Proposal

Mental health concerns among young adults attending universities are a significant challenge to the overall well-being of the academy. Research has indicated that university students experience higher rates of psychological distress and mental disorders than their non-student peers. Among US college students, approximately 50% experience some form of mental disorder annually. Moving away from home for the first time and starting a rigorous academic program can contribute to high stress, and stress tends to be higher at campuses with a student body that is in general more the type-A characteristic type of students. Despite growing recognition of this challenge, few university students seek help related to mental health — only 18% of students with a past-year mental disorder are known to receive treatment. This arises as a consequence of a variety of barriers, such as social stigma, paucity of knowledge about services and their location, failure to perceive themselves as in need of mental health support, and lack of time. Although many universities provision free short-term counseling in the campus, uptake of these services is far lower than would be expected on the basis of need.

This vision proposal focuses on *integrating non-conventional data sources, such as social media,* and *"big data" analytic methodologies alongside traditional mental health and well-being assessments,* which may be leveraged to understand, reason, and infer risks to *mental health and psychological distress in college student populations*.

The continual adoption of social networking sites (SNS) such as Facebook, Twitter, Tumblr, and Instagram is leading to a shift of social interactions to happen online beyond face-to-face settings. This new social interaction paradigm thus provides an opportunity to observe individuals' affective and cognitive states, as well as their access to social capital resources in a real-time, longitudinal fashion. Moreover, this real-time data stream of social information is often annotated with context including location information, cues about one's social environment, and rich collections of multimodal information beyond text, such as images and videos. Consequently, a rich body of research, including our own research, has emerged, intending to understand, model, and infer people's psyches and social milieus by utilizing continuing streams of evidence from social media. These include leveraging social media data to derive conditions and symptoms related to diseases, influenza propagation and prediction, psychological distress, drug abuse, postpartum depression, PTSD, depression, bipolar disorder and seasonal affective disorder, insomnia, eating disorders, and smoking and drinking addiction.

Note that such studies utilizing social media data and methods to understand and infer mental health risk have not specifically focused on college populations. The ubiquitous use of social media technologies in this segment makes this opportunity appropriate and timely. As of 2014, per Pew Internet, 97% of individuals in the age group 18-29 years use the Internet – a demographic that aligns with the bulk of university student population, out of which more than 95% use Facebook. Computational methods that can collect, process and analyze social media data and derive social media measures of mental health risk (e.g., suicidality) can complement state-of-the-science health assessment tools to give rich, detailed, difficult to measure observations about context and content of social interactions, accessible social capital, and exposure to online cues that may alleviate or mediate risk.

However, several significant challenges also lie toward effectively harnessing such social media data driven methods in practical settings that can directly impact health and wellbeing, specifically of college student populations. This proposed research direction intends to address a number of such challenges.

Methodology.

1. Most current efforts in using social media data for understanding health behaviors are limited in scale and methodology. We intend to develop and validate robust, scalable, principled and algorithmically sound "big data" methods that can derive generalizable markers of mental health states, out of complex social media data. Validation is important because while research has shown social media to indicate one's affective states, it is yet to investigated with rigor how this manifested affect actually relates to one's emotional status measured through self-reported or physiological response.

2. College populations use social media to not only share information about mundane and significant happenings in their daily lives, but social media also serves as a platform for humor, opinion sharing or for sarcastic reflections. In order to glean meaningful and useful information about one's health state from such data, we will develop systematic approaches that can discard noisy, irrelevant or ambiguous information that frequently appears in the social media feeds of college students.

3. Traditional assessment techniques, such as self-reported or survey instrument based methodologies, to support college student well-being are often limited in coverage, are expensive to administer, and have been known to lack a real-time component that can forecast future mental health risk. Social media data, such as measures of affect and access to social capital may help bridge such gap, however several activities may happen outside of what is measureable through social media, that can impact student health. We intend to develop innovative methods that can integrate survey and self-reported techniques with naturalistic social media data to capture comprehensive cues relating to adverse mental health events.

4. Over the past, big data investments have yielded significant new insights in genomics, electronic health records (EHR), informatics, and imaging, and newer investments are targeting solutions to mobile sensor big data. Big

data based health measurements from mobile sensors allow us to go beyond self-reported information relating to inference of health events, by measuring fine-grained physiological characteristics. However, they are often able to gauge only a limited set of cues, e.g., stress, heart rate, which are likely to be antecedents or acute precipitants of a certain health risk, and typically do not capture the short and long term rich psycho-social-interactional content and context that may be associated with one's propensity to a health risk. SNS big data complements such other big data sources, and we hope to develop careful and systematic solutions that can be integrated with the above mobile health data sources.

Data.

5. It has been recognized that social media usage in college student populations has been growing increasing complex; consequently negatively impacting the way existing methods that derive health risks from social media data may be readily and accurately applicable to this changing landscape of their use in this population of interest. For instance, per the Global Social Media Impact Study, college student populations have been known to be increasingly adopting various anonymous or semi-anonymous as well as ephemeral social networking platforms for broadcasting and sharing thoughts and events about their daily lives (e.g., Tumblr, Whisper, Snapchat, Reddit, Yik Yak, 4chan, WhatsApp). Such data is often difficult to collect (due to corporate data regulations and restrictions and ephemerality), and challenging to be associated with complementary forms of health assessment (due to anonymity and lack of identifiable information). In this proposed research, we intend to develop methodologies that can delineate relationship between typical (e.g., Facebook) and these emergent social media use and thereby mine mental health risk related information from both data sources together.

6. Next, recent observations have indicated the increasing prevalence of multimedia information in social media feeds of college populations, e.g., images and videos beyond text. We propose to extend current approaches that typically leverage textual data for deriving health markers, into including and mining meaningful cues embedded in the multimedia content shared by this population of interest.

Translation.

7. Due to the ubiquitous use of social media among college populations, social media can not only act as a continuous stream of naturalistic behavioral data, but can be leveraged to actually deliver timely, personalized, and accurate just-in-time interventions. Through this proposed research, we intend to build rigorous scientific ties in this topic to bring together social computing analytic researchers and intervention psychologists.

8. There is ongoing work into making mobile health sensing accessible to health research community. Through this proposed research direction, we hope to develop complementary efforts toward dissemination of social media big data analytics and social media health markers that can improve the accessibility of social media derived health assessment measures to the broader health research and biomedical community.

Background and Credentials of the Proposer

Munmun De Choudhury, Ph.D, is currently an assistant professor at the School of Interactive Computing, Georgia Tech. Munmun is an expert in computational social science. Her research has focused on the analysis and modeling of large-scale online data to make sense of behavior and social interactions. Specifically, Munmun is passionate about interdisciplinary approaches that can bring such big data analytic approaches to improving our well-being.

The proposed vision and identified opportunities and challenges stem from Munmun's recent extensive body of work, one of the first endeavors in the research community, focusing on developing computational, machine learning methods from our online social media that can model and infer a variety of health and well-being states, such as depression. Aside from numerous peer-reviewed publications at computational research venues, this line of research led to a Best Paper Award (2014) and two Best Paper Honorable Mention Awards (2013, 2014) at two of the premier conferences in human computer interaction (ACM SIGCHI, ACM CSCW). Her research has further invited considerable attention from the popular press, including the New York Times, NPR, the Wall Street Journal, Scientific American, and the TIME magazine. This rich and diverse experience, in both academic (Georgia Tech, Rutgers, Arizona State) and industrial research (Microsoft Research) setting has given her an opportunity to not only build ties across disciplines around practical problems, but also to translate ideas into actionable resources and software.

Since joining Georgia Tech in 2014 as a tenure-track assistant professor, Munmun has successfully competed for NIH support – she was awarded an R01 grant as PI in 2014 to develop social media based big data approaches to detect PTSD risk in crisis-embroiled communities. She is also serving as PI on four other industry, non-profit, and university funded projects, thus laying foundation of a strong research program in social media big data analytics and behavioral health. Currently, Munmun has nearly 50 peer-reviewed publications, an *h*-index of 20 and citation count of approximately 1,300. Last year she was recognized by the prestigious Berkman Center for Internet and Society at Harvard as a faculty associate, and by the Center for Data Innovation, a leading Washington DC based think-tank studying the intersection of data, technology, and public policy, as ``one of the 15 Women in Data to follow''. For more information, please refer to Munmun's webpage: http://www.munmund.net/