WLAN entropy with university APs 21=WLAN entropy with external APs

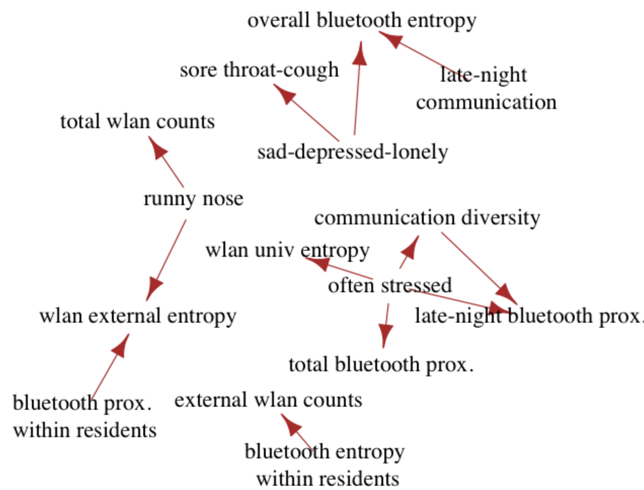


Figure 10. Highest-ranked PSI relationships across both data subsets. Directed ties represent temporal flux.

tools will improve the doctor-patient interaction model, such that health-workers and nurses can use diagnostic information for early detection of conditions, ultimately leading to better healthcare for individuals and lower costs for providers and insurers.

ACKNOWLEDGEMENTS

The authors would like to thank Devon Brewer and Iolanthe Chronis for help with experiment design and logistics, and Kate Farrahi and Anind Dey for comments. Research was 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 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.

REFERENCES

1. T. Abdelzaher, Y. Anokwa, P. Boda, J. Burke, D. Estrin, L. Guibas, A. Kansal, S. Madden, and J. Reich. Mobiscopes for human spaces. *IEEE Pervasive Computing*, pages 20–29, 2007.
2. S. Avancha, A. Baxi, and D. Kotz. Privacy in mobile technology for personal healthcare. Submitted to ACM Computing Surveys, December 2009.
3. A. Barabasi. Network Medicine–From Obesity to the "Diseasome". *New England Journal of Medicine*, 2007.
4. D. Brockmann. Human mobility and spatial disease dynamics. *Review of Nonlinear Dynamics and Complexity*, 2009.
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. N. Christakis and J. Fowler. The collective dynamics of smoking in a large social network. *New England Journal of Medicine*, 358(21):2249, 2008.
7. J. Cohen, P. Neumann, and M. Weinstein. Does preventive care save money? health economics and the presidential candidates. *The New England journal of medicine*, 358(7):661, 2008.
8. S. Cohen and G. Williamson. Stress and infectious disease in humans. *Psychological Bulletin*, 109(1):5–24, 1991.
9. 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.
10. D. Lazer, et. al. Computational Social Science. *Science*, 323(5915):721–723, 2009.
11. D. M. Musher. How Contagious are Common Respiratory Track Infections. *New England Journal of Medicine*, 2003.
12. P. Drummond and B. Hewson-Bower. Increased psychosocial stress and decreased mucosal immunity in children with recurrent upper respiratory tract infections. *Journal of psychosomatic research*, 43(3):271–278, 1997.
13. N. Eagle, A. Pentland, and D. Lazer. Inferring Social Network Structure Using Mobile Phone Data. *Proceedings of National Academy of Sciences*, 106(36):15274–15278, 2009.
14. J. Epstein, D. Goedecke, F. Yu, R. Morris, D. Wagener, and G. Bobashev. Controlling pandemic flu: the value of international air travel restrictions. *PLoS One*, 2(5), 2007.
15. R. Fletcher and S. Fletcher. *Clinical epidemiology: the essentials*. Lippincott Williams & Wilkins, 2005.
16. J. Fowler and N. Christakis. Dynamic Spread of Happiness in a Large Social Network: longitudinal analysis over 20 years in the Framingham Heart Study. *British Medical Journal*, 337(dec04 2):a2338, 2008.
17. G. Nolte et al. Robustly Estimating the flow of direction of information in complex social systems. *Physical Review Letters*, 100, 2008.
18. J. Ginsberg, M. Mohebbi, R. Patel, L. Brammer, M. Smolinski, and L. Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2008.
19. M. Gonzalez, C. Hidalgo, and A.-L. Barabasi. Understanding Individual Human Mobility Patterns. *Nature*, 453:779–782, 2008.
20. J. Imboden, A. Canter, L. Cluff, and et al. Convalescence from influenza: a study of the psychological and clinical determinants. *Archives of Internal Medicine*, 108(3):393, 1961.
21. A. Madan and A. Pentland. Modeling Social Diffusion Phenomena Using Reality Mining. In *AAAI Spring Symposium on Human Behavior Modeling*, 2009.
22. D. G. McNeil. Models Projections for Flu Miss Mark by Wide Margin. *New York Times*, June 2009.
23. S. Milgram. The Familiar Stranger: An Aspect of Urban Anonymity. *The Individual in a Social World: Essays and Experiments, Longman*, 1977.
24. MIT Media Lab. Social evolution project. `http://social.media.mit.edu`.
25. D. Olguin, P. Gloor, and P. A. Capturing Individual and Group Behavior Using Wearable Sensors. In *AAAI Spring Symposium, Human Behavior Modeling*, Palo Alto, CA, 2009.
26. O.S. Miettinen. *Theoretical Epidimiology: principles of occurence research in medicine*. Wiley New York, 1985.
27. P. Domingos. MetaCost: A General Method for Making Classifiers Cost-Sensitive. In *Fifth International Conference on Knowledge Discovery and Data Mining, KDD-99*, 1999.
28. P. Elliott, et al. *Spatial Epidimiology*. Oxford University Press, 2000.
29. H. Rang, M. Dale, J. Ritter, and P. Moore. *Pharmacology 5th Edition*. Churchill Livingstone, Edinburg, 2003.
30. R. Yirmiya, Y. Pollak, and et. al. Illness, Cytokines, and Depression. *Annals-New York Academy of Sciences*, 917:478–487, 1999.