Assessing Social Anxiety using GPS Trajectories and Point-Of-Interest Data

Yu Huang*, Haoyi Xiong*, Kevin Leach, Yuyan Zhang, Philip Chow, Karl Fua
Bethany A. Teachman, Laura E. Barnes
University of Virginia, Charlottesville, VA, USA
{yh3cf,hx6d,kjl2y,yz2fh,pic2u,kcf3st,bat5x,lb3dp}@virginia.edu

ABSTRACT

Mental health problems are highly prevalent and appear to be increasing in frequency and severity among the college student population. The upsurge in mobile and wearable wireless technologies capable of intense, longitudinal tracking of individuals, provide valuable opportunities to examine temporal patterns and dynamic interactions of key variables in mental health research. In this paper, we present a feasibility study leveraging non-invasive mobile sensing technology to passively assess college students' social anxiety, one of the most common disorders in the college student population. We have first developed a smartphone application to continuously track GPS locations of college students, then we built an analytic infrastructure to collect the GPS trajectories and finally we analyzed student behaviors (e.g. studying or staying at home) using Point-Of-Interest (POI). The whole framework supports intense, longitudinal, dynamic tracking of college students to evaluate how their anxiety and behaviors change in the college campus environment. The collected data provides critical information about how students' social anxiety levels and their mobility patterns are correlated. Our primary analysis based on 18 college students demonstrated that social anxiety level is significantly correlated with places students' visited and location transitions.

ACM Classification Keywords

H.1.2. Models and Principles: User/Machine Systems; J.4 Computer Applications: Social and Behavioral Sciences

Author Keywords

Mobile Sensing, Social Anxiety, GPS, Location Semantics

INTRODUCTION

Social anxiety is a widespread problem among college students, associated with impairment in academic functioning and relationships, and avoidance of many of the opportunities that college can offer. The American College Health Association reported that 40% reported feeling "overwhelming anxiety"

---

* These authors contributed equally.

at least once in the preceding year [1]. Among these mental health disorders, social anxiety is one of the most common disorders among college students.

Unobtrusive monitoring and modeling of real-time fluctuations in mental illness including symptoms of social anxiety, are conducted through surveys, skin temperatures, heart rate, etc. Such methodologies are often employed in a clinical laboratory setting, which consumes resources and time. To assess mental health disorders in a non-invasive fashion, engineers and scientists have worked to understand mental health through *in situ* data collection and analysis. The emergence of mobile sensing technologies shows promise in understanding of the relationship between human behavior and mental health status.

Mobile phones have become ubiquitous in daily life. By taking advantage of the various sensors embedded in commodity mobile phones, a rich set of data can be collected for researchers to analyze and relate one's daily life to mental health status as a way to study human emotions. Data from accelerometers, light sensors, gyroscopes, GPS, and other indicators provide a rich set of features that can shine a light on how mental health issues are associated with unique patterns in daily living. Among these sensors, GPS is one of the most commonplace sensors in commercially available mobile phones. The GPS trajectories of college students not only show where they have been, but also provides crucial information about daily life and behavior. This includes how long they spend on campus, how frequently they visit bars or clubs, or how frequently they visit the home of a relative. Thus, we are motivated to study the feasibility of leveraging mobile technologies to passively monitor college students' daily behavior in order to gain further insight into the relationship between social anxiety and these everyday behaviors.

Recently, seminal works [3, 14] have already started the investigation of using smartphone-based behavior sensing to collect users' health-related behavior data and correlate with their mental health status or stress levels. We take five steps that are also employed in these works: (1) recruiting participants and deploying a mobile application for data collection, (2) collecting data such as GPS locations, recognized activities, or self-reported data from participants during the study, (3) assessing participants' mental health status using validated clinical measures, (4) extracting meaningful features/metrics from participants' data (e.g., number of distinct locations visited during the study), and (5) correlating these features/metrics with participants' mental health status (e.g., Pearson correlation between number of distinct locations and clinically validated

measures). The above studies have already defined a new line of research and also demonstrated a set of methodologies that researchers could follow.

Proposed Study

In this paper, as a complementary data source, we propose a feasibility study to investigate the correlation between smartphone-sampled GPS locations and social anxiety. Utilizing the aforementioned approach, we perform the study as follows.

Data Collection and Pre-processing

Participants are high or low socially anxious college students recruited from undergraduate psychology courses who receive course credits and money as compensation. Then, participants' social anxiety levels are assessed in the pre-screening process using the Social Interaction Anxiety Scale (SIAS) [9]. SIAS is a well-validated measure of trait social anxiety that assesses individuals' level of distress and anxiety in a variety of socially evaluative situations (e.g., when meeting a stranger). The study lasted 10 days, during which participants' GPS locations were passively collected. Participants also completed brief questionnaires about mood and affect at random times each day. Given the GPS trajectories of each participant, we grouped common locations visited by the participant with a meaning, such as "home" or "work." These grouped significant locations were labeled using the *Points-of-Interest* (POI) information provided by Foursquare [15]. In total, five location types were labeled: "home," "work (study)," "religious," "food & leisure," and "transportation."

Metric Definition and Correlation Analysis

After pre-processing the data, we defined two types of metrics based on the labeled location trajectories. The first metric is the time that a participant spent in one type of location during the study period. The second metric is the frequency that each participant transits from one type of location to another. We measure each participant using these metrics, and perform a correlation analysis using these metrics and the participant's pre-assessed SIAS score. The correlation and significance analysis results indicate that participants' social anxiety level correlates with their mobility patterns. In other words, it is feasible to utilize mobility patterns to assess social anxiety levels.

Contributions

We make the following contributions in this paper.

1. We investigated the feasibility of studying the impact of social anxiety on mobility patterns. To the best of our knowledge, this paper is the first work studying the correlation between social anxiety and mobility patterns, by addressing the semantics (i.e., POI information) of locations.

2. Compared to existing studies characterizing a user's mobility pattern using the diversity and complexity of GPS trajectories, we propose to represent human mobility patterns by incorporating the semantics of the trajectory. Specifically, given the GPS trajectories and POI information, two novel metrics were defined—the first metric is the time that the participant stayed in different locations and the frequency of transitions between these locations.

3. A real-world study with 18 participants was organized for data collection. The correlation and significance analysis results indicate that the visiting patterns among "home", "work," and "religious" locations might be related to the social anxiety of college students.

RELATED WORK

There has been a significant body of work dedicated to monitoring and understanding human mental health status. Considering the research question of our study, we review the related literature from two main perspectives: 1) mental health monitoring frameworks, and 2) location-based sensing and analysis for human behavior.

Mental Health Monitoring

Psychological research on anxiety and depression has mostly relied on laboratory-based methods. Ecological Momentary Assessment (EMA) has been used to investigate socially anxious individuals' reactions to positive experiences and their ability to differentiate emotions. It found important differences in the social-emotional daily lives of anxious individuals and demonstrated the feasibility of this methodology to investigate social interaction and emotional regulation in the context of psychopathology [7, 8]. These studies relied on repeated sampling at discrete time intervals on a palmtop PC. Thus, data is limited by timing and objective bias. With the development of mobile sensing, a growing number of studies have been conducted by using personal smartphones and wearable sensors [10, 14, 3, 2, 5]. The StudentLife study uses a continuous sensing app [14] to monitor students in a smartphone programming class over the course of the semester and identified strong correlations between sensing data and PHQ-9 measures. Another study uses an app called Purple Robot to detect location, movement, phone usage, and other activities to assess a user's likelihood of depression [10]. In Canzian et al. [3], mobility trajectories sensed from smartphones also have a demonstrated correlation with the PHQ-8 depression measure.

Location-based Sensing

Location data have been used in some of the mental health monitoring studies described above. Each of these studies [10, 14, 3, 2, 5] explored the trajectories of GPS locations to correlate with human mental health status. Outside the mental health domain, as one of the most important indicators of human behaviors and daily life patterns, GPS locations are used to analyze and predict disease outbreak and transmission [12, 13, 17, 6, 4]. In these studies, GPS locations are used to track the movement and interactions between different objects. However, none of these studies explores the deeper information associated with GPS locations: the social context for the individuals. In other words, the geographical level analysis of GPS locations can only reveal the movement or interaction information, but loses the social context of locations by shallowly explaining the appearance at a certain location to be just an appearance. For example, if someone appears in a particular GPS location, the state-of-the-art work can only use this to infer the mobility of the person. However, the location class (e.g. restaurant, coffee shop, church) is potentially more

indicative of an individual's behaviors and social contexts. As a result, we hypothesize that presence in a location is not a good enough indicator to infer social interactions and that more information is needed about the context of a location.

In this section, we introduce the design of our study and data collection procedures. Then, we present the data pre-processing approach and study metrics. Finally, we discuss the setting of our analysis.

Study Organization and Data Collection

Participants are undergraduate students with varying levels of social anxiety recruited from undergraduate psychology classes who receive course credit as participation incentives. A subset of participants received monetary compensation. The decision to recruit university students was based on two reasons: (1) there are high rates of social anxiety among young adults, and (2) recruiting young adults in a university setting provides a relatively homogeneous sample in terms of life phase, psychological stressors, and life experiences, thereby eliminating a wide variety of potential nuisance factors.
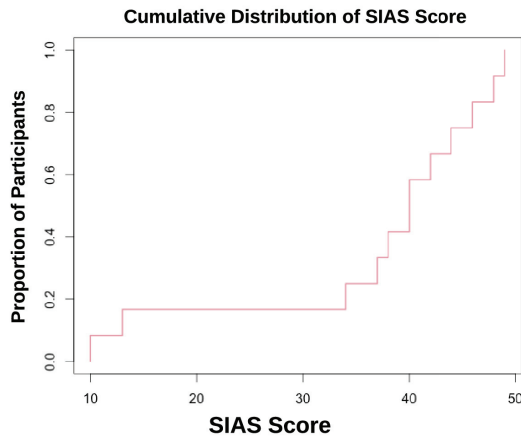
Figure 1. The cumulative distribution of SIAS scores for recruited participants

After receiving approval from the Institutional Review Board at our university, 18 participants were recruited. We first pre-assessed the social anxiety of each recruited participant using the Social Interaction Anxiety Scale (SIAS [9]). The SIAS measurement contains 20 items rated from 0 to 4. Generally, a higher SIAS score (specifically, higher than 40) indicates a higher risk of having social anxiety concerns; a low score indicates lower risk for social anxiety concerns. Figure 1 shows the cumulative distribution of SIAS scores among the 18 participants in our study. Our 18 participants' SIAS scores have a mean of 38.39 and a standard deviation of 14.85.

Next, we explained the study to participants, participants consented, and we installed a custom mobile app on their personal smartphone. As shown in Figure 2, participants were informed that the application passively collected the GPS location information every 150 seconds and uploaded it to an Amazon Web Services S3 server. The study length was 10-days. When the experiment was completed, researchers could then access all participants' raw GPS data for the further analysis.
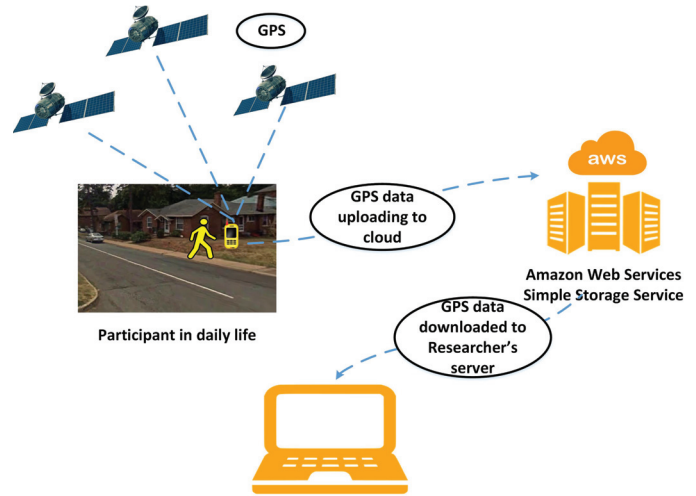
Figure 2. Data collection framework.

GPS Data Pre-processing and Feature Extraction

Given each participant's GPS trajectories, we first clustered the GPS trajectories by spatial and temporal locality, and then removed GPS points where the participant stayed less than 5 minutes (e.g., roads used for travel). Further, we labeled each location cluster in trajectories using the POI information (e.g., residence area, office area, campus area, and restaurants) obtained from Foursquare. With POI, we labeled the trajectories using five types of locations:

- **Work:** For the college students in this experiment, work locations are mostly within the campus with the exception of dormitories.

- **Food & Leisure:** This location set includes restaurants, grocery stores, and shopping centers.

- **Transportation:** Train and bus stations.

- **Religious:** Catholic and Christian churches.

- **Home:** Some participants have multiple home locations. In this study, we do not distinguish between them.

According to each participant's labeled GPS trajectories, we created two metrics to represent their mobility patterns:

1. **Cumulative Staying Time** in each type of location: Given a type of location and a specific participant, this metric characterizes the total time that the participant spent in one type of location during the study. For example, the total time a participant stayed in all "home" locations during the study might include her own apartment, her parents' house to which she returned every week, and her partner's apartment that she stayed overnight on occasion.

2. **Transition Frequency** from one type of locations to another: Given a specific participant and two types of locations (e.g., "home" and "work"), this metric characterizes the frequency that the participant transited from one type of location to another. This metric applies unidirectionally; for example, the transition frequency of "home"→"work" and the transition frequency of "work"→"home" are different.

Table 1. Correlation and Significance Analysis of Mobility vs. SIAS Score: Pearson Correlation (with p-values) between participants' SIAS Scores and time spent in specific types of locations. Additionally, the normalized mean ($\bar{x}$) and standard deviation (σ) of participants' stay at each location is shown.

Cumulative Staying Duration	Pearson	p-Value	$\bar{x}$	σ
Work	0.03	0.91	.37	.26
Transportation	0.11	0.68	.04	.10
Religious	**-0.44**	**0.06**	.04	.16
home	0.14	0.57	.46	.27
Food & Leisure	0.12	0.65	.10	.14

Given the five types of locations and the two metrics, we extracted a total of 25 features (i.e., Cumulative Staying Time in the five locations as well as the transition frequencies between any two different types of locations) for each participant using her labeled location trajectories.

Data Analysis

The goal of our study is to investigate (1) if extracted features (i.e., Cumulative Staying Time and Transition Frequency) are correlated with the pre-assessed SIAS score, and (2) if the extracted features can predict SIAS scores. To achieve the goal, we performed two sets of analysis:

- **Correlation and Significance Analysis** — In this study, we first estimated the Pearson correlation between each feature and SIAS score, to see if they are correlated positively or negatively, as well as the strength of correlation. To further verify the significance of the correlation among the recruited participants, we also calculated the significance level.

- **Linear Regression Analysis** — In this study, we first performed linear regression using extracted features as predictors and SIAS score as the target. We used two types of linear regression — Least Square Error (LSE) estimator and Least Absolute Shrinkage and Selection Operator (LASSO). The coefficient of each feature indicates the importance of the feature in the regression. Note that we are particularly interested in the coefficients learned by LASSO, when a limited number of samples are given for regression [16].

RESULTS AND ANALYSIS

In this section, we present the results of our analysis investigating the relationship between SIAS scores, Cumulative Staying Time, and Transition Frequency.

Correlation and Significance Analysis — Table 1 shows the Pearson correlations between Cumulative Staying Time and the SIAS score, as well as the corresponding p-Value. It appears the time spent in religious locations is negatively correlated with SIAS, which could mean that people who attend religious services are less likely to develop social anxiety, or that people with social anxiety are less likely to want to attend religious services, perhaps given their social nature. However, the p-value indicates that this result is not significant. The correlation to other locations appears weak. Thus, we cannot conclude that Cumulative Staying Time in their locations is correlated with social anxiety levels.

Table 2. Correlation and Significance Analysis of Mobility vs. SIAS Score: Pearson Correlation (with p-values) between each participant's SIAS Scores and her transition frequency from one significant location to another.

Transitions	Pearson	p-Value
Work → Food & Leisure	-0.412	0.088
Work → Transportation	0.088	0.725
Work → Religious	**-0.656**	**0.002**
Work→Home	-0.013	0.959
Food & Leisure → Work	0.421	0.080
Food & Leisure → Transportation	0.010	0.966
Food & Leisure → Religious	0.0003	0.998
Food & Leisure → Home	-0.050	0.841
Transportation → Work	0.090	0.722
Transportation → Food & Leisure	0.046	0.855
Transportation → Home	0.099	0.693
Religious → Work	**-0.575**	**0.012**
Religious → Food & Leisure	0.160	0.524
Religious → Transportation	0.094	0.709
Religious → Home	**-0.636**	**0.004**
Home → Work	0.002	0.991
Home → Food & Leisure	-0.077	0.759
Home → Transportation	0.134	0.595
Home → Religious	**-0.521**	**0.025**

Table 3. LSE and Lasso weights for cumulative staying time of POI locations. $\lambda = 4.4$ corresponds to the minimum MSE with 10-fold cross validation.

Locations	W_{LSE}	W_{Lasso} ($\lambda = 0.1$)	W_{Lasso} ($\lambda = 4.4$)
Work	34.2745	-0.3118	0
Transportation	45.6317	9.8432	0
Religious	**-4.2888**	**-38.5820**	**-12.5284**
Home	34.9082	0	0
Food & Leisure	43.0769	7.5765	0

Table 4. LSE and Lasso weights for transitions from one POI location to another. $\lambda = 3.4$ corresponds to the minimum MSE with 10-fold cross validation.

Transition	W_{LSE}	W_{Lasso} ($\lambda = 0.1$)	W_{Lasso} ($\lambda = 3.4$)
Work → Food & Leisure	51.8166	0.9826	0
Work → Transportation	93.6872	0	0
Work → Religious	32.5070	-8.2215	**-1.0637**
Work→Home	25.5635	0	0
Food & Leisure → Work	-50.0034	0	0
Food & Leisure → Transportation	25.6898	-2.2673	0
Food & Leisure → Religious	5.1067	0	0
Food & Leisure → Home	-0.4174	0	0
Transportation → Work	-89.9101	-2.2751	0
Transportation → Food & Leisure	-26.7613	0	0
Transportation → Home	-37.3759	0.8583	0
Religious → Work	-34.7113	-0.6419	0
Religious → Food & Leisure	0	7.2377	0
Religious → Transportation	84.2753	10.4818	0
Religious → Home	0	7.7794	**-4.3684**
Home → Work	-25.3932	0.0634	0
Home → Food & Leisure	0	-0.6911	0
Home → Transportation	0	3.8580	0
Home → Religious	-8.3211	0	0

Table 2 shows the results of Pearson correlation (with p-values) between transition frequencies and SIAS. Transitions between "Work" and "Home" to / from "Religious" locations are significantly correlated with social anxiety scores. These transitions are not surpising given that religious service attendance has been demonstrated to predict an increased well-being across a number of studies [11].

Linear Regression Analysis — Table 3 shows the linear regression analysis results when using Cumulative Staying Time in all five locations as predictors and the SIAS score as the target. The results include the coefficients of each location using both LSE and LASSO. Specifically, we address the results of LASSO using two $\lambda = 0.1$ and 4.4. Note that the minimal Mean Square Error was achieved when $\lambda = 4.4$. The result demonstrates the negative relationship between visiting religious locations and social anxiety level, which is similar to the correlation and significance analysis. The importance of religious locations becomes clearer when using LASSO. Religious location is the single variable with a non-zero coefficient when $\lambda = 4.4$. Compared to the results based on LASSO and LSE, we conclude that LSE model might be over-fitting due to the low sample size. It appears that the visiting patterns to "religious" locations is an important variable of linear regression in our study.

In addition to cumulative staying time, Table 4 shows the linear regression analysis results when using transition frequencies between two types of locations as predictors and the SIAS score as the target. When using LASSO with $\lambda = 3.4$, the minimal mean square error was achieved. In such case, only transitions between "home" and "religious" are with non-zero coefficients. This trend is as the same as in the corresponding significance analysis.

Limitations of Our Study — Due to the small number of recruited participants and their relatively simple life cycles as college students, the data we collected is biased. A larger sample size is needed to make valid inferences from the data. Further, roughly 6 of our 18 participants who frequently visited religious locations and who also have low social anxiety scores might have dominated the results.

CONCLUSION

With the increasing prevalence of social anxiety disorder and other mental health concerns, novel techniques for assessing psychological state have become increasingly important. Combined with the ubiquity of mobile phones and their increasingly powerful sensors, we can acquire a rich array of data from which we can derive a person's psychological or emotional state.

In this paper, we present a feasibility study assessing college students' social anxiety through GPS-based localization. Compared to existing studies characterizing mobility with GPS trajectories, we proposed to represent mobility by incorporating the semantics of the trajectory, such as a person's home or preferred leisurely location. We presented a study of 18 participants, demonstrating the significance that a person's risk for social anxiety correlates with their GPS trajectory. Further work will investigate the importance of location semantics in

a larger population and extend the approach to other mental health disorders.

REFERENCES

1. 2014. *American College Health Association-National College Health Assessment II: Reference Group Executive Summary Spring 2014*. Hanover, MD: American College Health Association (ACHA).

2. Michelle Nicole Burns, Mark Begale, Jennifer Duffecy, Darren Gergle, Chris J Karr, Emily Giangrande, and David C Mohr. 2011. Harnessing context sensing to develop a mobile intervention for depression. *Journal of medical Internet research* 13, 3 (2011).

3. Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 1293–1304.

4. Susan M Cooper, H Morgan Scott, Guadalupe R De la Garza, Aubrey L Deck, and James C Cathey. 2010. Distribution and interspecies contact of feral swine and cattle on rangeland in south Texas: implications for disease transmission. *Journal of Wildlife Diseases* 46, 1 (2010), 152–164.

5. Agnes Gruenerbl, Venet Osmani, Gernot Bahle, Jose C Carrasco, Stefan Oehler, Oscar Mayora, Christian Haring, and Paul Lukowicz. 2014. Using smart phone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients. In *Proceedings of the 5th Augmented Human International Conference*. ACM, 38.

6. Michael Irey, Tim R Gottwald, James H Graham, Tim D Riley, and Greg Carlton. 2006. Post-hurricane analysis of citrus canker spread and progress towards the development of a predictive model to estimate disease spread due to catastrophic weather events. *Plant Health Progress* 10 (2006).

7. Todd B. Kashdan and R. Lorraine Collins. 2010. Social anxiety and the experience of positive emotion and anger in everyday life: an ecological momentary assessment approach. *Anxiety, Stress, & Coping* 23, 3 (2010), 259–272. DOI:`http://dx.doi.org/10.1080/10615800802641950` PMID: 19326272.

8. Todd B. Kashdan and A. S. Farmer. 2010. Differentiating emotions across contexts: Comparing adults with and without social anxiety disorder using random, social interaction, and daily experience sampling. *Emotion* 14, 3 (2010), 629–638.

9. Richard P. Mattick and J.Christopher Clarke. 1998. Development and validation of measures of social phobia

scrutiny fear and social interaction anxiety1. *Behaviour Research and Therapy* 36, 4 (1998), 455 – 470. `DOI: http://dx.doi.org/10.1016/S0005-7967(97)10031-6`

10. Sohrab Saeb, Mi Zhang, Christopher J Karr, Stephen M Schueller, Marya E Corden, Konrad P Kording, and David C Mohr. 2015. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *Journal of medical Internet research* 17, 7 (2015).

11. Patrick R Steffen, Kevin S Masters, and Scott Baldwin. 2016. What Mediates the Relationship Between Religious Service Attendance and Aspects of Well-Being? *Journal of religion and health* (2016), 1–13.

12. Dallas Swendeman and Mary Jane Rotheram-Borus. 2010. Innovation in sexually transmitted disease and HIV prevention: Internet and mobile phone delivery vehicles for global diffusion. *Current opinion in psychiatry* 23, 2 (2010), 139.

13. Gonzalo M Vazquez-Prokopec, Donal Bisanzio, Steven T Stoddard, Valerie Paz-Soldan, Amy C Morrison, John P Elder, Jhon Ramirez-Paredes, Eric S Halsey, Tadeusz J Kochel, Thomas W Scott, and others. 2013. Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. *PloS one* 8, 4 (2013), e58802.

14. Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 3–14.

15. Dingqi Yang, Daqing Zhang, and Bingqing Qu. 2016. Participatory Cultural Mapping Based on Collective Behavior Data in Location-Based Social Networks. *ACM Transactions on Intelligent Systems and Technology (TIST)* 7, 3 (2016), 30.

16. Peng Zhao and Bin Yu. 2006. On model selection consistency of Lasso. *The Journal of Machine Learning Research* 7 (2006), 2541–2563.

17. Xiaoyong Zhou, Soteris Demetriou, Dongjing He, Muhammad Naveed, Xiaorui Pan, XiaoFeng Wang, Carl A Gunter, and Klara Nahrstedt. 2013. Identity, location, disease and more: Inferring your secrets from android public resources. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. ACM, 1017–1028.