Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Psychology

Social media, big data, and mental health: current advances and ethical implications

Mike Conway[1] and Daniel O'Connor[2]

Mental health (including substance abuse) is the fifth greatest contributor to the global burden of disease, with an economic cost estimated to be US $2.5 trillion in 2010, and expected to double by 2030. Developing information systems to support and strengthen population-level mental health monitoring forms a core part of the World Health Organization's Comprehensive Action Plan 2013–2020. In this paper, we review recent work that utilizes social media 'big data' in conjunction with associated technologies like natural language processing and machine learning to address pressing problems in population-level mental health surveillance and research, focusing both on technological advances and core ethical challenges.

Addresses
[a] Department of Biomedical Informatics, University of Utah, 421 Wakara Way, Salt Lake City, UT, United States
[b] Wellcome Trust, Gibbs Building, 215 Euston Road, London NW1 2BE, United Kingdom

Corresponding author: Conway, Mike (mike.conway@utah.edu)

Current Opinion in Psychology 2016, **9**:77–82

This review comes from a themed issue on **Social media and applications to health behavior**

Edited by **Melissa A Lewis** and **Clayton Neighbors**

For a complete overview see the Issue and the Editorial

Available online 21st January 2016

http://dx.doi.org/10.1016/j.copsyc.2016.01.004

Introduction

Mental illness (including substance abuse) is the fifth greatest contributor to the global burden of disease [1,2]. The economic cost of mental illness was estimated to be US $2.5 trillion in 2010, and is expected to double by 2030 [3]. A core goal of the World Health Organization's Comprehensive Mental Health Action Plan 2013–20 is to strengthen information systems for mental health, including increasing capacity for population health monitoring [4]. The widespread use of social media combined with the rapid development of computational infrastructures to support efficient processing of 'big data',[1] and crucially, the maturation of Natural Language Processing (NLP) and Machine Learning (ML) technologies offers exciting possibilities for the improvement of both population-level and individual-level health. Social media is well established as a data source in the political [6], business [7], and policy [8] contexts, is increasingly used in population health monitoring, and is beginning to be used for mental health applications. Social media analysis is particularly promising in the mental health domain, as Twitter, Facebook, etc., provide access to naturalistic, first person accounts of user behavior, thoughts, and feelings that may be indicative of emotional wellbeing.

An important feature of research in this domain is that it is inherently interdisciplinary and dispersed across health journals (e.g. PubMed), psychology journals (e.g. PsycINFO), and computer science conference and workshop proceedings (e.g. Compendex).[2] This review briefly surveys social media-based applications of NLP to the mental health domain, focusing on both recent technological advances and core ethical issues from the perspective of population-level mental health monitoring.[3]

Mining social media for health

The use of social media 'big data' for health applications — particularly public health applications — is a rapidly growing area of research [10,11] variously referred to as *infoveillance* [12], *digital epidemiology* [13], and *digital disease detection* [14]. Twitter in particular, due to its public Application Programming Interface,[4] and status as a 'broadcast' social network[5] has been used for population-level influenza surveillance [16–18], monitoring mass gatherings [19,20], understanding public sentiment towards vaccination [21], building pharmacovigilance applications (e.g. post-market surveillance of adverse drug events) [22,23], understanding public attitudes towards new and emerging tobacco products and e-cigarette marketing [24,25], and investigating prescription drug abuse [26].

[1] The term 'big data' lacks an agreed definition, but one common formulation characterizes the distinction between 'big data' and more traditional data in terms of *velocity*, *volume*, and *variety* (the 'three Vs') [5].

[2] Note that in computer science, peer-reviewed conference and workshop papers, as opposed to journals, are the preferred means of disseminating research results.

[3] Note that this review *does not* focus on intervention-based studies (e.g. Facebook's 2014 'emotional contagion' intervention study [9]).

[4] Twitter offers several freely accessible Application Programming Interfaces.

[5] Twitter's open status can be contrasted with sources of internet-derived public health data like *Google Flu Trends* [15] which are not easily accessible by researchers.

Mental health and natural language processing

Mental health has been a subject of research for NLP researchers since the early days of the discipline, as evidenced by Weizenbaum's ELIZA interactive Rogerian psychotherapist program [27] (1966), and Colby's 'paranoid' conversational agent, PARRY [28] (1972). As is to be expected, the field has moved on significantly since the development of these early chatbots. Recent work uses sophisticated NLP and ML methods to, for instance, assess suicide risk in pediatric populations based on writing samples [29], predict depression severity and optimal treatment based on narrative text derived from Electronic Health Records [30], identify linguistic features characteristic of early stage dementia [31], and predict the suicide risk of active duty military personnel based on Electronic Health Record data [32]. In parallel with these advances in NLP, there is a rich tradition in the psychology domain (exemplified by Pennebaker [33]) of using carefully developed and validated lexicons organized into various categories (e.g. *anxiety*, *insight*, *achievement*) in order to score texts according to the presence or absence of psychological terms.

Social media, natural language processing, and mental health

Social media has been used extensively in marketing for *sentiment analysis* (broadly, the ascription of positive or negative emotional valence to a text [34]) and for quantifying specific personality traits or dimensions. For example, predicting 'dark triad' traits (i.e. *narcissism*, *Machiavellianism*, and *psychopathy*) from tweets [35], detecting evidence of psychopathy [36], and the identification of 'Big 5' personality dimensions from Facebook data [37••]. Specifically focused on mental health, negative-emotion language on Twitter has been shown to correlate well with official United States suicide statistics at the state level [38••].

De Choudhury

With colleagues at Microsoft Research and Georgia Tech, De Choudhury has been responsible for a pioneering series of papers on applying computational methods to the investigation of mental health issues in a number of different social media platforms, including Twitter [39••,40,41], Facebook [42], and Reddit [43,44]. De Choudhury's work has focused on developing methods for both *monitoring population health* and *identifying risk factors for individuals*. In the *population health domain*, De Choudhury *et al.* [39••] describe the creation of a crowdsourced data set of tweets derived from Twitter users with depression-indicative CES-D (*Center for Epidemiological Studies-Depression*) scores. This data-set was then used to train a statistical ML algorithm capable of identifying depression-indicative tweets and then applied to geocoded Twitter data derived from 50 US states, with results correlating well with US Centers for Disease Control depression data. In the *identifying risk factors for individuals domain*, De Choudhury *et al.* [40] investigated new mothers' experiences of postpartum depression by automatically identifying birth announcements from public Twitter data using cue phrases (e.g. 'it's a boy/girl!'), then analyzing characteristics of the new mothers' Twitter stream before and after birth, discovering that using ML techniques in conjunction with an analysis of prebirth behavior patterns can predict postnatal emotional and behavioral changes with 71% accuracy.

CLPsych conference

The CLPsych — Computational Linguistics and Clinical Psychology — workshop series has provided an important forum for computer science researchers with an interest in clinical psychology, and for research psychologists and mental health clinicians with an interest in technology. While covering a wide range of mental health applications (e.g. automatically coding therapist/patient interactions [45], and automatically quantifying autistic childrens' repetitive linguistic behavior [46]) the workshop has had a specific focus on population mental health and social media. In particular, participants at the workshop introduced a novel method for developing data sets for specific mental illnesses: pulling tweets (via the public Twitter Application Programming Interface) from users with a self-disclosed, publicly-stated psychiatric diagnosis (e.g. 'I was diagnosed with having P.T.S.D', 'she diagnosed me with anxiety and depression'). The approach was first used to generate a data set for post-traumatic stress disorder, depression, bipolar disorder and seasonal affective disorder [47••], before extending the approach to other conditions (attention deficit hyperactivity disorder, anxiety, borderline, eating disorders, obsessive–compulsive disorder, and schizophrenia) [48]. Work has focused on characterizing language associated with particular mental health conditions on Twitter using variety of methods. Mitchell *et al.* [49] investigated linguistic characteristics associated with those Twitter users who had a self-disclosed schizophrenia diagnosis, discovering that — when compared to community controls — schizophrenia sufferers were more likely to use the first person, and less likely to use emoticons and exclamation marks — findings consistent with current understanding of schizophrenia (i.e. *preoccupation with self* and *flat affect*, respectively). Using the same dataset as [47••], Preoţiuc-Pietro *et al.* leveraged NLP techniques to examine 'Big-5' personality and demographic characteristics associated with a self-disclosed diagnosis of depression or PTSD [50], finding that PTSD sufferers were both older and more conscientious than depression sufferers. Resnick *et al.* [51] used a sophisticated topic modeling ML technique to identify themes in the depression Twitter data generated by Coppersmith *et al.* [47••], and discovered that the process of aggregating tweets — that is not treating individual tweets as atomic, but rather providing more context by processing data derived from a single user in weekly chunks — substantially improved the