Social Media as a Passive Sensor in Longitudinal Studies of Human Behavior and Wellbeing

Koustuv Saha[1*], Ayse E. Bayraktaroglu[2], Andrew T. Campbell[3], Nitesh V. Chawla[4], Munmun De Choudhury[1], Sidney K. D'Mello[5], Anind K. Dey[6], Ge Gao[2], Julie M. Gregg[5], Krithika Jagannath[2], Gloria Mark[2], Gonzalo J. Martinez[4], Stephen M. Mattingly[4], Edward Moskal[4], Anusha Sirigiri[3], Aaron Striegel[4], Dong Whi Yoo[1]

[1] Georgia Institute of Technology, {koustuv.saha, munmund, dongwhi.yoo}@gatech.edu
[2] University of California at Irvine, ayseelvan@gmail.com, {ge.gao, kjaganna, gmark}@uci.edu
[3] Dartmouth College, campbell@cs.dartmouth.edu, anusha.sirigiri@dartmouth.edu
[4] University of Notre Dame, {nchawla, gmarti11, smattin1, emoskal, striegel}@nd.edu
[5] University of Colorado, Boulder, {sidney.dmello, julie.gregg}@colorado.edu
[6] University of Washington, anind@uw.edu

[*] *All non-first authors are listed in alphabetical order.*

ABSTRACT

Social media serves as a platform to share thoughts and connect with others. The ubiquitous use of social media also enables researchers to study human behavior as the data can be collected in an inexpensive and unobtrusive way. Not only does social media provide a passive means to collect historical data at scale, it also functions as a "verbal" sensor, providing rich signals about an individual's social ecological context. This case study introduces an infrastructural framework to illustrate the feasibility of passively collecting social media data at scale in the context of an ongoing multimodal sensing study of workplace performance (*N*=757). We study our dataset in its relationship with demographic, personality, and wellbeing attributes of individuals. Importantly, as a means to study selection bias, we examine what characterizes individuals who choose to consent to social media data

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05. https://doi.org/10.1145/3290607.3299065.

sharing vs. those who do not. Our work provides practical experiences and implications for research in the HCI field who seek to conduct similar longitudinal studies that harness the potential of social media data.

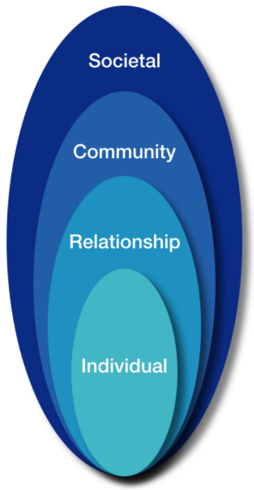

Figure 1: Social Ecological Model: Human behaviors and attributes can be considered to be deeply embedded in the complex interplay between an individual, their relationships, the communities they belong to, and societal factors. Social media provides a passive way to gather quantifiable signals about the social ecological dimensions relating to an individual's behavior [2, 3].

We recruited 757 participants who were information workers in cognitively demanding fields (e.g. software engineers, consultants, managers) across the United States. Participants were requested to remain in study for either upto a year or through April 2019. Enrollment was conducted from January 2018 through July 2018. Participants either received a series of staggered stipends totalling $750 or participated in a set of weekly lottery drawings (multiples of $250 drawings) depending on their employer restrictions.

Sidebar 1: Participant logistics in Tesserae.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**; • **Applied computing** → *Law, social and behavioral sciences*; Psychology.

KEYWORDS

social media, multimodal sensing, passive sensing, personality traits, workplace

ACM Reference Format:

Koustuv Saha et al.. 2019. Social Media as a Passive Sensor in Longitudinal Studies of Human Behavior and Wellbeing. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

INTRODUCTION

Studies of human behavior and wellbeing have typically largely relied on self-reported survey data from individuals. In recent years, a variety of limitations have been noted with these approaches. For instance, survey data suffers from subjective assessments, recall and hindsight biases. These surveys are often retrospective in nature—information is gathered after an event has occurred, or after an individual has experienced a specific change [11].

Recent research in studies of human behavior,has recognized the value of in-the-moment data recording and acquisition approaches. One prominent example centers around the proliferation of active sensing approaches, such as the use of ecological momentary assessments (EMAs) about an individual's momentary state and behavior [12]. However, active sensing methodologies suffer from compliance issues, are difficult to implement at scale and over extended periods of time. They also require careful and highly engineered study design, as well as continual, and proactive engagement of the user by requiring them to answer questions, which may pose a significant burden [14]. Consequently, researchers have begun to employ various forms of passive sensing [12], such as by logging an individual's phone usage and via wearable sensors. In particular, there has been significant success in employing these ubiquitous and increasingly popular sensors to study human behavior, wellbeing, and psychological attributes [12].

Social media provides an inexpensive and unobtrusive means of gathering both present and historical data of individuals in their natural settings [8]. This premise is built upon the findings of a growing body of work, which has employed social media data as a mechanism to identify markers and to assess risk with respect to a variety of psychological health and well-being concerns [4, 9]. Further, because social media data is recorded in the present by an individual, it also serves as a complementary *verbal* sensor to understand the psychological dynamics of an individual, beyond non-verbal passive sensors.

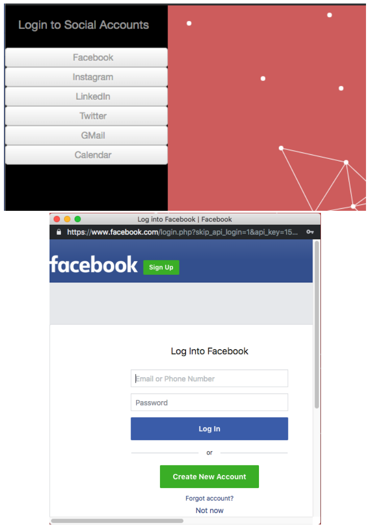

Figure 2: Screenshots of our application homepage, and the pop-up login window when a user clicks on Facebook. Once the user logs into their account, the application collects their data in the back-end.

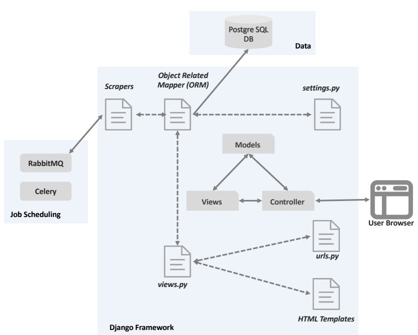

Figure 3: Schematic diagram of the Data Collection Infrastructure.

Its promise is also situated in the observation that many human behaviors and attributes have social underpinnings. When considered via the lens of the Social Ecological Model [2] (Figure 1).

This paper looks at the social media data through the lens of one of the largest in-situ longitudinal studies of human behavior to date. Our project called Tesserae, aims to understand whether and how workplace performance and productivity can be measured from volunteered passive sensing data of 757 individuals over a one year period. We present a case study highlighting our real-world experiences in attempting to passively collect social media data to achieve our research goals along with what we have learned about the specific individuals who voluntarily shared such data and the nature of the social media data itself. Specifically, we first introduce our data collection infrastructure and describe how we navigated several developmental challenges we encountered given the changing ecosystem of social media Application Programming Interfaces (APIs). Then, we provide descriptive statistics on the data we collected—who are the people who authorized social media data access and what does their data tell us when examined in the light of a variety of demographic (e.g., age, gender, education, occupation, and income), personality and wellbeing (e.g., affect and anxiety) attributes.

We believe this work will be instructive and of interest to the members of the HCI community involved in conducting similar studies or in harnessing the benefits and potential of social media more broadly. *Further, to facilitate reproducibility in research, we intend to publish a part of our social media dataset, with necessary privacy and ethical protocols in place, including de-identification measures and explicit consent from the individuals whose data is made available.*

SOCIAL MEDIA DATA COLLECTION APPROACH

The Tesserae Project. We deployed social media as a passive sensing modality of behaviors and wellbeing attributes in a large-scale multi-sensor study [6]. The enrollment process consisted of a set of initial questionnaires related to demographics, and job performance, intelligence, personality, affect, anxiety, and sleep that served as initial ground truth for these constructs. For the sensor streams, the participants were provided with 1) Bluetooth beacons—two static and two portable beacons to track their home and work location, commute and desk time, and social interactions, 2) Wearable—a fitness band based smartwatch to track health measures such as heart rate variability, stress, and physical activity, and 3) Smartphone application—we installed an application on their smartphones (both Android and iOS) to track phone usage. In addition, participants authorized to collect their social media data. We asked them to provide their Facebook and LinkedIn data, *unless they did not consent to do so, or did not have either account*. That is, we sought consent from only those participants who had existing Facebook or LinkedIn accounts from before the study. Additionally, they could optionally consent to their Instagram, Twitter, GMail (metadata only), and Google Calendar data. To collect the social media data from those who consented, we hosted a Python based web application that was developed by our team. This web application was built upon the Django framework and used an Open Authorization (OAuth) based data collection strategy. OAuth protocol is an open standard for access delegation, commonly used as a way for internet users to log in and grant third party

Facebook and Instagram application process involved submitting a document elucidating on what are the purposes of the app, what data will be collected, whether the app would be used for commercial purposes, and they also require a registered webpage for the app. The Instagram application also required preparing a simulated video of the entire pipeline. Once submitted, the applications are reviewed and the review verdict (approval or disapproval) is back within a week. In the month of November 2017, we submitted our approval requests for Facebook and Instagram. For Facebook, our app was approved at a single attempt within two days of submission. In the case of Instagram, our app was rejected twice, and was approved only in the third attempt, in which we additionally explained the broad objectives of our project, and how our data will be stored in the secured servers and databases in the backend.

Sidebar 2: Approval process of Facebook and Instagram API registration.

A user's request from the browser (click on a particular social media platform), is first handled by the *Controller* component, i.e, `urls.py`. This python script assigns the job to the particular *View* component, i.e, `views.py`, where every platform has their own view. It first instructs the controller to show the particular login screen of the platform. If the user successfully logs in, they are shown a "Thank You" page where they could also log out from the application. Their login enables the *View* component to obtain the OAuth information, and to call the *Model* component, whose role is to run the respective scraper of the platform. The scraper scripts use the registered application APIs of the social media platforms, which are saved in the `settings.py`. We used a RabbitMQ messaging queue, and a Celery based task scheduler to manage simultaneous requests of data collection. The collected data was stored in a postgre-sql database in separate tables per platform.

Sidebar 3: Back-end of social media data collection application.

access to their information, without giving them the passwords. Compared to other alternative data collection strategies such as downloading and sharing of social media archives, or scraping through webpage crawlers or smartphone applications, the OAuth protocol provides a more privacy-preserving, streamlined, and convenient means of data collection at scale. Additionally, this not only poses minimal burden to the participants, but also ensures data sharing over a secured channel without transfer of any personal credentials. The following paragraphs explain our data collection infrastructure.

Obtaining Social Media Authorization. The social media web application used the respective *official* API (Application Programming Interface) for each platform (Facebook Graph API, Instagram API, Twitter API, Linkedin API, GMail API, Calendar API). To use these APIs, the first step involved registering on their respective developer websites. Specifically, Facebook and Instagram APIs require special application approval (see Sidebar 2 for details). The process of registering (and being approved) an API let us obtain specific secret keys for each platform that could be used to collect participants' data from the same platform. During registration, we needed to specify the particular fields of data we choose to select, and the key is enabled to provide access to only those data field privileges.

Data Collection Infrastructure. To provide their data, the participants needed to log into their respective social media accounts that they had consented to (see Figure 2). On the back-end, a Django-based social media data collection web application was designed on a Models-Views-Controller (MVC) architecture (see Figure 3 and Sidebar 3). Per IRB approval, the entire infrastructure was hosted on a secure, encrypted servers located in one of our researcher institutions.

Tackling Developmental and Infrastructural Challenges. Given the long-term nature of participant enrollment (spanning several months) and the perpetually changing ecosystem of many of these APIs, we needed to continually debug the above infrastructure and the backend scripts to keep them updated with the changes and deprecations of the services provided by the APIs. For this purpose, three student researchers in the team experimented with several use-cases and automated testing scripts. Sidebar 4 outlines two such instances that complicated our task of debugging the app, and also required many trial-and-error based debugging based on the change in API function calls.

Our data collection encountered a major setback following the Cambridge Analytica data breach [1]. Facebook revoked the services of previously enabled data fields such as education history, number of friends, and location information in the user's profile. Facebook additionally required another round of more comprehensive app approval in August 2018. All of these changes required us to experiment and debug our scripts to continue data collection throughout the enrollment period, however eventually we were able to continue our social media data collection process.

Managing Participant Expectations. During the last phase of completing 757 enrollments, by July 13, 2018, we had 392 participants who authorized their Facebook data, and 262 who had originally consented but did not authorize their Facebook data yet—meaning the data of these 262 participants was not available to us. To obtain their data, we sent targeted emails requesting to provide their

The Instagram API deprecated the previously running Developer API and updated to Graph API in late January 2018. The Linkedin API underwent similar changes in early February 2018, but did not update their developer documentations.

Sidebar 4: Example instances of API changes that mandated debugging.

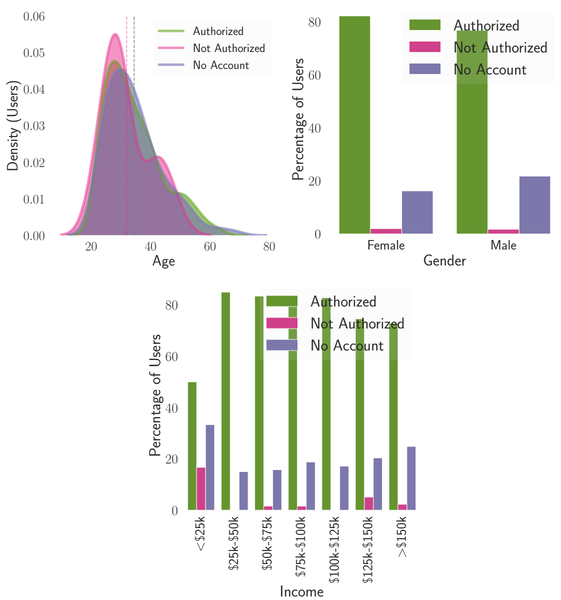

Figure 4: Relationship between demographic attributes and Facebook authorization choices.

Facebook data, along with instructions and a demo video. We followed them up with a reminder every 7-10 days if they did not respond or provide their Facebook data. While 36 participants never responded to our emails, a few responded with concerns such as that they are not active on Facebook, or that they disabled their accounts following the Facebook data breach episode [1]. Some of them offered to create new accounts, and some sought clarification if Facebook data is compulsory for participation. We clarified them that our study does not expect any sort of active social media use, or re-activation/creation of account only for the purposes of the study, or indicated that although they had initially consented, they are welcome to revoke it (one participant revoked their consent). From others, we mostly received positive responses and most of the consented participants provided their Facebook data. Finally, by August 27, 2018, we obtained the Facebook data of 195 additional participants. Therefore our Facebook participant pool consists of 587 who consented and authorized, 67 who initially consented but did not authorize, and 103 who did not have Facebook accounts.

FINDINGS AND INSIGHTS

The remaining part of the case study provides a deeper dive in the Facebook modality, considering that it provides us with a highly comprehensive dataset to understand user behavior in a longitudinal fashion. This is supported not only by the existing statistics that Facebook is the most popular social media platform [7], but also because it was the most prevalent social media stream in our study. During enrollment the participants provided us their demographic attributes, including their gender, age, education, type of occupation, role in company, and income, and also responded to validated surveys on personality traits and wellbeing measures. For personality traits, we used BFI-10's survey that scores the big five personality traits in the range of 1 to 5. For wellbeing, we measured— 1) affect using PANAS scale which scores positive and negative affect in the range of 0 to 50, 2) *anxiety* using the STAI-Trait scale which scores anxiety between 20 and 80, and 3) *sleep* using the PSQI scale which scores sleep between 0 and 21. Accordingly, we first study the distribution of these attributes of the participants from whom we were able to gather Facebook data and those from whom we were not. Next, within the participants who provided us their Facebook data, we examine the relationship of these traits and wellbeing attributes with their Facebook activity, as well as how the amount of their Facebook data varies across these attributes. Together, this also examines if there is self-selection of participants or if our sample of social media sensor is randomly distributed across all 757 participants.

Who Authorized Facebook Data?

We examine if there is any relationship between participant demographic attributes, and their authorization choice of Facebook data. A one-way analysis of variance (ANOVA) suggests no significant differences on the basis of age. In contrast, gender-wise, we find that female participants are more likely to authorize than males, whereas male participants are less likely to have an account compared to the females. Next, those (self-reported to be) born outside the U.S. were more likely to *not* have an account or did not share their data with us, compared to those born in the U.S. Across several income

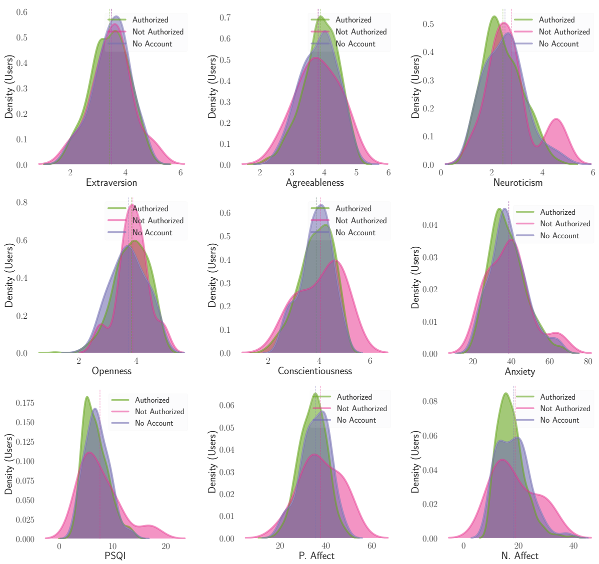

Figure 5: Relationship between data authorization choices, and personality traits and wellbeing attributes.

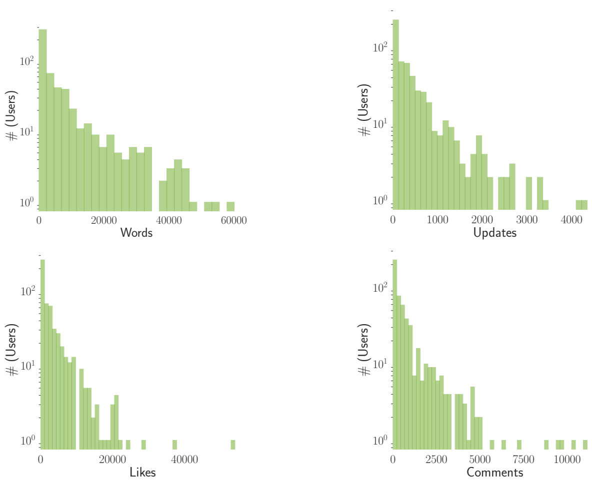

Figure 6: Distribution of words, updates, likes, and comments in our Facebook dataset.

criteria, we find that over 80% of those individuals earning between $25K USD and $125K authorized their Facebook data, whereas those earning the extremes (<$25K and >$150K) were least likely to authorize their data. For the personality traits, we found no significant differences (Figure 5). For the measures of wellbeing: anxiety, sleep, and affect, we find no significant differences in the participant behavior towards sharing their Facebook data. This is also evident per the density plots in Figure 6.

Descriptive Statistics of the Facebook Dataset

Out of the 587 participants who authorized their Facebook data, 544 had some form of Facebook data in their timelines. That is, the remaining 43 participants had no data on their timelines. The Facebook data can be broadly categorized into two types—those which were self-composed by the participants themselves (e.g., writing a status update or checking into a certain location), and the ones that they received on as shared updates on their timeline (e.g., a friend tagging them in a post). More comprehensively, the Facebook data consisted of all kinds of updates shared on the consented participants' timelines, including textual posts, Facebook apps (such as games) usage, check-ins at locations, media updates, being with someone, and the share of others' posts. Additionally, the likes and comments received on each of these updates on the participants' timelines were also collected. Note that as per our IRB approval, we did not collect any multimedia data or private messages.

Table 1 summarizes the descriptive statistics of our Facebook dataset. The median number of updates is 194, that of likes received is 1,151, and that of comments is 330 per participant. Figure 6 plots the distributions of updates, likes, and comments in our Facebook dataset. Temporally, our data ranges back to October 2005, and the number of days of data per participant averages at 1,917. Figures 7 show a temporal distribution of the updates in our dataset, and the number of days on Facebook per participant, giving us a sense of the historical data Facebook allowed us to capture.

Dataset Size and Participant Attributes

Next, within the 587 users who authorized Facebook data, we study the relationship of the quantity (size) of their data with their demographic, personality, and wellbeing attributes.

Demographic Attributes. We observe that the distribution of age over the number of updates, and number of likes and comments received on their timelines show a very varied distribution. Through an ANOVA 1-way analysis of variance, we find no significant differences in the data quantity across age, education, and income (eg. Figure 8). However, in case of gender, while there are negligible differences in their updates shared, there is a significant difference in the likes and comments received— females receive 85% more likes (4,223 vs. 2,289) and 57% more comments (1,019 vs. 649) than the males.

Personality. We then study the differences in the Facebook activity across personality traits. For every personality trait, we define two groups (high and low) based on a theoretically and empirically grounded measure of median threshold. Per Table 2, in extraversion and agreeableness, we find that although the high and low individuals do not show any significant differences in the length of posts

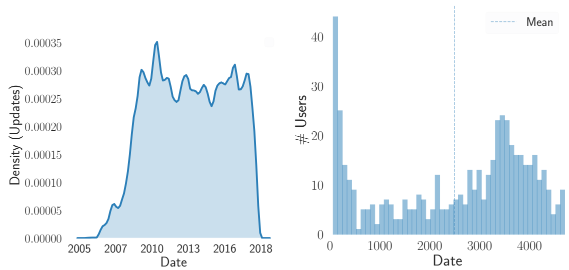

Figure 7: Temporal distribution of Facebook updates and the distribution of the number of days of data.

Table 1: Descriptive statistics of the Facebook dataset.

Type	Total	Range	Mdn	Stdev.
Timeline Updates	237,725	0-4,346	194.5	629.01
Likes Received	1,672,482	0-53,948	1151	5,046.43
Comments Received	452,003	0-11,145	330.5	1403.92
Self-posts	177,857	0-3,541	135	515.82
Self-comments	75,291	0-2,570	46	291.53
Self-Words	3,884,889	0-115,632	2,217	12,604.60

Table 2: Effect size measure (Cohen's d) between groups with high and low personality and wellbeing measures in their number of words, updates, likes, and comments. A positive d indicates the High group has greater values than the Low one, and a negative d indicates the opposite. Magnitudes more than 0.1 are bold-faced.

Attribute	#Words	#Updates	#Likes	#Comments
Personality				
Extraversion	0.03	-0.06	**0.26**	**0.11**
Agreeableness	0.09	-0.04	**0.21**	**0.15**
Neuroticism	**0.12**	-0.03	0.06	**0.11**
Openness	**0.33**	**0.12**	**0.26**	**0.29**
Conscientiousness	**-0.11**	0.07	-0.00	-0.04
Wellbeing				
Anxiety	0.07	-0.09	0.02	0.02
PSQI	**0.28**	0.06	0.05	**0.15**
P.Affect	**-0.21**	**-0.15**	0.05	**-0.11**
N.Affect	**-0.18**	**-0.14**	**-0.27**	-0.07

and the number of updates, they show significant differences in the likes received. That is higher extraversion or agreeableness is associated with greater likes received. Individuals high in neuroticism or openness show greater activity in their length of posts and the number of updates posted. Especially, in openness, we find huge differences both in their update posted and likes or comments received.

Wellbeing. For wellbeing measures, we again used a median split on the State Trait Anxiety scale for anxiety, and the PANAS-X scale for affect to find respective high and low in participants. We find that the individuals with high anxiety trait tend to post longer posts, and while there are no significant differences in the number of updates and likes, these individuals tend to receive a lot more comments on their posts. In terms of affective wellbeing, we find that the individuals who show high values in either of the affect scales (positive or negative affect), show greater tendency to post shorter posts, and lesser updates. While a causality is not established, it is interesting to note that the individuals showing greater negative affect tend to receive lesser likes. In sleep quality, higher PSQI indicates poorer sleep quality. A positive effect-size measure indicates that the individuals who show poorer sleep quality also tend to update more and receive more likes and comments on their posts.

DISCUSSION AND CONCLUSION

This case study gives a first of its kind report of the practical experiences we gained while collecting social media data of individuals in a large-scale multimodal sensing study.

Our analyses unpacked a common and valid question in many passive sensing and social media studies—who are the individuals who agree to share their social media data for research? From a theoretical perspective, we contribute to the literature on the social media use across demographic, personality, and wellbeing attributes of individuals. Except for gender, we do not find significant differences among participants who chose to authorize their Facebook data for our study. From a technical perspective, these differences, or the lack thereof, bear important implications for future research that seek to build quantitative models using social media data to assess behavior and wellbeing. For instance, the significant differences in the social media interaction received by females and males illustrates the need of holding gender as a control in such studies.

Further, we showed that the underlying behavioral differences of individuals, driven by their demographic attributes, personality traits, and wellbeing measures, are evident in their social media activity and interaction with others. Combined with the advantage that social media data can be unobtrusively obtained without active intervention, and that it captures long-term and historical human behavior, we demonstrate that social media can potentially function as a viable passive sensor in longitudinal studies of behavior and wellbeing—although the actual value of this passive sensor, especially in the context of the Tesserae study is yet to be established in future research.

Many of our empirical insights also align with the literature in psychology and behavioral science [5, 10, 12]. For instance, agreeableness is often used as an attribute of being well-mannered, and this construct aligns with our observation that the posts of highly agreeable individuals are liked and commented more than the others. Similarly, greater openness implies greater self-actualization and

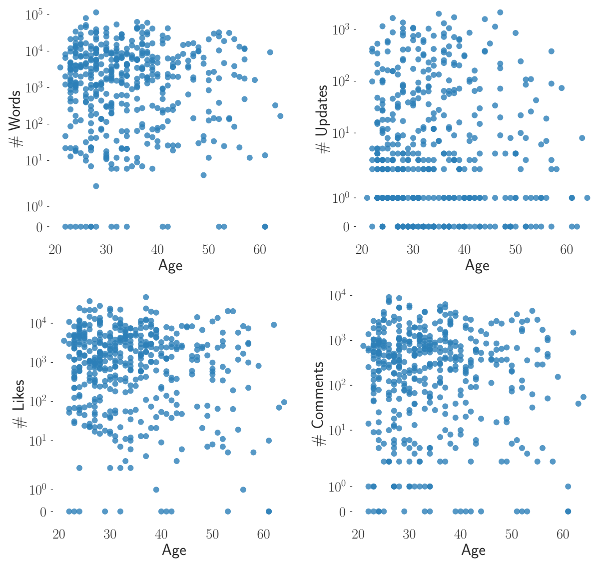

Figure 8: Relationship between age and the number of updates shared, and likes and comments received.

Acknowledgements. This research is supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No. 2017-17042800007. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. We thank Eva Sharma and Sarmistha Dutta for supporting the development of the data collection infrastructure, and all the external proctors who helped in the enrollment process.

expressiveness of intense experiences, a trait that probably drove the high openness individuals to post more updates and lengthier posts (as revealed in our dataset). Aligning with prior work [13], we also find that individuals with poorer sleep quality show greater social media activity.

Finally and importantly, we showed that it is possible to build an automated infrastructure to passively gather social media data in large-scale studies of human behavior and wellbeing. We believe this will be useful to the research community as both a validation demonstrating the feasibility of such studies in the future, as well as serve as a example case scenario where multiple developmental challenges, over the course of a long enrollment period, were tackled carefully, diligently, and in a privacy-preserving way. By releasing a sample corpus of our dataset, we envision the research community will benefit further in terms of being able to conduct replicable and reproducible research.

REFERENCES

[1] Carole Cadwalladr and Emma Graham-Harrison. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian* (2018).

[2] Ralph Catalano. 1979. *Health, behavior and the community: An ecological perspective.* Pergamon Press New York.

[3] Linda L Dahlberg and Etienne G Krug. 2006. Violence: a global public health problem. *Ciência & Saúde Coletiva* (2006).

[4] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *International AAAI Conference on Web and Social Media.*

[5] Gabriella M Harari, Nicholas D Lane, Rui Wang, Benjamin S Crosier, Andrew T Campbell, and Samuel D Gosling. 2016. Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. *Perspectives on Psychological Science* (2016).

[6] Stephen M Mattingly, Julie M Gregg, Pino Audia, Ayse Elvan Bayraktaraglu, Andrew Campbell, Nitesh V Chawla, Vedant Das Swain, Munmun De Choudhury, Sidney K D'Mello, Anind K Dey, Ge Gao, Krithika Jagannath, Kaifeng Jiang, Suwen Lin, Liu Qiang, Gloria Mark, Gonzalo J Martinez, Kizito Masaba, Shayan Mirjafari, Edward Moskal, Raghu Mulukutla, Kari Nies, Manikanta Reddy, Pablo Robles, Koustuv Saha, Anusha Sirigiri, and Aaron Striegel. 2019. The Tesserae Project: Experiences with Large-Scale, Longitudinal, In-Situ, Multimodal Sensing of Information Workers. In *CHI Ext. Abstracts.*

[7] Pew. 2018. pewinternet.org/2018/03/01/social-media-use-in-2018/.

[8] Koustuv Saha, Larry Chan, Kaya De Barbaro, Gregory D Abowd, and Munmun De Choudhury. 2017. Inferring Mood Instability on Social Media by Leveraging Ecological Momentary Assessments. *Proc. ACM IMWUT* (2017).

[9] Koustuv Saha and Munmun De Choudhury. 2017. Modeling Stress with Social Media Around Incidents of Gun Violence on College Campuses. *Proc. ACM HCI* 1, CSCW (2017).

[10] H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8, 9 (2013), e73791.

[11] Roger Tourangeau, Lance J Rips, and Kenneth Rasinski. 2000. *The psychology of survey response.*

[12] 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 *Ubicomp.* ACM, 3–14.

[13] Heather Cleland Woods and Holly Scott. 2016. # Sleepyteens: social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. *Journal of Adolescence* (2016).

[14] Xiaoyi Zhang, Laura R Pina, and James Fogarty. 2016. Examining unlock journaling with diaries and reminders for in situ self-report in health and wellness. In *CHI.* ACM.