Social sensing: Obesity, unhealthy eating and exercise in face-to-face networks

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Social Sensing: Obesity, Unhealthy Eating and Exercise in Face-to-Face Networks

Anmol Madan  
MIT Media Lab  
Cambridge MA  
anmol@media.mit.edu

Sai T. Moturu  
MIT Media Lab, MA  
Arizona State University,  
Tempe, AZ  
smoturu@media.mit.edu

David Lazer  
Northeastern Univ  
Boston MA  
david_lazer@harvard.edu

Alex (Sandy) Pentland  
MIT Media Lab  
Cambridge MA  
pentland@media.mit.edu

ABSTRACT
What is the role of face-to-face interactions in the diffusion of health-related behaviors- diet choices, exercise habits, and long-term weight changes? We use co-location and communication sensors in mass-market mobile phones to model the diffusion of health-related behaviors via face-to-face interactions amongst the residents of an undergraduate residence hall during the academic year of 2008-09. The dataset used in this analysis includes bluetooth proximity scans, 802.11 WLAN AP scans, calling and SMS networks and self-reported diet, exercise and weight-related information collected periodically over a nine month period. We find that the health behaviors of participants are correlated with the behaviors of peers that they are exposed to over long durations. Such exposure can be estimated using automatically captured social interactions between individuals. To better understand this adoption mechanism, we contrast the role of exposure to different sub-behaviors, i.e., exposure to peers that are obese, are inactive, have unhealthy dietary habits and those that display similar weight changes in the observation period. These results suggest that it is possible to design self-feedback tools and real-time interventions in the future. In stark contrast to previous work, we find that self-reported friends and social acquaintances do not show similar predictive ability for these social health behaviors.

1. INTRODUCTION
According to the World Health Organization, we are currently in the midst of a global obesity epidemic, with over a billion overweight and over 300 million clinically obese adults worldwide[26]. This increasing trend is attributed to lifestyle changes in our society, including increased consumption of energy-dense, nutrient-poor foods with high levels of sugar and saturated fats, and reduced physical activity.

In addition to these factors, recent work suggests that obesity and other health-related lifestyle decisions spread through social networks, and in particular long-term face-to-face networks may play an important role. Longitudinal studies based on the Framingham Heart study social network indicate that health-related behaviors from obesity [5] to happiness [11] can spread through social ties. The effects of social networks and social support on physical health are well-documented [1, 18]. However, these studies depend on self-reported information and do not quantify actual face-to-face interactions that may lead to changes in behavior. The availability of such data would be useful in answering many open questions in the context of social contagion of health behaviors. For example, to what extent are eating habits of an individual influenced by those of their spouse, roommate, close-friend or casual acquaintance? Is the adoption of social behaviors a characteristic of the person being influenced, or the influencer or simply the context of the relationship? Is the underlying causal process social contagion or is it practically impossible to disambiguate homophily and confounding due to the limitations imposed by measurement and estimation techniques?

In the past, due to the absence of pervasive wireless sensors (mobile phones, sensor badges, or others), and accuracy limitations of self-reported surveys and human recall, it was impossible to build fine-grained models of human interactions. However, with ubiquitous mobile phones, we can now use short-range bluetooth radios, cellular-tower identifiers, Global positioning system (GPS) sensors and other location technologies to model specific interactions, relationships, and homogeneity of behaviors amongst people. In this paper, we use wireless sensing techniques to automatically capture these interactions and estimate their effects on health behaviors. In addition to face-to-face interactions, phone and email communication also deserve to be studied as alternate modalities for diffusion of behaviors between people.

Popular models of diffusion phenomena do not account for continuous individual exposure. Cascade and threshold models treat interactions between nodes as a point estimate of tie-strength and not as a continuous multi-dimensional interactions. SEIR (susceptible-exposed-infected-recovered)
models are an extension of SIR epidemiological models that include an exposure state, but only as an averaged incubation period parameter common to the sample population and not as a unique value for every node. In this paper, exposure to different opinions for individuals as measured using mobile phone sensors is used to explain BMI changes. The analysis in this paper is based on social interaction data and health related behavior self-reports collected using mobile phone sensors at an undergraduate university dormitory over an entire academic year. In the next section, the experiment design, mobile data collection platform, and dataset characteristics are described in more detail. Understanding the role of social network interactions in the ongoing obesity epidemic will enable the design of novel technologies and better interfaces to control adverse spreading, and facilitate social support real-time interventions for positive reinforcement of a healthy lifestyle.

2. RELATED WORK

Until the start of this century, most of the data collected on human interactions was through self-reported surveys and experience sampling. More recently, long term monitoring has been implemented using a variety of technologies including video [24], smartphones [8, 9, 10, 16] and wearable sensing devices [23, 4, 14, 20, 21].

Social interactions in the real world present a great avenue for understanding user behaviors. Using multiple sensors (infrared sensors, proximity to Bluetooth devices, proximity to Wi-Fi or Bluetooth base stations), we can detect social proximity and face-to-face interactions between people and quantify such interactions over time. Communication patterns between individuals over the phone and the web can be studies using smartphones.

Physiological sensing is being increasingly used to study health [15, 22, 19]. However, the focus on the social determinants of health is limited. Longitudinal studies based on the Framingham Heart study social network indicate that health-related behaviors from obesity [5] to smoking [6] to happiness [11] can spread through social ties. This work has generated greater interest in the study of peer effects on health [12]. Further, the effects of social networks and social support on physical and mental health and the powerful role that they can play in health promotion are well documented [1, 2, 3, 13, 7, 18].

While these studies clearly indicate the importance of social determinants on health, there is limited work studying real-world interactions and their impact on health. The latest sensing technologies provide us with the capabilities to collect such fine-grained data and gather new insights. In this work, we undertake such a mission and study a closely connected network of individuals and how their interactions are correlated with their behavior.

3. EXPERIMENT DESIGN

The dataset described below was collected as part of longitudinal study with seventy residents of an undergraduate dormitory. These residents represent eighty-percent of the total population– most of the remaining twenty percent of residents that declined to participate citing privacy concerns were spatially-isolated. The dormitory is known within the university for its pro-technology orientation and the decision of students to reside was based on self-selection. The students were distributed roughly equally across all four academic years (freshmen, sophomores, juniors, seniors) and 60 percent of the students were male. The study participants also included the graduate resident tutors responsible for each floor.

This overarching experiment was designed to study the adoption of political opinions, diet, exercise, obesity, eating habits, epidemiological contagion, depression and stress, dorm political issues, interpersonal relationships and privacy. A total of 320,000 hours of human behavior data was collected in this experiment. In this paper however, we only discuss the mobile platform, dataset and analysis related to changes in dietary habits, weight changes and physical exercise. The overall experiment is described in more detail here [16].

For training labels, participants completed social health related survey instrument for the months of March, April and June, shown in Table 1. Participants also listed their close friends and social acquaintances (binary responses) while completing each monthly survey. The histograms of BMI changes and weight changes for all participants from March to June is plotted in Figure 4. In addition, Figure 4 also shows the Pearson’s correlations between the dependent variables.

### Table 1: Monthly Social Health Survey for Dependent Labels

<table>
<thead>
<tr>
<th>Survey Question</th>
<th>Possible Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current weight and height (weighing scales provided in common areas)</td>
<td>Numeric values</td>
</tr>
<tr>
<td>Servings of salads consumed, on average per week</td>
<td>0 to 7 (or more) servings</td>
</tr>
<tr>
<td>Servings of fruits and vegetables consumed, on average per week</td>
<td>0 to 7 (or more) servings</td>
</tr>
<tr>
<td>Self reported level of healthiness of diet</td>
<td>6-point Likert scale, Very Unhealthy to Very Healthy</td>
</tr>
<tr>
<td>Instances of aerobic exercise (20 mins or more), on average per week</td>
<td>0 to 7 times</td>
</tr>
<tr>
<td>Instances of active sports, on average per week</td>
<td>0 to 7 times</td>
</tr>
</tbody>
</table>

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, partici-
pants were compensated $1 per day that they completed the
on-device symptom survey.

4. MOBILE SENSING PLATFORM

The mobile phone based platform for data-collection was
designed with the following features and long-term sensing
capabilities.

4.1 Device Selection

The platform is based on Windows Mobile 6.x devices, as
they can be deployed with all four major American opera-
tors. Software was written using a combination of native-C
and managed-C#. The software-sensing package was sup-
ported for six different handset models in the Windows Mo-
 bile product range. All supported devices featured WLAN,
EDGE and SD Card storage, and most featured touch screens,
flip-out keyboards. The HTC Tilt, a popular GSM phone in
our experiment is shown in Fig 1.

4.2 Proximity Detection (Bluetooth)

The software scanned for proximate Bluetooth wireless de-
vices every 6 minutes. The Windows Mobile phones used in
our experiment were equipped with class 2 Bluetooth radio
transceivers, which have a realistic indoor sensing range of
approximately 10 feet. Scan results for two devices in prox-
imity have a high likelihood of being asymmetric, which is
accounted for in our analysis. Due to API limitations with
Windows Mobile 6.x, signal strength was not available to

<table>
<thead>
<tr>
<th>Table 2: Raw Logged Data Formats</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bluetooth</strong></td>
</tr>
<tr>
<td>UTC timestamp</td>
</tr>
<tr>
<td><strong>Scans</strong></td>
</tr>
<tr>
<td>1-way hash of remote device MAC</td>
</tr>
<tr>
<td><strong>WLAN</strong></td>
</tr>
<tr>
<td>UTC timestamp</td>
</tr>
<tr>
<td><strong>Scans</strong></td>
</tr>
<tr>
<td>1-way hash of Access Point MAC</td>
</tr>
<tr>
<td>AP ESSID</td>
</tr>
<tr>
<td><strong>Call</strong></td>
</tr>
<tr>
<td>UTC Start timestamp</td>
</tr>
<tr>
<td>Records</td>
</tr>
<tr>
<td>UTC end timestamp</td>
</tr>
<tr>
<td>1-way hash of remote phone number</td>
</tr>
<tr>
<td>Incoming vs. outgoing flag</td>
</tr>
<tr>
<td>missed call flag</td>
</tr>
<tr>
<td>user roaming flag 0-1</td>
</tr>
<tr>
<td><strong>SMS</strong></td>
</tr>
<tr>
<td>UTC timestamp</td>
</tr>
<tr>
<td>Records</td>
</tr>
<tr>
<td>1-way hash of remote phone number</td>
</tr>
<tr>
<td>incoming/outgoing flag</td>
</tr>
</tbody>
</table>

the sensing application. Table 2 show the logged formats
for bluetooth data.

4.3 Approximate Location (802.11 WLAN)

The software scanned for wireless WLAN 802.11 Access
Point identifiers (hereafter referred to as WLAN APs) ev-
ery 6 minutes. WLAN APs have an indoor range of 125ft
and the university campus has almost complete wireless cov-
erage. Across various locations within the undergraduate
residence, over 55 different WLAN APs with varying signal
strengths can be detected. WLAN logs were captured in the
format shown in Table 2.

4.4 Communication (Call and SMS Records)

The software logged Call and SMS details on the device
every 20 minutes, based on recent events. These logs in-
cluded information about missed calls and calls not com-
pleted. The format for logging calls and SMS messages is
shown in Table 2.

4.5 Battery Impact

The battery life impact of periodic scanning has been pre-
viously discussed [10]. 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). Win-
dows Mobile 6.x devices have relatively poorer battery per-
formance than other smartphones, and WLAN usage (web
browsing by user) had a bigger impact on battery life than
periodic scanning.

4.6 Backend Database

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 finally stored in
various MySQL tables for analysis.

4.7 Open Source Availability

This sensing software platform for Windows Mobile 6.x
has been released under the LGPLv3 open source license for
public use, and is available for download here[25].

5. DATASET CHARACTERISTICS
In the following analysis, social interaction data for the entire spring semester is considered. The mobile phone dataset for this period includes 20609 phone calls, 11669 SMS messages and 2291184 scanned bluetooth devices, which includes communication with non-residents. As seen in Figure 2, even with the same set of individuals, differences in interaction modalities produce different interaction networks. Clear weekly and daily temporal structure is observed in the interactions amongst individuals as seen in figure 3. For example note the spike in SMS communication on Friday nights compared to other nights of the week. Similarly, the daily distribution of sample counts indicates that phone calls subside around 5am, reflecting the practice of sleeping late common to this community. In prior work [10, 16], this temporal interaction structure has been used to identify friendship ties within the interaction network. In [17], it is found that discriminating between interactions during different hours can be used to identify self-reported political discussant ties for an individual.

6. ANALYSIS
The effect of social influence on behavior has been well established in the literature, and behavior is at the root of the obesity problem with dietary choices and exercise habits playing a significant role. Christakis and Fowler’s recent work [5] on the effect of social ties on obesity has generated greater interest in the study of peer effects on health [12]. While these studies have used networks of self-identified social contacts, they lacked data about real face-to-face interactions that occur on a more regular basis. With the belief that this information is useful in studying this phenomenon, we set out to analyze the effect that friends, acquaintances and face-to-face interactions have on weight change in our study population.

6.1 Features that Reflect Exposure

Mobile phone interactions reflect exposure to opinions and behaviors in physical proximity and phone communication. For each individual, we compute exposure as the number of alters reflected in the interaction modalities listed below, conditioned upon health-related behaviors described in the next section.

- Total Bluetooth exposure: alters reflected in Bluetooth co-location data
- Late-Night & Early-morning Bluetooth exposure: alters reflected in Bluetooth co-location data between the hours of 9am and 9am the next morning
- Weekend Bluetooth exposure: alters reflected in Bluetooth co-location data only for Saturdays and Sundays
- Total Phone and SMS exposure: alters reflected in phone communication and SMSs exchanged; both incoming and outgoing communication are included. Duration of calls is currently not considered
- Weekend Phone and SMS exposure: alters reflected in phone communication and SMS exchanged only for Saturdays and Sundays

As an alternative, in the following analysis we also considered the interaction counts for the ego-alter pair as exposure features. However, the number of auto-detected alters conditioned by their behaviors outperformed the more complex features in our dataset.
Body Mass Index (BMI) is a commonly used metric to estimate healthy body weight, based on an individual’s height. It is equal to 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.

In this work, we use an individual’s change in BMI as a dependent variable and study the influence of various exposure-based independent variables described above using linear regression. BMI is used as it is a better indicator of healthy weight because it takes into account some differences in an individual’s physical stature. However, similar results as described below were obtained while using an individual’s change in absolute weight as a dependent variable (see table 4) serving to depict that the results hold even if a different measure is used for weight change.

6.2.1 Exposure to Overweight and Obese Peers

As per the definition of obesity, participants with BMI $\geq 30$ are considered obese in our dataset. The independent variables used in this analysis were the number of obese persons that had actual interactions with the individual in question in the form of Total Bluetooth exposure and Late-Night & Early-morning Bluetooth exposure. A significant correlation that explained about 17% of the variation was found as reported in Table 3.

For comparison, this analysis was repeated using the number of self-reported close friends and social acquaintances that were obese as independent variables. However, no significant correlation was found between the self-reported independent variables and the dependent variable measuring change in an individual’s BMI from March to June 2009.

As the dataset contains only a small number of obese people, the previous analysis was repeated using exposure features to include both overweight and obese individuals (BMI $\geq 25$) as opposed to just obese individuals. In this case, Total Bluetooth exposure and Late-Night & Early-morning Bluetooth exposure to individuals that are overweight or obese explained about 25% of the variation (see Table 3).

As earlier, when this analysis was repeated using exposure to self-reported close friends and social acquaintances that were either overweight or obese, no significant correlation was found between the self-reported independent variables and the dependent variable measuring change in an individual’s BMI.

6.2.2 Exposure to Peers with Unhealthy Diets and Poor Exercise Habits

So far, we have looked at exposure features that focused on the physical aspects indicative of peer health. In this section, exposure features indicative of healthy or unhealthy diets and poor exercise habits are considered.

As explained previously, in monthly surveys, participants reported their diet on a 6-point Likert scale ranging from ‘Very unhealthy’ (1) to ‘Very healthy’ (7). Based on the distribution of responses, a response of 3 or less on this scale is considered as unhealthy eating behavior. BMI Change for

Figure 5: Pearson correlation coefficient matrix for all dependent self-report variables for June 09, all $p < 0.01$. From top to bottom, left to right, the variables are salads per week (mean $= 1.5$, sd $= 1.4$), veggies and fruits per day (mean $= 1.9$, sd $= 1.3$), healthy diet category (mean $= 3.8$, sd $= 1.3$), aerobics per week (mean $= 2.1$, sd $= 2.1$), sports per week (mean $= 0.8$, sd $= 1.5$), BMI change compared to March 09 (mean $= 0.11$, sd $= 0.68$). BMI Change does not show a strong correlation with either the eating or exercise habits. A healthy diet shows some correlation with the playing sports.
the period of March to June 2009 was once again the dependent variable and a similar pattern of results as earlier were observed. Total Bluetooth exposure and Late-Night & Early-morning Bluetooth exposure to peers with unhealthy eating habits explained approximately 17% of the variation in the dependent variable. The exposure to close friends and social acquaintances with unhealthy eating habits did not show significant correlation with BMI change in this period.

Then, we try to understand the role of exposure to individuals who tend to be less physically active. Total activity is the sum of self-reported responses for aerobics per week and sports per week, from the survey responses in the previous section. Based on the distribution of responses, an individual is considered inactive if the total activity is less than or equal to 3. The results were again consistent with the previous section, where Total Bluetooth exposure and Late-Night & Early-morning Bluetooth exposure to peers who were physically inactive explained about 23% of the variance in BMI change.

6.2.3 Exposure to Peers Who Had Substantial Weight Gain in the Same Period

Finally, we looked to see if there was a correlation between BMI change for an individual from March-June 2009 and exposure to peers who gained substantial weight during the same period. Only individuals who gained more than 4 pounds were considered.

The Total Bluetooth exposure and Late-Night & Early-morning Bluetooth exposure features show the most significant correlation to BMI change of all our analyses, and these features together explain about 35% of the variability in the independent variable. It is also interesting to note that when the same analysis was repeated using individuals who lost more that 4 pounds, none of the features showed significant correlations suggesting that a different dynamic might be at play for exposure to good behaviors. Consistent with the above analysis, exposure to close friends and social acquaintances who gained weight did not show significant correlation.

7. CONCLUSIONS & DISCUSSION

In this work, we study the impact that social interactions in real world face-to-face networks have on BMI and weight changes in a co-located student community. Our approach allows us to understand, the role of exposure to different types of peers—those that are obese, overweight, have unhealthy dietary habits, and inactive lifestyles. We find that exposure measured using bluetooth proximity to peers that are overweight or obese and to peers that have unhealthy dietary habits or inactive lifestyles, can influence weight changes in an individual as opposed to exposure to close friends and social acquaintances. The largest correlations observed are with social exposure to peers with large weight gains during the same period. In all cases, we find that exposure measured via self-reported close-friend and acquaintance relationships is not statistically significant. These results are intuitive in that they suggest that we are affected by the behaviors of those that we interact with. However, there is potential for further validation with larger studies in the future.

It is also important to note that these results are based on a relatively small population of students that may not be representative of large real-world communities. However, they provide a starting point for the discussion on the importance of studying social networks based on real world interactions. We see statistically significant results indicating that face-to-face interactions might actually have a much larger effect on individual behaviors affecting health. In related work, we have seen similar results with face-to-face interactions affecting how individuals form political opinions [17].

These results paint a bright picture for future studies of social networks fueled by the latest advances in mobile sensing technologies. These technologies allow us to collect fine-grained data on a larger scale that would not have been possible earlier. For our part, we will continue to study this large observational dataset to tease out more interesting results around health and other behavioral aspects. We hope that these results will guide the next generation of mobile phone-based studies to gather much more interesting data leading to far greater insights.

8. ACKNOWLEDGMENTS

The authors would like to thank Iolanthe Chronis and other undergraduate contributors for their help with the experimental deployment. Dr. Sai Moturu was partially sponsored by an AFOSR award to Prof. Huan Liu at Arizona State University under Award Number FA9550-08-1-0132. Research was partially sponsored by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0052, 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.

9. REFERENCES


