Dissertation

Tweeting About Mental Health

Big Data Text Analysis of Twitter for Public Policy

Mikhail Zaydman

This document was submitted as a dissertation in January 2017 in partial fulfillment of the requirements of the doctoral degree in public policy analysis at the Pardee RAND Graduate School. The faculty committee that supervised and approved the dissertation consisted of Douglas Yeung (Chair), Luke Matthews, and Joie Acosta.

PARDEE RAND GRADUATE SCHOOL

For more information on this publication, visit http://www.rand.org/pubs/rgs_dissertations/RGSD391.html

Published by the RAND Corporation, Santa Monica, Calif.

Abstract:

This dissertation examines conversations and attitudes about mental health in Twitter discourse. The research uses big data collection, machine learning classification, and social network analysis to answer the following questions 1) what mental health topics do people discuss on Twitter? 2) Have patterns of conversation changed over time? Have public messaging campaigns been able to change the conversation? 3) Does Twitter data provide insights that match the results obtained from survey and experimental data? This dissertation finds that Twitter covers a wide range of topics, largely in line with the impact that these conditions have on the population. There is evidence that stigma about mental illness and the appropriation of mental health language is declining in Twitter discourse. Additionally the conversation is heterogeneous across various self-forming communities. Finally, I find that public messaging campaigns are small in scale and difficult to evaluate. The findings suggest that policy makers have a broad audience on Twitter, that there are communities engaged with specific topics, and that more campaign activity on Twitter may generate greater awareness and engagement from populations of interest. Ultimately, Twitter data appears to be an effective tool for analysis of mental health attitudes and can be a replacement or a complement for the traditional survey methods depending on the specifics of the research question.

Table of Contents

ABSTRACT: III

LIST OF FIGURES VII

LIST OF TABLES VII

ACKNOWLEDGMENTS IX

CHAPTER 1 : INTRODUCTION 1

Policy and research questions 4

CHAPTER 2 : BACKGROUND AND MOTIVATION 6

Introduction 6
Mental health and stigma 6
State of social media analysis 9
Network and cascade analysis 10
Machine learning and supervised classification 12
Sentiment analysis 12
Topic modeling 14

CHAPTER 3 : METHODS 15

Data acquisition 16
Gathering Tweets about mental health *16*
Gathering campaign-relevant Tweets *19*
Search results and working data sets *20*
Development of coding scheme 20
Reliability of coders *25*
Automated coding 26
Model performance and selection 28
Modeling approach comparison 33
Network analysis 37

CHAPTER 4 : CHARACTERIZING THE MENTAL HEALTH CONVERSATION 40

Introduction 40
Most mental health content on Twitter is self-focused and few Tweets share mental health resources 40
Approximately 10 percent of mental health-relevant Tweets are stigmatizing 41
Network community conversations about mental health vary in types and topics of Tweets 42
Among large communities with mental health conversations, 71 percent demonstrated low levels of stigma in community conversations 44

CHAPTER 5 : CHANGES OVER TIME 47

Introduction 47
Longitudinal trends in mental health discourse on Twitter 47
Campaign Twitter presence 49
The volume of campaign-related Twitter activity is too low to assess whether the campaign activity is affecting the overall Twitter conversation about mental health *49*
Three of the four campaigns show positive signs of engaging other Twitter users to reTweet messages *51*

Engagement with the campaign is visible in the social network *54*

CHAPTER 6 : COMPARING DISSERTATION FINDINGS WITH THE LITERATURE **58**

POPULARITY OF TOPICS IN SOCIAL MEDIA TRACKS IMPACT OF CONDITIONS ON PUBLIC HEALTH 60
CONDITION SPECIFIC STIGMA INFORMATION IS LIMITED AND AGREEMENT WITH SOCIAL MEDIA IS MIXED 61
SELF-DISCLOSURE IS IMPORTANT TO INDIVIDUALS' MENTAL HEALTH, BUT THERE IS LITTLE LITERATURE ON THE KIND OF DISCOURSE PEOPLE HAVE ABOUT MENTAL HEALTH 64
COMMUNITY IDENTIFICATION AND SUPPORT HELP TO REDUCE STIGMA AND SELF-ESTEEM ISSUES, BUT DO NOT ALWAYS IMPROVE CLINICAL OUTCOMES 65
PUBLIC MESSAGING CAMPAIGNS UNDERUTILIZE TWITTER; DOD PERFORMANCE NOT AN OUTLIER 66
TWITTER APPEARS TO BE A ROBUST DATA SOURCE BUT FURTHER VERIFICATION IS NEEDED 72

CHAPTER 7 : CONCLUSIONS, POLICY IMPLICATIONS, LIMITATIONS, AND FUTURE WORK **73**

FINDINGS AND POLICY RELEVANCE 74
Twitter discourse contains mental health discourse *74*
People are getting better at discussing mental health *75*
Mental health communities exist and are heterogeneous *75*
Public messaging campaigns are performing well on Twitter but need greater volumes of content *76*
Social media data correlates with traditional research data but may lend itself to different questions *77*
Methods for analyzing text and social media are available, robust, and reproducible *78*
LIMITATIONS 78
Insufficient time and resource for methodological refinement *78*
Omitted some topics of interest *79*
Insights not differentiated by demographics *79*
Some Relevant Data Is Not Accessible *79*
Campaign effects are small and hard to capture *80*
FUTURE WORK 80
Richer measurement of mental health discourse *80*
Deeper and broader community analysis, richer longitudinal analysis *81*
FINAL THOUGHTS 82

BIBLIOGRAPHY **85**

List of Figures

Figure 3.1 Twitter Data Analysis Process 16
Figure 4.1 Distribution of Tweets by Type of Content and Topic 41
Figure 4.2 Tweets by Type and Topic 42
Figure 4.3 Characteristics of Tweet Types and Topic for Four Communities with High Proportions of Tweets About Addiction, Depression, PTSD, and Suicide 43
Figure 4.4 Variation in Proportion of Stigmatizing Tweets Across Communities 44
Figure 4.5 Characteristics of Tweet Types and Topic for Four Highly Stigmatizing Communities 46
Figure 5.1 Time Trends in Tweet Volume by Tweet Type and Topic 49
Figure 5.2 Tweet Volume by Month from Official and Unofficial Accounts 51
Figure 5.3 Twitter Engagement with the Campaigns 52
Figure 5.4 Centrality of Campaign Engaged Users within Mental Health Focused Social Network by Campaign 53
Figure 5.5 Identifying a Community of the Campaign Engaged Users 55
Figure 5.6 Centrality of Campaign Engaged Users within Mental Health Focused Social Network by Community 56
Figure 5.7 Network of Actively Engaged Users 57
Figure 6.1 Prevalence of Mental Health Conditions 60
Figure 6.2 Relative Levels of Stigma 63
Figure 7.1 Summary of Dissertation Process, Findings, Implications, and Future Directions 83

List of Tables

Table 3.1 Initial Search Strategy 17
Table 3.2 Final Twitter Search Strategy: Strings 18
Table 3.3 Final Twitter Search Strategy: Hashtags 19
Table 3.4 Campaign-Related Twitter Search Strategy 20
Table 3.5 Tweet Coding Scheme 21
Table 3.6 Inter-rater Reliability for 500 Tweets Coded by Three Coders 25
Table 3.7 SVM-Based Automated Coding Model Performance for Each Tweet Characteristic 30
Table 3.8 Volumes of Tweet Predictions in Working Data Set 33
Table 3.9 Neural Net AUC Performance 36
Table 3.10 Support Vector Machine AUC Performance 36
Table 3.11 Comparison of Support Vector Machine and Neural Network Optimum Models 36
Table 5.1 Official Twitter Accounts 50
Table 6.1 Ranking of Conditions by Journal Article 62
Table 6.2 Key Performance Indicators from Recent Twitter Public Messaging Campaigns Evaluations 68

Acknowledgments

I am grateful for the incredible support of my dissertation committee: Douglas Yeung, Joie Acosta, and Luke Matthews. They believed in my ideas and allowed me to pursue my interest. They spent countless hours mentoring me on methodological and substantive questions in addition to reading enumerable drafts and proposals. Their willingness to support and teach as well as to allow me to pursue my passions was an amazing part of my Pardee RAND experience and will serve me well in the future.

Additionally, I am grateful for the research support I received while working through this dissertation. Jennifer Cerully was an amazing and generous support through the earlier parts of this dissertation. Ilana Blum, Nupur Nanda, and Rachel Ross were wonderful research assistants who worked tirelessly to hand code my data set. I would also like to extend my appreciation to the RAND Twitter data team. In particular, Andrew Cady and Benjamin Batorsky, who dedicated many hours to helping me obtain the very large volume of data that was used in this dissertation.

This work could not have been possible without the generous financial support provided by the Arthur S. Wasserman Prize for Reducing Social and Economic Disparities, the Anne and James Rothenberg Dissertation Award, and ARA Research Support Funding. These awards allowed me to dedicate the needed attention to my research. I would like to thank the donors, Pardee RAND, and the RAND National Security Research Division.

Finally, I would like to thank the entire Pardee RAND community: peers, faculty, administration, and the many research staff who included me in their work. This community was instrumental in making the last four year such a formative and amazing experience.

Chapter 1: Introduction

This dissertation aims to demonstrate the value of social media for understanding and affecting attitudes towards mental health. This is done by using Twitter to capture the state of positive and negative attitudes across various mental health topics and identify the areas of positive change. These results are compared against traditional academic sources of research in this area to demonstrate the robustness of social media analysis. The value of understanding and working towards improving popular perceptions of people with mental health conditions is that positive change leads to a greater utilization of mental health care services and an improvement in public health[1].

An individual's attitudes and beliefs towards mental health and mental health treatment likely influence how that individual will interact with anyone who is struggling with mental illness. These interactions include close relationships, like in the workplace as well as incidental contact, as well as interaction through social media. Additionally, those beliefs and attitudes will impact how individuals perceive their own mental health and play a role in determining whether they seek care. Willingness to seek care can be a substantial barrier to mental health treatment. Given that one in four US adults struggles with some form of mental illness in a given year[2], these attitudes play a large role in public health.

Mental health stigma is the result of a combination of attitudes and beliefs that lead people to reject people with mental illness[3]. This rejection may be overt – refusing to work with someone because they have a diagnosed condition, or covert – assuming disability that extends far beyond the diagnosis. It is also something that is internalized by the individual and leads to a self-perception of 'otherness' or separation[4]. This creates a normalization of stigma where both the individual with an illness and those surrounding him or her see isolation and diminished capacity as the norm. Overall, mental health stigma can lead to social exclusion and discrimination. The effects of this are far reaching, but concretely stigma can lead to unequal access to key resources like education, employment, and community[5].

Individuals' attitudes towards mental health arise based on the sum total of the individual's personal background[6]. This means that factors such as history, family, personal experience, mass media, personal interaction, and social media interaction all play a role in determining how an individual will relate to the notion of mental health. Individuals change their attitudes over time (and may even hold multiple context specific attitudes). As such, these and other factors continue to shape and influence individual behavior[7]. As people are exposed to the various factors that shape

[1] (Corrigan, 2004)
[2] (Kessler et al., 2005)
[3] (US Department of Health and Human Services 2003), page 4
[4] (Acosta et al., 2014)
[5] (Centers for Disease Control and Prevention et al., 2012)
[6] (Ajzen, 2001)
[7] (Ajzen, 2001)

their attitudes, their attitudes may change. An aggregation of these individual changes may alter prevailing social attitudes as well. These attitudes play a role in how much stigma an individual assigns to someone else's or their own mental health status. This stigma affects the probability that a person experiencing mental illness will receive the treatment they need and the support of their community, or remain untreated for the fear of bearing a stigmatized label[8].

US veterans are population where mental health challenges are significant and mental health stigma is a major concern. The burden of disease is higher than the general population: one third of veterans returning from Iraq and Afghanistan experience severe mental health issues[9]. Yet, only 23% to 40% percent of veterans identified as suffering from serious mental health challenges reported seeking treatment and less than half of the same group indicated interest in treatment[10,11]. As part of the ongoing effort to address these challenges the White House issued Presidential Executive Order #13625 in August 2012[12] which directed relevant agencies to pursue strategies to improve mental health of veterans, active duty military and their family, invest greater resources in suicide prevention programs, and strengthen interagency cooperation. As part of this work RAND was given the opportunity to perform an evaluation of four ongoing public messaging campaigns aimed at improving the utilization of mental health services by veterans and active duty military[13]. These campaigns work to improve awareness and overcome negative perception of mental illness which are seen as barriers to veterans and active duty military getting the care they need. As the Department of Defense (DoD) is the agency most associated with this population, these campaigns will be referred to as DoD campaigns going forward. This dissertation comes from that RAND project but examines population attitudes towards mental health (as opposed to exclusively veteran and active duty populations) and uses the DoD campaigns as a case study of how to correlate applied efforts with population attitudes.

Four public messaging campaigns were analyzed. The DoD, Veterans Health Administration (VHA), and the Department of Health and Human services (HHS) are responsible for one or more of these campaigns. These are the Real Warriors Campaign (DoD), Make the Connection (VHA), Veterans Crisis Line (VHA), and National Recovery Month (HHS). DoD instituted the Real Warriors Campaign (RWC) in 2009. RWC is a large-scale multimedia public health awareness campaign focused on resilience of returning veterans. Make The Connection (MTC) is a program launched by the VA in 2011 as a public awareness campaign aimed to encourage veterans to seek any mental health treatment they need. Veteran Crisis Line (VCL) is a VA affiliated confidential hotline meant to assist veterans in mental health crisis situations. National Recovery Month, sponsored by HHS, is aimed to educate Americans about substance abuse and has been occurring every September since 2011.

[8] (Goffman, 1963)
[9] (Milliken, Auchterlonie and Hoge, 2007)
[10] (Hoge et al., 2004)
[11] (Vogt, 2011)
[12] (Obama, 2014)
[13] Forthcoming

These campaigns all include a degree of social media activity. This includes Facebook, YouTube and Twitter. This is not surprising as social media has seen application in a variety of contexts. There have been recent publications on the use of social media for teachers[14], students[15], teens[16], med-students[17], doctors[18], pharmacist[19], e-government[20], and veterans[21]. These papers find that many groups are using the digital tools and are deriving value from them. People suffering from mental illness may be more likely than other individuals to rely on social media – either as a way to engage socially while maintaining a comfortable distance and avoid stigma or as a way to overcome the marginalization and isolation they experience[22]. Social media has penetrated into many aspects of society, is a valued channel of communication for many people, and a potential source of empathy and connection for those struggling with mental illness[23]. The prevalence of social media also makes it a potentially valuable source of data to understand both current attitudes and perceptions towards mental health as well shifts in those attitudes over time.

Twitter is an excellent platform for social science research as it provides a public record of peoples' attitudes, beliefs, and activities over a long time horizon (every public Tweet ever is in a searchable archive). The longitudinal nature of Twitter enables research about how discourse is changing over time with a degree of continuous observation which would be difficult to achieve by other methods. Additionally, the public network of Twitter makes it possible to see who is influencing the conversation. Also, the content on Twitter is created by individuals for his or her own purposes and allows for large scale observation of naturally occurring dialogue that was not previously possible. Looking at this content may allow us to capture a more accurate representation of public discourse by avoiding many of the human biases that a research setting could introduce[24]. However, using Twitter may create different biases, such as issues of population representativeness and differences between online and offline behavior. Understanding the biases of Twitter requires testing the data source and conducting projects such as this one. To improve the understanding of Twitter as a data source, this dissertation captures overall sentiment, tracks longitudinal shifts, explores online community formation and examines the role of public policy intervention on this discourse. Of the four campaigns discussed above, Real Warriors and Recovery Month both have dedicated Twitter accounts, while the rest of the campaigns rely on larger institutional accounts to disseminate relevant information on Twitter. The presence of DoD campaign efforts online allows us to ask if they are working, whom they are affecting, and how they are engaging with various users.

[14] (Tess, 2013)
[15] (Kushin and Yamamoto, 2010)
[16] (Boyd, 2014)
[17] (Roy et al., 2014)
[18] (Panahi, Watson and Partridge, 2014)
[19] (Benetoli, Chen and Aslani, 2015)
[20] (Mainka et al., 2014)
[21] (Ruiz and Stadtlander, 2015)
[22] Ibid.
[23] (Park, McDonald and Cha, 2013)
[24] (Sarniak, 2015)

Policy and research questions

To address the above policy concerns three key areas are investigated: the nature of the mental health discourse online, the change in that discourse over time, and the reliability of Twitter as a data source. By understanding the online discourse the hope is to shows which topics are currently or persistently stigmatized and would benefit from greater government intervention. By analyzing the time trends in the discourse policy makers should gain insight into which areas are changing positively, and which areas may be challenges in need of greater resources. Finally, the comparison of the results found in this Twitter analysis to traditional data sources should offer insight to policy analysts and social science researchers into the value of social media data. The specific questions being answered in this work are as follows:

P1. Which mental health topics are characterized by positive and supportive discourse and which are discussed in a stigmatizing way?

P1R1. What is the type, topic, and tone of mental health content on Twitter?

P1R2. Does the type and tone of mental health focused content vary across the network communities of users posting mental health focused tweets?

P2. What are the longitudinal trends in content and sentiment of Twitter conversation around mental health and mental health treatment?

P2R1. Is there a change in the sentiment of the conversation? The topics? The type of content?

P2a. Do sentiment and topic changes correlate with activities of DoD public messaging social media campaigns?

P2aR1. Have there been changes in the online conversation that correlate with the DoD public messaging campaign efforts?

P2aR2. Does engagement with DoD social media campaigns vary across the communities of engaged Twitter users?

P3. Can social media research substitute for traditional academic research that relies on survey, secondary data, and experimental study?

P3R1. Do the attitudes towards mental health found in online conversation correlate with the mental health attitudes found in the academic literature?

In the following, the background (Chapter 2) consists of a literature review detailing the state of social media analytics as well as major findings in research into mental health stigma. The methods section (Chapter 3) presents the data collection methods, details of machine learning

methods implemented, and a discussion of the social network analysis tools utilized. The research questions are answered in Chapters 4, 5, and 6. Chapter 4 contains details on the categorization of the overall Twitter discourse by topic, type, and stigma level as well as exploring the general network structure of relevant Twitter users. Chapter 5 looks at the longitudinal trends in the Twitter discourse and analyzes the performance of the DoD public messaging campaigns at connecting with these trends. Then, Chapter 6 focuses on comparing the findings of Chapter 4 and 5 with the most comparable academic research that relies on traditional data sources, like, surveys and not social media. Finally, Chapter 7 presents the key results from this work, identifies areas of policy relevance, discusses limitations, and makes recommendations for future work.

Chapter 2: Background and Motivation

Introduction

In this chapter I present a survey of literature that informs the methods and ideas of this dissertation. I cover the state of research on mental health and stigma as motivation for the policy focus of this work. I then present a range of papers that are either cutting edge or foundational work in social media analysis. This includes machine learning techniques, sentiment analysis, and network analysis which are central to this work. Machine learning will be used to classify the large volume of data collected and identify stigma in a variation of sentiment analysis. Network analysis allows for the differentiation of individuals and communities within the social network of Twitter. Additionally, I cover methods that were crucial to my conceptualizing of this approach: topic modeling and cascade analysis even though the methods themselves were not ultimately used. All of these inform the methods that are developed in Chapter 3.

Mental health and stigma[25]

Defining mental health involves several overlapping concepts. At the highest level mental health includes the notions of mental illness and mental wellness[26], which are interrelated but distinct. Typically, popular discussion of mental health centers on mental illness. Mental illness is 'collectively all diagnosable mental disorders' or 'health conditions that are characterized by alterations in thinking, mood, or behavior (or some combination thereof) associated with distress and/or impaired functioning.'[27] The list and descriptions of mental illnesses most commonly recognized by practitioners is codified are the Diagnostic and Statistical Manual of Mental Disorders which is compiled and amended by the American Psychiatric Association[28]. Mental illness is also found to be strongly related to the development and treatment of various chronic diseases like diabetes, cancer, cardiovascular disease, asthma and obesity[29]. Despite this relationship, mental illness is still considered a distinct category of illness.

Mental wellness (sometimes called eudaimonic well-being or eudaimonia) refers to the positive description of mental health, where in 'mental health is defined as a state of well-being in which every individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community.'[30] The question of how to operationalize this positive description has been addressed in different ways. Typically the earliest answers to this question are attributed to Aristotle who described good life 'as the end result of "a virtuous activity of soul of a certain kind" (Aristotle, 1925/1998, p. 18) and described the path to eudaimonia as a conscious and lifelong active exercise of intellect and

[25] Elements of this section is based on unpublished work (Acosta et al., 2013)
[26] (2013)
[27] (US Department of Health and Human Services, 1999)
[28] (American Psychiatric Association, 2013)
[29] (Murray and Lopez, 1996)
[30] (World Health Organization, 2014)

character virtues.'[31] Later, John Stuart Mill articulated the notion of well-being as the maximum amount of pleasure (a reframing of the Aristotelian virtue) with the least amount of pain[32]. This formulation was adapted by social physiology with happiness being defined as the balance between positive and negative affect[33]. However, it was found that positive and negative affect were not independent phenomena[34] which led to the rejection of the theory. The most prevalent modern measure of wellness is the Scales of Psychological Well-Being[35]. However, other dimensions such as emotional well-being and social well-being are often considered[36].

Psychological well-being is an empirically tested measure of individuals' capacity for positive functioning. This measure is based on six areas: 'positive evaluations of oneself and one's past life (Self-Acceptance), a sense of continued growth and development as a person (Personal Growth), the belief that one's life is purposeful and meaningful (Purpose in Life), the possession of quality relations with others (Positive Relations With Others), the capacity to manage effectively one's life and surrounding world (Environmental Mastery), and a sense of self-determination (Autonomy).'[37] Emotional well-being focuses on capturing the subjective metrics of goal satisfaction, coping strategies, and disposition. Emotional well-being relates these areas to overall life satisfaction[38]. Additionally there are conceptions of well-being that are defined by external facing features of a person's life as opposed to their inner health, this is often referred to as social well-being[39].

Stigma affects all of the elements of mental health discussed above. Stigma refers to the process that results in negative perceptions and discriminatory behavior towards specific individuals. Goffman describes the process of stigma creation in terms of an individual's performance (actual social identity) in comparison to social expectations (virtual social identity).

> While a stranger is present before us, evidence can arise of his possessing an attribute that makes him different from others in the category of persons available for him to be, and of a less desirable kind--in the extreme, a person who is quite thoroughly bad, or dangerous, or weak. He is thus reduced in our minds from a whole and usual person to a tainted, discounted one. Such an attribute is a stigma, especially when its discrediting effect is very extensive [.][40]

Goffman's classic definition focuses on public stigma, the process by which individuals are rejected by other people. Current literature also discusses institutional and self-stigma[41]. Institutional stigma refers to the process by which policies of institutions may intentionally or

[31] (Archontaki, Lewis and Bates, 2013)
[32] (Mill, 1901)
[33](Bradburn, 1969)
[34] (Diener et al., 1985)
[35] (Ryff, 1989)
[36] (Centers for Disease Control and Prevention: Program Performance and Evaluation Office, 2013)
[37] (Ryff and Keyes, 1995)
[38] (Diener et al., 1999)
[39] (Keyes, 1998)
[40] (Goffman, 1963)
[41] (Acosta et al., 2014)

unintentionally restrict people with mental illness[42]. Self-stigma is the process by which individuals suffering from mental illness internalize socially prejudicial attitudes and grow to accept and normalize their own reduced status.[43]. Self-stigma can significantly impact treatment outcomes as well personal feelings of an individual[44]. Both self-stigma and social stigma can create social isolation[45]. As such, stigma impacts all of the above discussed facets of mental health.

Attitudes are typically defined as a "summary evaluation of a psychological object captured in such attribute dimensions as good-bad, harmful-beneficial, pleasant-unpleasant, and likable-dislikable."[46] It is fundamentally an evaluative construct. While attitudes themselves fail to describe specific actions of an individual, they do play a role in determining the likely outcome in an aggregate set of actions[47]. The expectancy-value model, the most prevalent framing of attitude formation, argues that this evaluation arises spontaneously based on the full prior experience of the individual[48]. By the theory of planned behavior, attitudes so formed become part of the forces (along with subjective norms and perceived behavioral control) shaping the likely course of action of individual[49].

The literature on attitudes towards people with mental illness (PWMI) demonstrates the severity and prevalence of stigma. PWMIs are perceived as dangerous and violent[50,51,52]. In a 2013 study, 46% of respondents agreed that "people with mental illness are, by far, more dangerous than the general population"[53]. There is also the common perspective that the individual is responsible for their mental illness[54,55], and that stress and lack of willpower leads to mental health problems[56]. This perspective may explain why some people feel that PWMIs are inferior[57], that they are to blame[58], and are a burden to society[59]. These attitudes, along with a plethora of other negative associations (unattractive[60], unreliable[61], irresponsible[62], incompetent[63], etc.) create an "us vs. them"

[42] (Corrigan and O'Shaughnessy, 2007)
[43](Davey, 2013)
[44] (Livingston and Boyd, 2010)
[45] (Yanos, Roe and Lysaker, 2010)
[46] (Ajzen, 2001)
[47] (Epstein, 1983)
[48] (Ajzen, 2001)
[49] (Ajzen, 1991)
[50] (Schomerus et al., 2012)
[51] (Kassam et al., 2011)
[52] (Reavley and Jorm, 2011)
[53] (Barry et al., 2013)
[54] (Spagnolo, Murphy and Librera, 2008)
[55] (Bathje and Pryor, 2011)
[56] (Angermeyer and Dietrich, 2006)
[57] (Corrigan, Kerr and Knudsen, 2005)
[58] (Boardman et al., 2011)
[59] (Beldie et al., 2012)
[60] (Wood and Wahl, 2006)
[61] (Lakeman et al., 2012)
[62] (Barke, Nyarko and Klecha, 2011)
[63] (Lakeman et al., 2012)

distinction between people with and without mental illness[64]. The distinction may reinforce the belief held both by people with and without mental illness that treating PWMI negatively is normal and acceptable[65].

Attitudes towards treatment are also important for the improvement of public health. There is uncertainty in the literature on how people view mental health treatment. In the 2006 General Social Survey 97% of respondents believed that mental health conditions improve with treatment. But in 2013 a survey showed that only 56% of people agreed that "most people with serious mental illness can, with treatment, get well and lead productive lives"[66]. There are studies that indicate that the public has positive opinions about counseling, but negative attitudes towards pharmaceutical treatment[67,68]. This trend suggests that there is a lack of consensus on the efficacy of treatment and this may create attitudes of treatment being futile. However, there is consistent evidence of improvement in attitudes over time[69,70] and that attitudes differ by condition[71].

State of social media analysis

Social media is pervasive in American life. 62% of all American adults (aged 18 and older) use Facebook, with the majority of users (70%) using the site daily with a large portion (43%) using the site multiple times a day. While Facebook is by far the largest social media platform in terms of raw users and user penetration, there many others: 26% of adults use Pinterest, 21% use LinkedIn, and 23% use Twitter.[72] For this work I will be focusing on using Twitter to find policy relevant insights. While Twitter is smaller than other services it is still very large and has several advantages. Twitter is a microblogging service, users post short (140 or fewer characters) blog posts meant to be read by anyone with an interest in the content or author. This content is designed to be public, carrying no expectation of privacy, which allows researchers to search and analyze the data while complying with human subjects' protection protocols. Additionally, the volume of content is enormous — 500 million posts per day[73] — and fully accessible by researchers.

Social media can be used for analyses with many policy implications. Insight has been gained from Twitter, traditional blogs, Weibo, Instagram, Facebook, FourSquare and many others. These have covered the emergence of protest networks[74], disease surveillance[75], air pollution

[64] (Schulze et al., 2003)
[65] (Link et al., 1989)
[66] (Barry et al., 2013)
[67] (Angermeyer and Dietrich, 2006)
[68] (Crisp et al., 2005)
[69] (Mojtabai, 2007)
[70] (Schnittker, 2008)
[71] (Lincoln et al., 2008)
[72] (Duggan, 2015)
[73] (Oreskovic, 2015)
[74] (Tremayne, 2014)
[75] (Corley et al., 2010)

monitoring[76], calorie intake analysis[77], prediction of onset of mental illness[78], and food deserts[79]. Mostly these papers utilize a combination of network analysis, topic modeling, cascade analysis, and supervised classification.

As social media is a new and rapidly changing phenomenon (Twitter was founded in 2006 but did not experience explosive growth until early 2009[80]) there has not been a centralized and incremental accumulation of the best methods for understanding the data. However, there is a diverse range of work on improving social media analyses. These typically fall into a few categories – how to generate computational efficiency[81,82], how to ensure data robustness[83], and philosophical and conceptual implications of social media[84]. However, the lack of meta analyses or even meta-topics means that for the purposes of my dissertation I will be focusing on the techniques used in specific articles which contain well designed and well regarded research. Below, I discuss the techniques that I found to be repeatedly used in effective social media analysis: network analysis, machine learning, sentiment analysis and topic modeling.

Network and cascade analysis

Network analysis, the study of relationships between individual entities using graph theory has had numerous applications – disaster response[85], transmission of HIV[86], key actors in terrorist networks[87], key actors in academic networks[88], protein interactions[89], transportation[90], etc. It is also an obvious tool for making sense of social networks like Twitter. Twitter's tipping point for growth is often considered to be the 2007 SXSW festival in March[91], and possibly the first network analysis came out in August of the same year[92]. This early analysis tackled many of the same questions that subsequent authors explore – network properties like centrality, geographic distribution, and user intentions within networks. The major difference is that in 2007 a paper could hope to capture the entire social graph of Twitter. Three years later a paper touts the

[76] (Wang, Paul and Dredze, 2015)
[77] (Mejova et al., 2015)
[78] (De Choudhury et al., 2014)
[79] (De Choudhury, Sharma and Kıcıman, 2016)
[80] (Frommer and Angelova, 2009)
[81] (Bakshy et al., 2011a)
[82] (González-Bailón et al., 2014)
[83] (Ruths and Pfeffer, 2014)
[84] (McFarland and Ployhart, 2015)
[85] (Chatfield and Brajawidagda, 2012)
[86] (Liljeros et al., 2001)
[87] (Koschade, 2006)
[88] (Jonnalagadda, Peeler and Topham, 2012)
[89] (Jeong et al., 2001)
[90] (Guimera et al., 2005)
[91] (Douglas, 2007)
[92] (Java et al., 2007)

speediness of a particular algorithm for processing a small portion of the network data with a Cray supercomputer[93].

There are two main ways to construct a network on Twitter. The platform allows users to form relationships of followers (unidirectional) and friends (bidirectional)[94,95]. Mapping the network graph of these connections can illuminate who is able to see what content. This allows researchers to observe how content spreads across the network. Alternatively, a network can be built by analyzing the actual communication between users[96,97]. This includes messages sent to (@) someone as well as reposts (re-tweets) of someone. What these different types of interactions actually imply is not solidified. For example a re-tweet does not necessarily imply agreement, and sharing in general does not imply endorsement but could be conversational or informational in a specific context[98]. Any network research has to grapple with the meaning of these kinds of communications. However the communication structure is defined, various features of the network can be used to describe the dynamics of the relationships captured. Common analytic tools include degree centrality – how many entities (or 'nodes' in network analysis terminology) a specific user is directly connected to; betweenness centrality – the proportion of shortest paths in the network that pass through a particular node; closeness – The mean distance between the node and every other node in the connected component; and eigenvector centrality – a measure that reflects how many highly connected nodes a given node connects to.[99,100,101]. The various centrality measures allow for a description of the network structure as well as identification of nodes which have the capacity to impact the conversation the most. Eigenvector centrality is particularly useful for this as it considers the network several degrees removed and gives a sense how far a message from a given node could theoretically be transmitted.

A specific analytic approach of interest to this work is cascade analysis — the analysis of how content propagates through a social network. This analysis has been used to predict what content is likely to spread widely over a communication network[102], identify key nodes for diffusion (which may not be present in many networks)[103], and understand where to intervene to facilitate the spread of useful emergency information and prevent the spread of false rumors[104]. The typical approach for conducting cascade analysis on Twitter is to map the network (either through follower lists or from content) and then identify the appearance of specific content (a URL, an image, or a tweet) chronologically within the network. In conducting the analysis it is difficult to rule independent co-generation of content in the network, however it is unlikely that such a

[93] (Ediger et al., 2010)
[94] (Java et al., 2007)
[95] (Salathé and Khandelwal, 2011)
[96] (Cheong and Cheong, 2011)
[97] (Tremayne, 2014)
[98] (Freelon, 2014)
[99] (Jonnalagadda, Peeler and Topham, 2012)
[100](Freeman, 1979)
[101] (Bonacich, 1972)
[102] (Galuba et al., 2010)
[103] (Bakshy et al., 2011a)
[104] (Hui et al., 2012)

situation occurs frequently[105]. These studies have had relevant findings, for example, the size and type of network matters in the initial spread of content[106], that for content to be taken up by a user repeated exposure is valuable[107] and that is very difficult to predict which user will cause a cascade[108].

Machine learning and supervised classification

Machine learning refers to the computer application of induction algorithms, which take specific instances as input and produce a model that generalizes beyond those instances[109]. There are two commonly discussed approaches to machine learning: supervised and unsupervised. Supervised learning describes techniques used to learn the relationship between independent attributes and a specific dependent outcome[110]. Unsupervised learning describes methods that organize data based on independent attributes without a specific outcome variable[111].

Supervised learning is commonly used to understand online text, including Twitter. A classic example is the detection of spam messaging: it is impossible to label all conceivable spam content but by labeling example emails as either spam or not spam an algorithm can identify future messages as more similar to either category and apply the correct label[112]. This same approach has been used to identify tweets related to alcohol consumption[113], apply general labels (like sports and news) to trending Twitter topics[114], and many others[115,116].

Sentiment analysis

Sentiment analysis is a term for the computational treatment of opinions, feelings, and subjectivity of texts[117]. The goal is to systematically extract value judgments of text in a way that is more consistent and replicable than a holistic human judgment. This method comes from a psychological tradition dating back to Freud where the choice of language is inherently meaningful (e.g., Freudian slips) and has since been extended to more sophisticated analysis of terms, grammar, and reference frames[118]. This approach may be applied to documents, sentences or

[105] (Bakshy et al., 2011b)
[106] (Lerman and Ghosh, 2010)
[107] (Romero, Meeder and Kleinberg, 2011)
[108] (Bakshy et al., 2011a)
[109] (Kohavi and Provost, 1998)
[110] Ibid.
[111] Ibid.
[112] (Benevenuto et al., 2010)
[113] (Aphinyanaphongs et al., 2014)
[114] (Lee et al., 2011)
[115] (Cole-Lewis et al., 2015)
[116] (Sriram et al., 2010)
[117] (Pang and Lee, 2008)
[118] (Tausczik and Pennebaker, 2010)

entities[119] (specific objects within sentences) and of course Tweets. One of the earliest notable examples of sentiment analysis on Twitter came a year after the start of Twitter's explosive growth with the analysis of Tweets on the day of Michael Jackson's death[120] and found a spike in sadness words over Standard English language usage.

Sentiment analysis can be done in three ways: the lexical approach, the machine learning approach and a hybrid approach. Lexical approach refers to the use of a dictionary coded for certain sentiment features – positivity/negativity, anger, fear, joy, etc.[121] The text to be analyzed is split into either individual words or groupings (referred to as unigrams, bigrams, n-grams depending on size). The n-grams are matched up against the dictionary and the values of all the grouping are totaled up to identify positivity, negativity, or any other coded emotion score. More sophisticated lexicons will incorporate part of speech tagging, giving different sentiment values to words when they are used as a noun, a verb, an adjective, etc. This approach has been used to evaluate sentiment in anonymous online disclosure of mental health status[122], and to correlate anger on Twitter posts with community rates of heart disease[123], among others.

The machine learning approach to sentiment analysis is simply a supervised classification method as described above. A sample of the content is first manually coded for sentiment features of interest. Then a classifier is built to correctly identify the coding based on the presence or absence of text elements (words, punctuation, author, anything inherent to the text). Finally, the constructed classifier is used for text that has not been manually coded[124]. The advantage of supervised classification is that it will always be domain specific whereas most lexicons are developed for general study of linguistics. Supervised classification can be used to distinguish whether people are expressing concern over someone else's illness or disclosing their own,[125] whereas a lexical approach grounded in linguistic theory may be better able to distinguish concern from a different emotion such as anxiety. Twitter often contains non-standard English that evolves rapidly (e.g., abbreviations, intentional misspellings) which can be parsed by a supervised classification approach and would be difficult to interpret with a lexicon-based method.

It is also possible to conduct sentiment analysis with a hybrid approach[126]. A lexical approach is first applied to get a basic polarity of a text (e.g., tweets).Then, any opinion words (those with a non-zero emotional content according to the lexicon) are stripped out, and a classifier is applied to predict positive and negative tags based on the remaining domain specific features. In the case of Twitter, domain specific features would be emoticons, abbreviations, and

[119] (Pawar, Shrishrimal and Deshmukh)
[120] (Kim et al., 2009)
[121] (Pawar, Shrishrimal and Deshmukh)
[122] (De Choudhury and De, 2014)
[123] (Eichstaedt et al., 2015)
[124] (Liu and Zhang, 2012)
[125] (Ji et al., 2015)
[126] (Mohammad, Kiritchenko and Zhu, 2013)

misspellings[127]. This approach allows sophisticated linguistic analysis to be combined with a problem specific classifier, creating a richer model.

Topic modeling

Topic modeling is an unsupervised machine learning algorithm to discover topical groupings in text data. This is done by creating grouping that contain often-related words and separating groupings of rarely related words. Using this approach is difficult with Twitter data as every tweet is short, so a given 'document' does not contain much information. A common topic modeling technique used with social media data is Latent Direchlet allocation (LDA)[128,129,130,131], first proposed in 2003[132]. LDA is based on a generative model of linguistics. The goal is to model text as a set of topics with distinct probabilities, and a set of terms within each topic. LDA imagines text to be generated as follows: an author that chooses to construct a document selects a random topic from a set of possibilities based on set probabilities and then begins to draw random terms that make up that topic, and in so drawing words the author generates a text. LDA is a statistical method that uses the Direchlet distribution (a multinomial probability distribution) to model the probability of a given topic and of a given word in that topic. In LDA the parameters of the Direchlet function are optimized to maximize the likelihood that the text that is being analyzed was generated by said distributions.

LDA has been extended and customized many times. There have been demonstrations of the utility of marrying LDA with author topic models where tweets from a single author are grouped into a single document to create a more robust model fit[133]. There are methods to analyze topic emergence over time using Markov-chain processes as well as a Bayesian approach of treating time slices of topics as priors for future topic models[134]. LDA has been modified to maximize information gain in health related discourse[135], and time dimensions have been added to understand health status (treating topics of being sick, getting sick and having been sick as distinct)[136]. Even the simplest applications of LDA can be useful to generate hypotheses about the underlying structure of Twitter conversation.

[127] (Zhang et al., 2011)
[128] (Paul and Dredze, 2011)
[129](Chen et al., 2015)
[130] (De Choudhury and De, 2014)
[131] (Hong and Davison, 2010)
[132] (Blei, Ng and Jordan, 2003)
[133] (Hong and Davison, 2010)
[134] (AlSumait, Barbará and Domeniconi, 2008)
[135] (Paul and Dredze, 2012)
[136] (Chen et al., 2015)

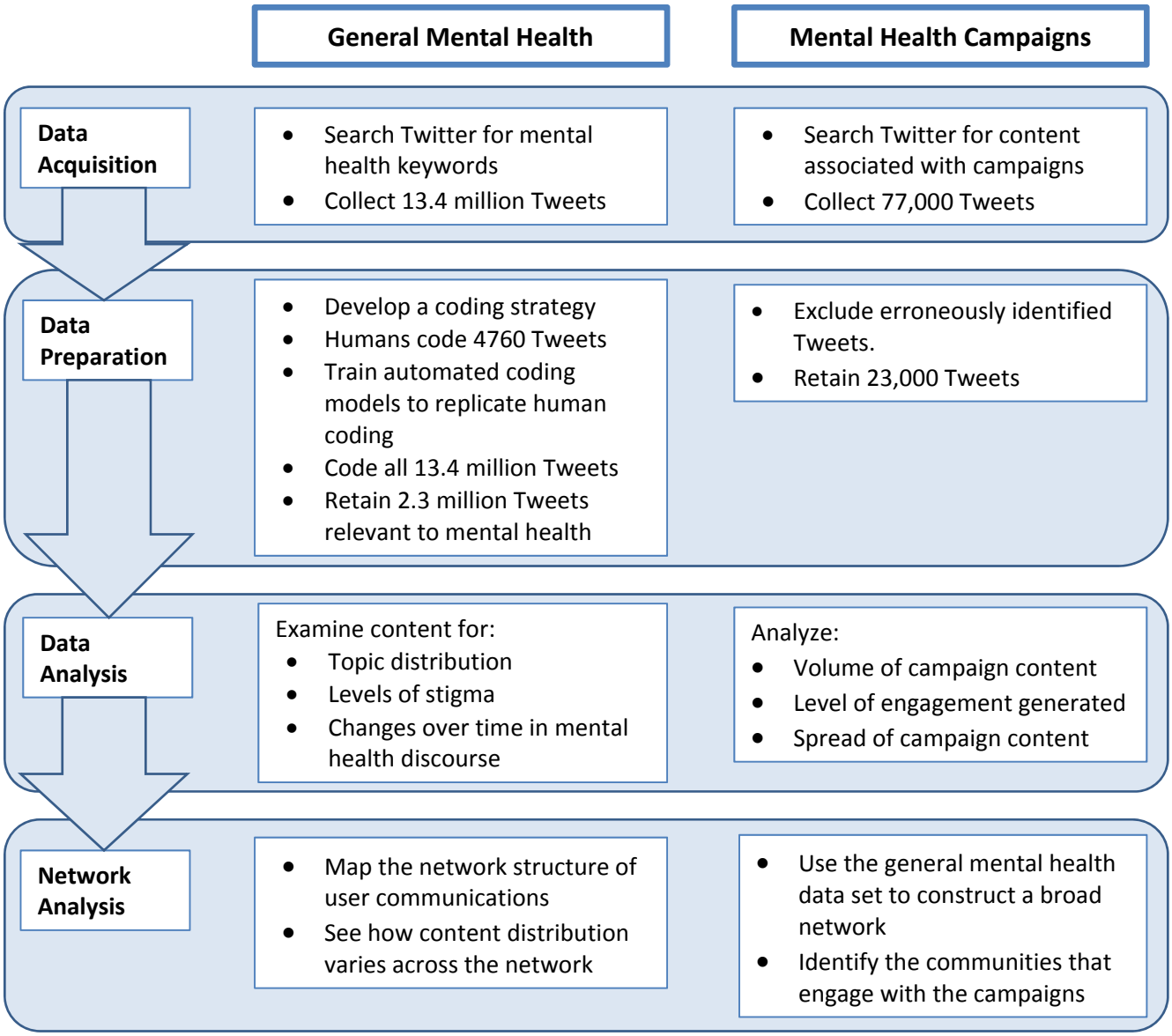
Chapter 3: Methods

This social media analysis follows two parallel streams (see Figure 3.1). To analyze general mental health discourse on Twitter, I developed a search strategy of mental health related keywords, sampled approximately 100 days between January 2009 and January 2016, and collected 13.2 million Tweets. I then developed a qualitative coding scheme to describe the content of the general mental health Tweets. A sample of Tweets was coded by members of the research team. Then I applied machine learning algorithms to allow a computer to code the remainder of the Tweets. This allowed me to classify the categories of content that are shared online, how stigmatizing the content is, and how the relative volumes of these categories shift overtime. Finally, I used the machine learning filtered data to perform network analysis to identify "communities" using instances of Twitter users communicating directly with each other.

In parallel, I developed a search strategy to collect campaign related Tweets. The campaign search covered January 2012 through January 2016 and sampled 200 days. The timeline for the campaign search was selected to cover the time period when all four campaigns were active (MTC began public messaging in late 2011). I then filtered and cleaned the campaign data and answered questions about volumes of campaign content and the level of popular engagement with the campaigns on Twitter. Finally, I used the resulting data set of general mental Tweets to identify which communities the campaigns are reaching.

By having these two data sets I was able to create metrics of the online discourse on mental health and measure the activity of the campaigns. Once the two data sets were analyzed, I combined the two to draw insight into how the campaign activity related to the overall conversation and what elements of the conversation might be useful for improving the design and activity of the campaigns.

Figure 3.1 Twitter Data Analysis Process



Data acquisition

Gathering Tweets about mental health

To understand Twitter discourse about mental health and the campaigns, I first developed a set of search terms to identify Tweets related to mental health. I then used this search strategy to sample 200,000 Tweets per day for 100 time points, approximately one day per month from January 2009 to February 2016[137]. I sampled one random day per month, and two days for May and

[137] The number of days sampled was the maximal possible given the budget and the cost structure of the data as managed by Twitter Corporation. The daily volume of Tweets collected was selected to maximize data size while remaining within bounds for budget, processing capacity, and reasonable query run time.

September. The additional days were selected to correspond to Mental Health Month in May and National Recovery Month in September. In 2009, I sampled 3 days per month because the data volumes were low at the start of the year and gradually grew to where 200,000 Tweets could be reliably collected in a single day. I also identified campaign-related terms and accounts and sampled 200 days of Tweets by and about the campaigns. I sampled 4 days a month from January 2012 through January 2016. For the majority of the analyses I focus on the text content of Tweets, but date information (along with other metadata) is retained and used for longitudinal analysis. Twitter data was gathered using Gnip, a commercial data gathering tool. The details on how these steps were performed are below.

To collect data on the overall mental health discourse I created search strings designed to capture Tweets with mental health-relevant content. These strings were developed based on the expertise of my collaborators and a review of literature and other sources that discuss language used to describe mental health and illness. The initial set of search terms (see Table 3.1) included common terms for general mental illness, as well as disorder-specific language.

Table 3.1 Initial Search Strategy

Hashtags Included in Search	Strings Included in Search
#mentalhealth	Mental health
#depression	depression
#worldmentalhealthday	stigma
#WMHD2015	eating disorder
#nostigma	suicide
#nostigmas	ptsd
#eatingdisorders	mental illness
#suicide	addiction
#ptsd	bipolar
#mentalhealthawareness	
#mentalillness	
#stopsuicide	
#IAmStigmaFree	
#suicideprevention	
#MH	
#addiction	
#bipolar	
#stigma	

Using the initial set of search terms, I drew a sample of 400,000 Tweets from three days of activity. I cleaned the content of the Tweets by removing non-roman letter characters and combining variants of words to a single root (e.g., reducing ‘sandy’ to ‘sand’). Using the cleaned sample, I computed the relative frequencies of individual words and used the R package *qdap* to

compare those frequencies to the distribution found in a reference collection of 1 million Tweets[138]. By comparing the two sets of frequencies, I identified the words that appeared 70 times more frequently in the sample of Tweets than in the reference corpus. These words were reviewed by my collaborators and, if deemed relevant to mental health, were retained as search terms in the final search strategy. If it was unclear whether the word was relevant to mental health, it was retained in the final set of search terms with the understanding that the classification algorithm would eliminate unrelated content. After this review, I also added recovery-focused language in addition to disorder-focused language. The literature I reviewed included negative terms used by teenagers to describe individuals with mental illness[139] as well as several phrases recommended by advocacy groups to describe individuals with mental illness[140]. Using this updated search strategy I collected another sample of Tweets and identified problematic terms. I found a number of terms that returned very large volumes of Tweets that were not actually relevant to mental health. As such I chose to remove the search terms 'crazy', 'stressed', 'stress' and 'freak'. The final search strategy is presented in Table 3.2 and Table 3.3.

Table 3.2 Final Twitter Search Strategy: Strings

Abusive	Distressed	Lived experience	No-one upstairs	Psycho	Stressed
Addict	Distressing	Living with addiction	Not all there	Psychopath	suicide
Addiction	Disturbed	Living with alcoholism	Not quite there	ptsd	Therapist
Alzheimers	Disturbing	Living with depression	Not the sharpest knife in the drawer	Recovery	Therapy
Asylum	eating disorder	Living with PTSD	Numscull	Retard	Wheelchair jockey
Autism	Escaped from an asylum	Loony	Nutcase	Schizo	Window licker
Bipolar	Few sandwiches short of a picnic basket	Loony bin	Nuts	Schizophrenia	You belong in a home
Bonkers	Freak	Lunatic	Nutter	Schizophrenic	
Brain damage	Gone in the head	Madness	Nutty as a fruitcake	Screw loose	
Brain dead	Halfwit	Manic depression	OCD	Screwed	
Breakdown	Hallucinating	Mass murderers	Off their rocker	Self-control	
Coping	Hallucinations	Mental	Operational stress	Self-determination	
Crazy	Hand fed	mental health	Out of it	Self-harm	

[138] (Sanders, 2015)
[139] (Rose, 2007)
[140] (Disability Rights California, 2014)

Demented	Handicapped	Mental health challenges	Padded cells	Shock syndrome
Depressed	Head case	Mental hospital	Paranoid	Sick in the head
Depression	Hurting yourself	Mental illness	Pedophile	Simpleton
Deranged	In recovery	Mental institution	Perverted	Split personality
Difficulty learning	Insane	Mentally challenged	Psychiatric	Stigma
Dignity	Intellectually challenged	Mentally handicapped	Psychiatric health	Strait jackets
Disabled	Learning difficulties	Mentally ill	Psychiatrist	Stress

Table 3.3 Final Twitter Search Strategy: Hashtags

#abuse	#eatingdisorders	#nostigma	#presspause
#addiction	#endthestigma	#nostigmas	#mentalhealthmatters
#alzheimers	#IAmStigmaFree	#1SmallAct	#ocd
#anxiety	#mentalhealth	#psychology	#suicideprevention
#bipolar	#pts	#mhchat	#therapy
#bpd	#anxiety	#schizophrenia	#trauma
#Operationalstress	#therapy	#ptsd	#WMHD2015
#mhsm	#endthestigma	#psychology	#worldmentalhealthday
#trauma	#AA	#schizophrenia	#stress
#spsm	#mentalhealthmatters	#stigma	#wellbeing
#alcoholism	#mentalhealthawareness	#stopsuicide	#adhd
#depressed	#mentalillness	#suicide	#bpd
#depression	#MH	#shellshock	

Gathering campaign-relevant Tweets

In order to identify Tweets related to the campaigns being evaluated, I searched for the Twitter handles of all campaign accounts, the campaign names, and campaign-associated hashtags (see Table 3.4). MTC and VCL do not have campaign-associated Twitter handles and rely on the accounts of various government entities (such as the VA) to disseminate their message. I identified the key accounts that are intended to disseminate the content of those two campaigns, but I did not actively sample the data of those accounts because the majority of the content shared on those accounts is not focused on the campaigns being evaluated. I randomly sampled four days per month from January 2012 to January 2016 for a total of 200 days.

Table 3.4 Campaign-Related Twitter Search Strategy

Real Warriors Campaign	**@realwarriors**	
Make the Connection		**#ConnectWith**
Veterans Crisis Line		**#VeteransCrisisLine, #ThePowerof1**
National Recovery Month	**@RecoveryMonth**	**#RecoveryMonth**

Search results and working data sets

I collected a working data set of 13.4 million Tweets possibly related to mental health by applying the search strategy across 128 days, from January 2009 through February 2016. I intended to sample 100 days but the lower data volumes in 2009 (when Twitter was still primarily an early adopter network) meant that I sampled time periods greater than a single day to get more than 10,000 Tweets (an approximate volume where, based on prior experience, I felt that some classification could be done). Once I used the machine learning methods developed in the analytic step of this work to code every Tweet for every variables of interest I was able to filter the data to 2.3 million Tweets which were relevant to mental health.

For the campaign-relevant Tweet data set, I collected 77,000 Tweets across 200 days between 2012 and 2016. After manual examination of the data set, I realized that the phrase "Make the Connection" is both the name of one of the campaigns of interest as well as a common colloquial phrase. This resulted in over 50,000 Tweets that were incorrectly identified as campaign-related but that were, in fact, parts of unrelated conversations. To address this I excluded all Tweets that were selected into the data set based on the 'Make the Connection' string which did not also include the roots 'vet' or 'mil' in the Tweet. The resulting body of 23,000 campaign-relevant Tweets came from 196 days sampled from January 2012 through December 2015. Among the campaign-related Tweets, 2,800 originated from official government-affiliated accounts that publicize campaign content, and 20,800 from users who were not identified as having an affiliation with campaign-related channels.

Development of coding scheme

Having used the string-based search strategy to establish working data set of 13.4 million Tweets that were likely to be relevant to mental health, I wanted to filter the data to determine which were relevant and which were erroneously identified, as well as characterize the content. I first developed a hand coding scheme for categorizing the content of the Tweets. Building a hand coded example data set is how large volumes are typically classified with the use of supervised machine learning[141],[142]. I identified the kind of Tweet characteristics that were of interest,

[141] For a definition of supervised learning see Kohavi, Ron, and Foster Provost. "Glossary of terms." *Machine Learning* 30.2-3 (1998): 271-274.

developed a guideline to identify those characteristics, and then had a team of three coders apply the qualitative coding guidelines to a subset of Tweets (n = 4760). I worked with the coding team iteratively, clarifying and refining the coding guidelines while working to maximize inter-rater reliability. Table 3.5 contains the guidelines and examples of coding for mental health relevance, type of content, mental health stigma, and topic. The 4760 Tweets constituted a training and validation data set for the machine learning algorithms. The overall volume was the maximum that could be coded within the time constrains. The goal was to code a volume which would allow for over 20 examples of each category as I hoped that this would increase the chances of building a successful model for identification of those categories. Table 3.6 includes the average volume of instances actually found for each characteristic.

Table 3.5 Tweet Coding Scheme

Mental health-relevance			
Relevance	• Ambiguous, colloquial mental-health related language (somewhat relevant) • Explicit, clinically appropriate mental-health related language (highly relevant)	• Not relevant = 0 • Relevant (somewhat or highly) = 1	• RT @psychologicaI: Talking to your BEST FRIEND is sometimes all the therapy you need • RT @Geli: Another bright light snuffed out by the deadly disease of addiction... There is help! #RipCoryMonteith #GoneTooSoon
Type of Content (only applied for Tweets coded as mental health-relevant)			
Appropriation	• Misuse of mental health language - describing non-mental health related states, describing concepts or things.	• Absent = 0 • Present = 1	I am becoming a tums addict again
Information	• Information, data, opinion, advocacy info on mental health • Impartial and factual information related to mental health	• Absent = 0 • Present = 1	We need #PeerSupport workers in all MH team who are the voice that speaks positively of hope and recovery, regardless of what has gone before

[142] For a classic example of supervised learning for spam message identification and the need for training data, see Benevenuto, Fabricio, et al. "Detecting spammers on twitter." *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)*. Vol. 6. 2010.

Mental health resources	• References to resources that are useful to individuals suffering mental health conditions	• Absent = 0 • Present = 1	RT @EcheMadubuike: Suicide Hotline: 1-800-273-8255 a simple reTweet, might save someone's life.
Other-focused	• Tweets that are discussing specific person or persons other than the author • A general abstract 'other' is not coded in this category	• Absent = 0 • Present = 1	Even Though Jordan Is Like 3 Days From Insane, She's The Best Girl I Know.
Self –focused	• Tweets that discusses the author	• Absent = 0 • Present = 1	Still #depressed. Can't bring myself to do anything #bipolar
Topic (only applied for Tweets coded as mental health-relevant)			
Addiction	General discussion of addiction states	• Absent = 0 • Present = 1	Addiction is a serious disease; it will end with jail, mental institutions, or death if you do no get professional help.
Anxiety	Anxiety as mental health status or anxiety as descriptor	• Absent = 0 • Present = 1	RT @BenBuckwalter: Who struggles with depression or anxiety? I feel like more people do than we all realize.
Autism	Autism disorder or autism as a descriptor	• Absent = 0 • Present = 1	RT @DrBrocktagon: Fantastic pulls-no-punches article from Mayada Elsabbagh on A Global Vision for Autism Research https://t.co/UUgxv0NjdQ
Bipolar disorder	Bipolar disorder or bipolar disorder as descriptor	• Absent = 0 • Present = 1	@CamzCaplaya yeh..ppl are bipolar with there opinions of her
Depression	Depression as disorder or depression as descriptor of emotional state	• Absent = 0 • Present = 1	The only problem is that staying in leads to making me all depressed and shizz.

Developmental disability	Mental or learning disability or mental or learning disability as a descriptor	•Absent = 0 •Present = 1	Have you ever been discriminated against at Wilson for your mental illness/learning disability/physical disability?
General mental health	No explicit condition, focus on mental health in general	•Absent = 0 •Present = 1	RT @BacaBaca2012: BACA fights again racism in mental health services at the grassroots and the new DSM.
Military or veteran-related mental health concerns	Any mental health topic related to military service members or veterans	•Absent = 0 •Present = 1	RT @Patrick_Baz: MoD confirms more British soldiers commit suicide than are killed in battle http://t.co/1ElLDKrg83
OCD	Obsessive compulsive disorder or use of as a descriptor	•Absent = 0 •Present = 1	My boss has OCD. Lord..
PTSD	Post-Traumatic Stress Disorder of use of as a descriptor	•Absent = 0 •Present = 1	@v_pendleton PTSD effects all the family, please support @SurfAction, supporting Veterans, vital work being done http://t.co/YVa50lCVcB
Recovery	Focus on a return to health after mental illness or addiction disorder	•Absent = 0 •Present = 1	We need #PeerSupport workers in all MH team who are the voice that speaks positively of hope and recovery,regardless of what has gone before
Substance use (alcohol)	Use and misuse of alcohol	•Absent = 0 •Present = 1	Survey of 663 published in J Addict Dis April 2012: Great Recession begun in 2010 linked w/problematic drinking from no job & bad job woes.

Substance use (illicit drugs)	Use and misuse of illicit drugs	•Absent = 0 •Present = 1	RT @wilson0804: RT @eonline: Cory Monteith tapped into his troubled past to play a drug addict in his final role--watch a clip here: http:/
Suicide	Discussion of suicide, clinical or colloquial	•Absent = 0 •Present = 1	This broke my heart RT @InlawsOutlaws: Bullied NM gay teen posts suicide note before taking his life http://t.co/v6lfwmHby9
Stigma (only applied for Tweets coded as mental health-relevant, coded by expert)			
Mental Illness Stigma	Derogatory use of mental health language or negative reference to mental health treatment	•Absent = 0 •Present = 1	Pointless closing the borders. It's like having a house party and you let the complete mental, twitching nutter in then say nobody else in.

Coders first coded the Tweets in the training and validation data set (henceforth referred to as the training data set for brevity)_for mental health relevance, and if a Tweet was identified as not relevant to mental health, no further coding was conducted for that Tweet. Tweets coded as somewhat or highly relevant were coded for all subsequent categories. Tweets were then coded for their type of content and the topic of the Tweet (see Table 3.5). The type of content and topic categories were not mutually exclusive, and a Tweet could be coded as a 1 for multiple categories. The type-of-content codes focused on capturing the nature of the content of the Tweet (e.g., whether it described self-relevant information or provided information of general interest). The topic codes were designed to capture whether or not Tweets discussed a wide variety of conditions and disorders. Several topics that were of interest (ADHD, domestic and child abuse, traumatic brain injury) were initially included in the coding scheme but did not appear in the training data set. In the early phases of coding, I included a 'write-in' category for topics to allow us to capture and iterate on the coding strategy with the team. As a result of this iteration, I added a general 'addiction' code to the coding scheme. The coders did not identify any other topics that occurred with a high enough frequency to be added to the list.

The stigma code was designed to capture whether a Tweet contained stigmatizing content about mental health. In early rounds of coding, I observed that the coding team had difficulty reaching sufficient levels of inter-rater reliability on this code. As a result, I opted to identify an expert (one of the principal investigators on the larger DoD campaigns project) and have her code 500 Tweets from the training data set that were identified as mental health-relevant.

Reliability of coders

Approximately 10 percent of the Tweets were coded by all three coders, allowing me to compute the mean of pairwise Cohen's Kappas to assess inter-rater reliability for each binary presence/absence score[143] (see Table 3.6). Cohen's Kappa can range from -1 to 1, with higher scores indicating better agreement among raters. I interpreted the Kappa values using labels of poor for Kappa <0.40, fair for 0.41 – 0.60, good for 0.61 – 0.80, and excellent for 0.81 – 1.00[144]. Mean pairwise Cohen's Kappa indicated fair to excellent agreement across most codes. There was poor agreement among coders for only three codes (mental health resources, recovery, and traumatic brain injury). These codes also correspond with categories for which few Tweets were identified by coders (as indicated by lower figures in the average number of observations column in Table 3.6). I did not eliminate codes from the coding scheme based on inter-rater reliability because I was interested to see the modeling performance for those codes. The consideration is that codes with low agreement across coders should not be able to produce accurate models. Though it is important to consider that appropriation, mental health resources, other-focused, developmental-disability, general mental health, OCD, and recovery are all codes that the coders had varying degrees of trouble identifying reliably.

Table 3.6 Inter-rater Reliability for 500 Tweets Coded by Three Coders

Code	Mean Cohen's Kappa	Average number times feature coded as "present"
Mental health relevance		
Relevance	0.77	185.00
Type of content		
Appropriation	0.55	61.33
Information	0.60	58.00
Mental health resources	0.23	5.67
Other-focused	0.51	33.00
Self-focused	0.72	79.00
Topic		
Abuse[b]	0.33	0.67
Addiction	0.81	28.33

[143] Inter-rater reliability coefficients were computed using the R package *irr,* Gamer et al. (2012)
[144] (Hallgren, 2012)

ADHD[b]	N/A	0.00
Anxiety	0.61	5.67
Autism	0.87	7.67
Bipolar disorder	0.94	12.00
Depression	0.92	43.33
Developmental disability	0.41	6.00
General mental health	0.48	48.67
Military or veteran-related mental health concerns	0.90	3.67
OCD	0.50	0.67
PTSD	1.00	1.00
Recovery	0.22	1.00
Substance use (Alcohol)	0.77	1.33
Substance use (Illicit drugs)	0.80	6.67
Suicide	0.85	14.00
Traumatic brain injury[b]	0.00	0.33

[a]Average number of observations refers to mean of the number of instances in which each of the three coders coded a Tweet as having the feature being coded.
[b]Code later removed from the coding scheme due to low volume of Tweets.
NA = not applicable. Could not be calculated because no Tweets were present.

Automated coding

The goal was to classify the 13.4 million Tweets in the working data set. This volume of Tweets would require over a decade of labor time to code manually. Instead, I develop an automated system of classification. I used the 4,760 human-coded Tweets in the training data set to build an automated coding model that can replicate the human hand-coding on a large scale. To create an automated coding model, I developed models for each characteristic of interest so that I could use the model to predict the presence or absence of each of these characteristics for any Tweet[145].

[145] For an example of such classification for topic relevant study, see Aphinyanaphongs et al. (2014) and Cole-Lewis et al. (2015). Cole-Lewis et al. utilize the same underlying mathematical model as this work.

In order to create a numerical representation of the text, I first transformed the coded training data set of Tweets into a spreadsheet known as a document matrix. In the matrix, every Tweet is represented as a single row, and every column represents a possible word. The number of columns is equal to the total number of different words in the entire sample of human-coded Tweets. The cells of the spreadsheet are populated with zeroes and ones representing the presence or absence of the word corresponding with a given column. To simplify the matrix, I first cleaned the documents in order to eliminate non-standard characters and reduce multiple versions of a word to a common spelling, for example turning 'go' and 'going' into a single word. There are also many words which appear very rarely in the data set: for example, a particular misspelling may occur only in a single Tweet. Such words are unlikely to provide any predictive power for identifying future tweets. Words that appear less frequently than once in 1000 Tweets were not included in the matrix[146]. The resulting matrix had 1,297 columns representing 1,297 distinct words present in the training data set and served as the input for the automated coding models.

The results of the modeling work produce independent prediction on the presence or absence of each characteristic of interest. The output of each automated coding model is binary and indicates the presence or absence of the Tweet characteristic of interest analyzed in that model (e.g., mental health relevance, PTSD, stigma). Each of the model predicts a single Tweet characteristic, so a distinct predictive model for each Tweet characteristic is constructed. It is of note that the coding scheme called for a primary evaluation of whether the Tweet was in fact relevant to mental health, and if it was not, no further categories were coded. This means that all 4,760 Tweets were used in the training data set to build a model predicting whether a Tweet is mental health-relevant or not and the 500 Tweets in the training data set coded for stigma by an expert were used to build a model predicting whether a Tweet contains stigmatizing content or not. Only the 1,618 mental health-relevant Tweets in the training data set were used to construct models for all other characteristics.

I compared two possible analysis strategies for use in the automated coding models - logistic regression and support vector machines (SVM). Logistic regression is a common statistical approach that provides information on how variables contribute to the outcomes. In this case, each word represented in the document matrix served as a variable that predicted whether a Tweet had the characteristic of interest being tested in that model. As the logistic model is valuable for its simplicity and interpretability only a basic regression was used. No interaction terms were tested nor was there any variable selection approaches implemented. SVM is a machine learning method designed to maximally separate observations into similar groups based on their quantifiable properties. In this case, this meant separating Tweets with a characteristic present from Tweets absent that characteristic, based on the presence or absence of specific words in the Tweet. As SVM gauges performance by the distance between classified groups, there is a penalty parameter to the distance metric based on misclassification. This penalty parameter allows for some tuning of the classifier. Different penalty parameters were tested and ultimately a penalty of C=50 was found to

[146] Excluding words that have a low frequency — referred to as 'sparse terms', see Feinerer (2015) — helps with generalizability and prevents overfitting, Hawkins (2004). There is no standard to assess the level of frequency which is too low to be meaningful. In this case, the level was selected to reduce the number of words in our matrix while also maintaining over 1000 words to be used in the predictive model.

provide the most consistently predictive results. Given the constrained timeline that was available for this research modeling exercise it was not possible to fit a different penalty parameter for outcome variable, so 50 was used for all the variables modeled. In testing, SVM produced more accurate prediction as determined by comparing several statistics that serve as indicators of model performance (area under the curve, true positives, and true negatives) that are described in detail in the next section. As a result, I elected to use SVM for the automated coding models, and I present the results and performance below[147].

Model performance and selection

In this section I review the quality of performance of our SVM-based automated coding models in predicting from the training data set whether a Tweet has a characteristic that is relevant to the research question. When considering model performance, I review the following statistics, presented in Table 3.7 for each automated coding model:

- **Area under the Curve (AUC)** - a summary statistic of the number of accurate predictions that a model makes. In other words, it is a summary metric of the number of times the automated coding model codes a Tweet in the same way that a human coder coded it. An AUC of 1 indicates perfect agreement between the value predicted by the model and the human-assigned value, and an AUC of 0.5 indicates that the model predictions are no better than chance[148],[149].
- **True positives** - the number of Tweets that were predicted to contain the code of interest by the automated coding models and were coded as such by human coders
- **False positives** - the number of Tweets that were predicted to not contain the code of interest by the automated coding models but were identified as having that code by the human coders
- **True negatives** - the number of Tweets that were predicted to not contain the code of interest by the automated coding models and were coded as not having that coded by the human coders.
- **False negatives** - the number of Tweets that were predicted to contain the code of interest by the automated coding models but were identified as not having the code of interest by human coders.
- **Total actual positives** – the number of Tweets that were identified by the human coders as having the code of interest. Total actual positives are always the sum of true positives and false negatives.
- **Total actual negatives** – the number of Tweets that were identified by the human coders as not having the code of interest. Total actual negatives are always the sum of true negatives and false positives.

[147] Interested readers may find more details and theory on automated coding model approaches in Bishop (2006)
[148] (Huang and Ling, 2005)
[149] For a discussion of the challenges of AUC and the need to report the types of errors and not simply overall error see Lobo, Jiménez-Valverde and Real (2008)

Though no formal procedure exists for determining the quality of model performance, I relied on several rules of thumb to examine the models listed in Table 3.7. I examined AUC to identify models that had lower values relative to other models. Based on personal experience and the range of values across the models, AUC values below 0.700 appears to represent models that are low performing relative to the rest of the models[150]. Also, I tried to examine models to determine where there might be a high number of false negatives relative to true negatives or a high number of false positives relative to true positives. Combining the AUC analysis with the false positive/false negative analysis allowed for the selection of effectively modeled characteristics.

[150] I note that the baseline model for topic classification for Twitters internal researchers has an AUC of .72, so I elect to use .7 as a baseline for which to judge our models. See Yang et al. (2014)

Table 3.7 SVM-Based Automated Coding Model Performance for Each Tweet Characteristic

Characteristic	Model retention ● = retained ● =not retained	AUC	True Positives	False Positives	True Negatives	False Negatives	Total Actual Positives	Total Actual Negatives
Mental health relevance								
Relevance	●	0.864	1026	331	2811	592	1618	3142
Type of content								
Appropriation	●	0.785	305	143	916	241	546	1059
Information	●	0.750	400	263	726	217	617	989
Mental health resources	●	0.869	20	0	1574	23	43	1574
Other-focused	●	0.668	72	106	1213	219	291	1319
Self-focused	●	0.734	393	272	721	226	619	993
Topic								
Abuse	●	1.000	1	0	1610	6	7	1610
Addiction	●	0.972	207	19	1368	20	227	1387
ADHD	●	0.940	0	0	1608	9	9	1608
Anxiety	●	0.913	23	11	1556	26	49	1567
Autism	●	0.927	51	1	1540	25	76	1541
Bipolar disorder	●	0.983	109	3	1499	6	115	1502
Depression	●	0.997	418	4	1190	5	423	1194
Developmental disability	●	0.972	36	2	1565	14	50	1567
General mental health	●	0.907	184	60	1256	110	294	1316
Military or veteran-related mental health concerns	●	0.947	5	1	1593	17	22	1594

OCD	🟢	1.000	20	1	1592	4	24	1593
PTSD	🟢	0.968	20	0	1586	11	31	1586
Recovery	🔴	0.589	0	0	1610	5	5	1610
Substance use disorder (alcohol)	🔴	0.953	3	1	1604	9	12	1605
Substance use disorder (illicit drugs)	🟢	0.930	14	4	1570	29	43	1574
Suicide	🟢	0.991	176	5	1430	6	182	1435
Traumatic brain injury	🔴	0.977	5	0	1608	4	9	1608
Stigma								
Mental illness stigma	🟢	0.804	42	9	384	68	110	393

The total actual positives and total actual negatives columns contain the counts of presence or absence of a code in the training data set that was used to train the automated coded models.

A review of Table 3.7 indicates that the automated coding models performed well for most categories, including mental health relevance, many of the topic codes, and stigma. For these categories, AUCs were sufficiently high, and examination of false positives compared to true positives and false negatives compared to true negatives did not yield an indication of poor performance. I elected to apply these models (i.e., the models with a green circle in the model retention column) to the larger Twitter data set.

Several models performed too poorly to consider using (those with red circles in the model retention column). I also examined models for the following characteristics for potential poor performance but ultimately decided to retain them: type of content codes for information and self-relevant and the topic code for military and veteran-related mental health concerns. Examining the model performance revealed that many of the poorly performing models were those for which that topic of interest occurred infrequently in the training data set. Specifically, many of the models that failed have only 5 to 22 occurrences. This makes it unlikely that an accurate model can be developed to predict such a rare event without significantly expanding the size of the training data.

The type-of-content codes were among those that performed somewhat poorly. The presence of appropriation, informational content, and self-focused content was difficult to predict, and these models had an AUC of 0.75. Because these categories were critical to understanding Tweet content and the AUCs for these categories were at threshold, I retained these models. Another type of content code capturing whether Tweet content is other-focused was not retained because of the AUC below 0.70 and because it yielded more false positives than true positives.

Modeling type of content is likely more challenging than modeling, for example, topic, due to the nature of the category. While there are likely keywords for identifying mental health topics that clearly indicate whether a feature is present or absent (e.g., most Tweets about depression will have some version of the word 'depression'), there may not be equivalent keywords for considering type of content. For example, if trying to identify whether a Tweet is self- or other-focused, words like "I" and "you" are too common to be unique keywords. Thus the models for predicting type of content rely on patterns of co-occurrence among words to make predictions.

Applying each model I opted to retain yielded a prediction of the number of Tweets in the working data set of 13.4 million Tweets that would have that characteristic (see Table 3.8).

Table 3.8 Volumes of Tweet Predictions in Working Data Set

Characteristic	Predicted Count in Working Data Set
Mental health relevance	
Relevant	2,277,092
Type of content	
Appropriation	681,514
Information	697,206
Mental health resources	16,837
Self/Personal	924,806
Topic	
Addiction	151,059
Anxiety	3,404
Autism	41,672
Bipolar disorder	141,817
Depression	249,397
Development disability	13,579
General mental health	321,477
Military or veteran-related mental health concerns	473
OCD	26,128
PTSD	9,468
Substance use (illicit drugs)	11,687
Suicide	111,318
Stigma	
Mental illness stigma	220,991

Working data set, N = 13,432,321

Modeling approach comparison

Neural Nets (NN) are a popular machine learning approach that I was not able to test during the initial classification work because of time constrains. In this section I explore the potential use of this method for the problem of classifying Tweets according to type and topic (see table 3.5 for definitions and examples). A NN is a machine learning approach that uses the logic of human neurons to construct a predictive model and has been gaining attention for its ability to solve very complex problems. Recently, NNs were used by Google subsidiary Deep Mind to build Alpha Go, the first computer engine that was able to best a top human Go player[151]. This was a feat that until recently was not considered feasible due to the incredible large number of possibilities in Go.

[151] (Moyer, 2016)

NNs may also be an effective method for the complexity of a task such as language interpretation and text classification. A neural net is a system of networked neurons. A neuron in this case is an algorithm that takes the sum total of inputs and based on threshold rules produced a single output. The function that determines this threshold is predetermined in designing the architecture. The inputs to the neuron are weighted, and it is these weights which are iteratively optimized in the machine learning approach to make sure that the neuron outputs the correct value for a given case of training data. The strength of a neural network is that by using multiple neurons any relationship can be effectively approximated. This feature of interconnected neurons is what makes NNs ideal for complex problems.

The downside of neural networks is their complexity and relative immaturity of the field. Because of the structure of NNs it is exceedingly difficult to understand how inputs drive the outputs. There is little ability to extrapolate which inputs have the largest influence and which have none. There is no empirical rule for what network size a given problem requires. The amount of interactions between the various nodes means that best numerical methods for minimizing the error terms are unclear and vary from problem to problem. Additionally, as a result of the novelty of this method there are few mature tools to implement and interpret them. Neural nets were first implemented computationally in 1954[152], but because of the amount of computation that is required there was little work using the approach until 2009 when the method produced high performance for pattern recognition[153]. The relative recent uptake of the methods means that there is little work comparing various structural features of networks, the optimization procedures, or tools for parsing out the internal workings of the algorithm. Additionally, as I discovered in this work, the implementation of neural nets that do exist can vary significantly.

The approach used for this comparison was initially to compare the performance of various neural net architectures to versions of the SVM classifier for all the outcome variables studied in this dissertation. This was done by examining SVM with a radial basic function kernel and several penalty parameters and a single layers neural network with 1 to 7 nodes. Both models were tested with a 10-fold cross validation approach. However, the run time for the total process was several weeks of computing time[154]. After completing the run an error was discovered in the model formulation and the results had to be discarded. I elected not to repeat the process because of the large run time and the limited value of the comparison of all outcome variables. That exercise demonstrated the challenge of working with neural nets: the huge amounts of computation required means that neural nets are poorly suited to iterative modeling and experimentation.

Instead of focusing on how NNs compare with SVM across multiple outcome variables I focused on understanding how to make a neural net that would be best suited for my task of classifying Tweets. To do this I utilized two R libraries, *nnet*[155] and *neuralnet*[156]. I found that the two libraries under default parameters (set to identify categorical not continuous variables) produce

[152] (Farley and Clark, 1954)
[153] (Graves and Schmidhuber, 2009)
[154] The machine used was a server with 16 CPU cores, and 128gb of RAM.
[155] (Ripley and Venables, 2011)
[156] (Fritsch, Guenther and Guenther, 2012)

consistently different results. *Nnet* performed approximately 5 points better on the AUC metric across all architectures. This result was consistent for multiple runs and various levels of cross validation. The consistency of the difference indicates that the discrepancy was not a result of variance of fit but was an actual performance difference.

In investigating the difference between the two libraries I found a limitation of current tools available for neural nets and challenge of implementing this approach. The *nnet* library, which classified data more accurately, provided very little documentation and very few parameters that could be tuned, while the *neuralnet* library allowed for much finer control. *Neuralnet* provides the ability to set the algorithm used to compute the optimal weights within the neural networks as well as tune the parameters within the algorithm. It is difficult to judge how the type of algorithm used for optimization will affect the model performance so simply testing various features is required. After testing the algorithms offered I was able to find a set of options that allows *neuralnet* to have a performance result comparable to *nnet*. I used traditional back-propagation with a learning rate of 0.01[157]. This produced model performance within 2 points of *nnet.* Based on the documentation included in *nnet*, I believe that the library is based on a back-propagation approach though I cannot be certain.

However the speed of calculation was still drastically different between the two approaches. *Nnet* was able to obtain better results in approximately 1/20 of the time of *neuralnet*. Examining the log files from the calculations it appears that *neuralnet* has much stricter rules for when it defines a function to have converged at the optimum, which means that the *neuralnet* implementation continues to search for an optimum for many more calculation iterations than *nnet. Nnet* typically settles on an optimum in 100 iteration steps while *neuralnet* will often calculate to 100,000 iterations. It is not clear from the documentation what each library defines as iteration (a cycle of optimization of parameters). When I have attempted to restrict the number of iterations used by *neuralnet* or alter its threshold for determining convergence to the optimum the model could not be successfully calculated.

Ultimately, I did a model comparison using the default parameters of *nnet,* examining the various sizes of neural network architecture and compared them to the SVM results. The testing was done using 5-fold validation to minimize variance in performance while maintaining practical computation time. The testing was done for a six outcome variable, see tables 3.9 and 3.10. Additionally, the largest, 100 node network, was run for only one variable: relevance. The reason is that even with *nnet*'s comparatively fast computation speed large networks still take a significant amount of time and the largest networks that I tested had a runtime of over a day. Given time constraints I elected to use the largest NN to analyze the variable with the greatest number of observations with the greatest amount of variation and a handful of other variables with fewer cases. The other five variables – appropriation, other person focus, depression, general mental

[157] Back-prorogation is the standard approach for constructing supervised (known outcome) NNs. Back-propagation is the method where the error of the final estimate is fed back into the network so that each node gets assigned a value for its relative contribution to the error. Learning rate refers to increment at which the relative importance of each neuron is adjusted, with smaller learning rates resulting in a faster network construction but less accurate results. For additional details, see Hecht-Nielsen (1989).

health, and stigma- were selected to include a range of performance values with SVM. The best performing SVM model was depression and the worse was other person focus. All neural networks considered were single layer as it is theoretically sufficient to approximate any function of inputs to outputs[158]. Additionally, *nnet* does not allow for multilayered architecture. Below, Tables 3.9 and 3.10 show the performance of the models, table 3.11 compares the two approached.

Table 3.9 Neural Net AUC Performance

	Number of Nodes							
Characteristic	**1**	**2**	**3**	**4**	**5**	**10**	**40**	**100**
Relevance	0.86536	0.85380	0.87434	0.88959	0.89740	0.89947	0.90362	0.90240
Appropriation	0.81296	0.81851	0.82822	0.83578	0.83676	0.84003	0.83639	
Other Person	0.66096	0.67311	0.67631	0.68267	0.68341	0.68031	0.68204	
Depression	0.99783	0.99780	0.99779	0.99781	0.99781	0.99776	0.99777	
General MH	0.95162	0.95007	0.95024	0.94955	0.94936	0.94891	0.94839	
Anxiety	0.95042	0.95049	0.95141	0.95147	0.95167	0.95251	0.95282	

Note: Model is a single layer NN, performance is evaluated with a 5 fold cross validation approach. Highest AUC values for each variable are highlighted.

Table 3.10 Support Vector Machine AUC Performance

	Cost Parameter							
Characteristic	**0.001**	**0.01**	**0.1**	**1**	**10**	**50**	**100**	**1000**
Relevance	0.88847	0.88937	0.88988	0.89641	0.88696	0.86851	0.86241	0.83940
Appropriation	0.79581	0.81792	0.81926	0.81765	0.80031	0.78263	0.77743	0.76227
Other Person	0.65093	0.67339	0.67764	0.67663	0.68221	0.67819	0.67380	0.67094
Depression	0.97683	0.97119	0.97030	0.97348	0.98306	0.98663	0.98723	0.98412
General MH	0.83740	0.84869	0.85150	0.84918	0.85974	0.86639	0.85870	0.84820
Anxiety	0.80054	0.91549	0.91616	0.91588	0.90934	0.89369	0.89483	0.89391

Note: Model is an SVM with Radial Basis Function Kernel; performance is tested with a 5 fold cross validation approach. Highest AUC values for each variable are highlighted.

Table 3.11 Comparison of Support Vector Machine and Neural Network Optimum Models

Characteristic	**NN Optimal**	**SVM Optimal**	**NN AUC – SVM AUC**	**Best Performing**
Relevance	0.90362	0.89641	0.01	NN, 40 nodes
Appropriation	0.84003	0.81926	0.02	NN, 10 nodes
Other Person	0.68341	0.68211	0.00	NN, 5 nodes
Depression	0.99783	0.98723	0.01	NN, 1 node
General MH	0.95162	0.86639	0.09	NN, 1 node
Anxiety	0.95282	0.91616	0.04	NN, 40 nodes

[158] (Hornik, 1991)

In reviewing the comparison of results, Table 3.11, NN outperformed SVM for all categories tested. However the classification performance improvement varied greatly across the different categories: there was almost no improvement in the classification of 'other person' focused Tweets, and a large improvement in classification performance of 'general mental health'. The variability in performance makes it difficult to assess the value of the significant computational time and methodological difficulty required for the construction of a NN classifier. Additionally there is little pattern in the NN architecture that performed the best. There is no rule for what the optimum number of nodes in a network might be. A review of programming forums found that the recommended number lies somewhere between 1 and the number of predictors in the data set. It may be that at much higher volumes of nodes there would be a significant performance improvement but it is computationally prohibitive to test in this use case. Additionally, 100 nodes did not perform 40 nodes when I tested it for the relevance variable; though, this only is a single test case, see Table 3.9. Similarly, there is no consistent cost parameter that optimized SVM for all problems, Table 3.10.

The best statistical approach for modeling text data appears to be dependent on the outcome variable of interest and the time and computational resources available. There is no single architecture that performs best for every category of text classification for mental health Tweets. Finding the highest performing model requires iteratively tuning the parameters of a given approach for the specific outcome. Additionally the best method for a given task is not necessarily the best performing one. This work illustrates the power of neural network but also the very high complexity of implementation and the very large computational overhead. For the task of classifying a large volume of text data to understand overall sentiment and topics of interest SVM appears to be preferable to NN.

Network analysis

To better understand how mental health engaged Twitter users communicate with each other and the campaigns, I conducted a network analysis. I focused on directed Tweets in which one user addresses another (as opposed to undirected Tweets which are posts meant to be read by anyone on Twitter and reTweets that involve reposting another other users' Tweets). I used the directed Tweets to identify instances of users communicating with others, and used these connections as inputs into the *igraph* package for R to construct a social network of how users are linked by communication about mental health[159]. This analysis focuses on the communication ties between users who are engaged with mental health. This means the conclusions are focused on understanding the structure of the mental health conversation, not the ways in which the mental health conversation fits into the overall network structure of Twitter.

Social networks often give rise to groups of people, known as "communities," which can be defined as nodes connected more densely inside the group than to individuals outside of the group. Constructing a social network based on directed Tweets allows me to understand if communities of users communicate about mental health in ways that are distinct from overall patterns observed in

[159] (Csardi and Nepusz, 2006)

the 2.3 million mental health-relevant Tweets in the working data set. Doing so also allows me to determine if these communities are distinct from each other in meaningful ways. I can also understand whom campaign-related Twitter content reaches, how far campaign messages are spreading, and if there are groups of people that converse about topics that are relevant to the campaigns[160].

Understanding how the conversation about mental health may vary for an individual based on with whom they communicate and understanding how the campaigns are spreading online required two different methods of network construction. To understand the variation in the overall conversation about mental health on Twitter, I had to reduce the number of observations that did not have meaningful mental health information. The general mental health conversation includes data from spam accounts (which post in ways intended to avoid detection by Twitter's automated detection methods) and users which are influential outside of the mental health context (e.g., celebrities) but are not central to the mental health conversation being studied. As a result, I focused on mental health-relevant Tweets from our working data set that were directed Tweets. Undirected Tweets were not included as they do not mention another user, making such tweets irrelevant for the goal of our social network analyses – to identify connections among users. ReTweets were omitted because often the audience for the reTweet is unclear, and reTweets are often associated with spam accounts. After omitting undirected Tweets and reTweets, a large volume of Tweets was available to construct a network (534,000 connections among 777,000 users).

Understanding the conversation network around the DoD campaigns required a different social network construction than for the overall conversation. I was unable to detect communities focused on military or veteran mental health issues associated with the campaigns inside the network constructed of direct messages (as discussed above). However, including reTweets, in addition to directed messages did allow for the detection of such communities. The conversation related to the campaigns was a much smaller and much more clearly defined than overall conversation about mental health. This meant that the concerns that led to the exclusion of reTweets for the general conversation – spam accounts, unrelated celebrity Tweets, computational restrictions – were not relevant. Instead maximizing the information available was more important. To understand the network of campaign-engaged users I created a network from the total data set of 2.3 million mental health-relevant Tweets (directed messages and reTweets), located the campaign-engaged users within that network, and identified which communities they belonged to. The data set of campaign related-Tweets contained a small number of users (n = 10,134) who engaged with the campaigns. Of those 10,134 users, I identified 9,123 users present within the data set of 2.3 million mental health-relevant Tweets.

Having constructed these two networks of mental health communication I focused on detecting communities within the networks. Community detection involves identifying clusters of users that

[160] For an early discussion of community structure on Twitter and the insights that community analysis offers see Java, Akshay, et al. "Why we twitter: understanding microblogging usage and communities." *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*. ACM, 2007.

are densely interconnected with each other but separate from users outside of the group. To identify clusters, I used the walktrap method as it has been shown effective for large networks such as the network we were analyzing[161]. I identified communities made up of users that sent or received directed Tweets about mental health to each other. Community detection is a computationally intensive task, and thus it was not feasible to perform community detection on the network constructed with directed Tweets and reTweets. To address this, I performed community detection on the subset of all nodes with two or more network connections. I eliminated all users with only one connection – the users that were most peripheral to the networks and users that were part of very small clusters. The community detection algorithm identifies communities within larger network components and also labels isolated components as communities. By eliminating the nodes with only a single connection, clusters of two users were not identified as communities but other small clusters of users remained and were identified as communities. This step reduced the size of the network from 1.6 million users with 1.4 million connections to 352,000 users and 418,000 connections, which made the task computationally feasible. The network that I constructed without the use of reTweets (534,000 users, 777,000 connections) was computationally tractable without a size reduction and therefore community detection was performed on the full 534,000 user network.

I used the community structure of networks to understand the characteristics of the mental health-relevant content shared within different communities. I examined how mental health discourse varies among communities and if there are outlier communities with particular high volumes of a particular type of content, content focused on specific mental health topics, or stigmatizing content. I examined communities with particularly high levels of stigma in their posts, as well as high volumes of posts related to the topics of PTSD, depression, substance use disorders, and suicide. I also explored the distribution of message by type and topic for communities that are engaged with the campaigns in order to understand the kind of users that the campaigns are reaching.

I also examined some measures of centrality to understand the importance of users within the social network of mental health discourse. I opted to explore eigenvector centrality because it serves as a measure of importance that considers the number of connections that a user has as well as the number of connections that those connections have. Users with higher eigenvector centrality scores are connected to many well-connected users, and thus are likely to serve as a central hub for effective distribution of content. This approach aligns well with the notion that effective dissemination of mental health public awareness campaign content on Twitter depends upon the ability of a user to propagate messages and influence the conversation around mental health. This analysis of centrality does not address the centrality of user within the broader Twitter network, focusing instead on the active members of the mental health conversation.

[161] (Pons and Latapy, 2005)

Chapter 4: Characterizing the Mental Health Conversation

Introduction

This chapter is the first chapter discussing the results of this analysis. It focuses on answering the first policy question and its constituent parts. The core question is how to describe what is being said on Twitter about mental health.

P1. Which mental health topics are characterized by positive and supportive discourse and which are discussed in a stigmatizing way?

P1R1. What is the type, topic, and tone of mental health content on Twitter?

P1R2. Does the type and tone of mental health focused content vary across the network communities of users posting mental health focused tweets?

In this chapter I discuss the types of messages that people post on Twitter and find that self-focused are the most prevalent but other types are also present in large volumes. In analyzing the stigma content of the mental health relevant Tweets I find that approximately 10% contain stigmatizing content. General mental health is the most common topic of discussion, while depression, suicide, anxiety and addiction are also common. In analyzing the social network structure of Twitter communication I find that some users do engage with mental health in groups and that these groups are heterogeneous, focusing on distinct topics and have distinct types of conversation.

Most mental health content on Twitter is self-focused and few Tweets share mental health resources

Panel A in Figure 4.1 depicts the number of Tweets that feature different types of content. The most common type of message are self-focused. Tweets containing informational content account for about one-third of total Tweet volume. Tweets that feature appropriation (that is, using mental health terms to describe things and not people [e.g., 'the weather is bipolar'] and using mental health language to describe emotional states [e.g., 'I am going insane watching this football game']) account for another third. Few Tweets involved the provision of mental health resources. These findings suggest that Twitter users are often discussing mental health in a self-relevant way. However, the prevalence of Tweets that are informational suggests that they are using Twitter for other purposes as well.

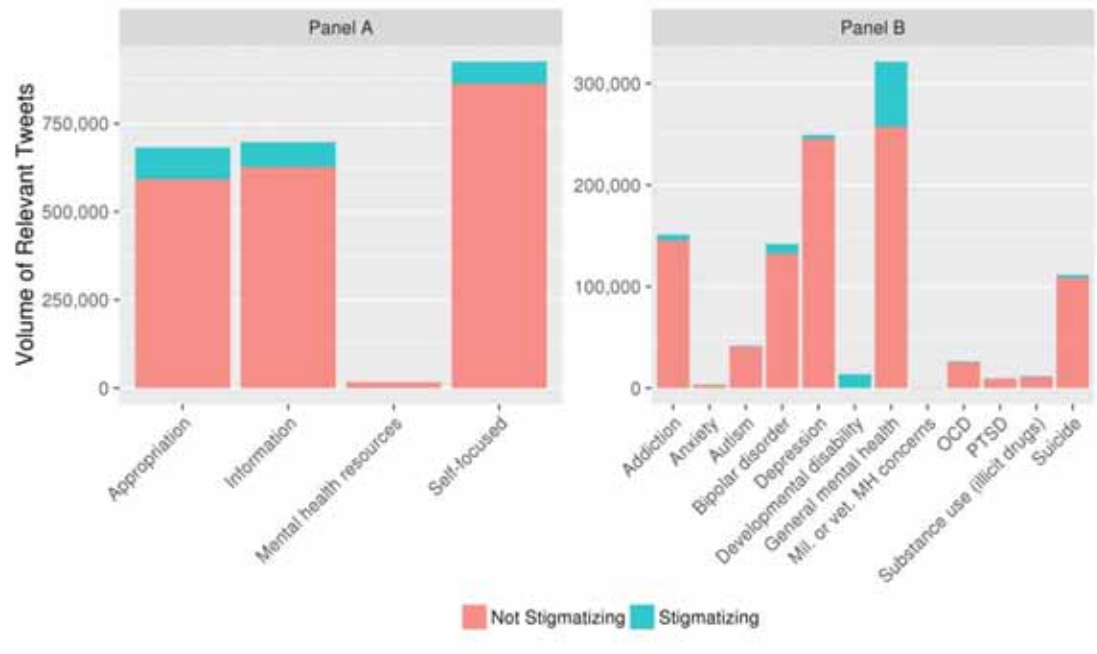
Panel B in Figure 4.1 depicts the number of Tweets that focus on each of the different topics included in the coding scheme. General discussion of mental health (without focus on a specific disorder) is the topic most commonly discussed. Other frequently discussed topics include addiction, bipolar disorder, depression, and suicide.

Approximately 10 percent of mental health-relevant Tweets are stigmatizing

Most Tweets were not coded as being stigmatizing. Ten percent of the 2.2 million mental health relevant Tweets were stigmatizing. Figure 4.1, Panel A shows that the proportion of content that was coded as stigmatizing is similar across Tweets coded as containing appropriation (13 percent), information (10 percent), or that were self-focused (7 percent). Figure 4.1, Panel B shows that most Tweets about most topics were not stigmatizing, with the exception of Tweets coded as discussing developmental disabilities, of which, 99.6 percent were stigmatizing. A larger proportion of Tweets focused on general mental health (20 percent) were stigmatizing (when comparing to the proportion of stigmatizing Tweets present for most other topics).

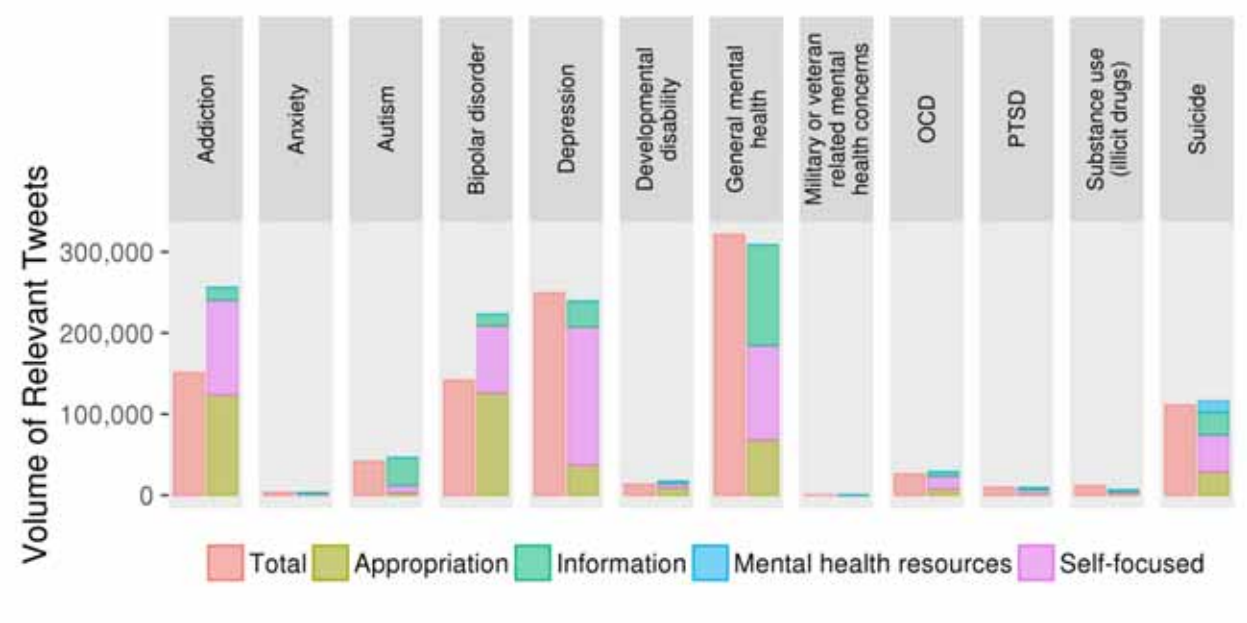
Tweet types and topics were related to each other in several different ways, see Figure 4.2. More Tweets about addiction and bipolar disorder contained content coded as appropriation relative to Tweets about other topics. An informal review of Tweets about addiction and bipolar disorder suggest that these findings may be driven by casual and colloquial use of terms like "addicted" and "bipolar" that are not in line with correct clinical use of the terms. I also note that a large volume of Tweets about depression are also self-focused, which is likely reflective of the positive correlation between self-focus and negative affect documented in psychological literature (Mor and Winquist, 2002). Finally, I note that Tweets focused on general mental health are most often coded as informational, which may be reflective of a general conversation about mental health.

Figure 4.1 Distribution of Tweets by Type of Content and Topic



Note. Panel A shows the distribution of type of content. Panel B shows the distribution of Tweets by topic. The distributions are independent each other, and codes are not mutually exclusive. The bars are stacked; the blue bar represents the part of each feature that is stigmatizing. See table 3.5 for definitions of type and topic categories.

Figure 4.2 Tweets by Type and Topic



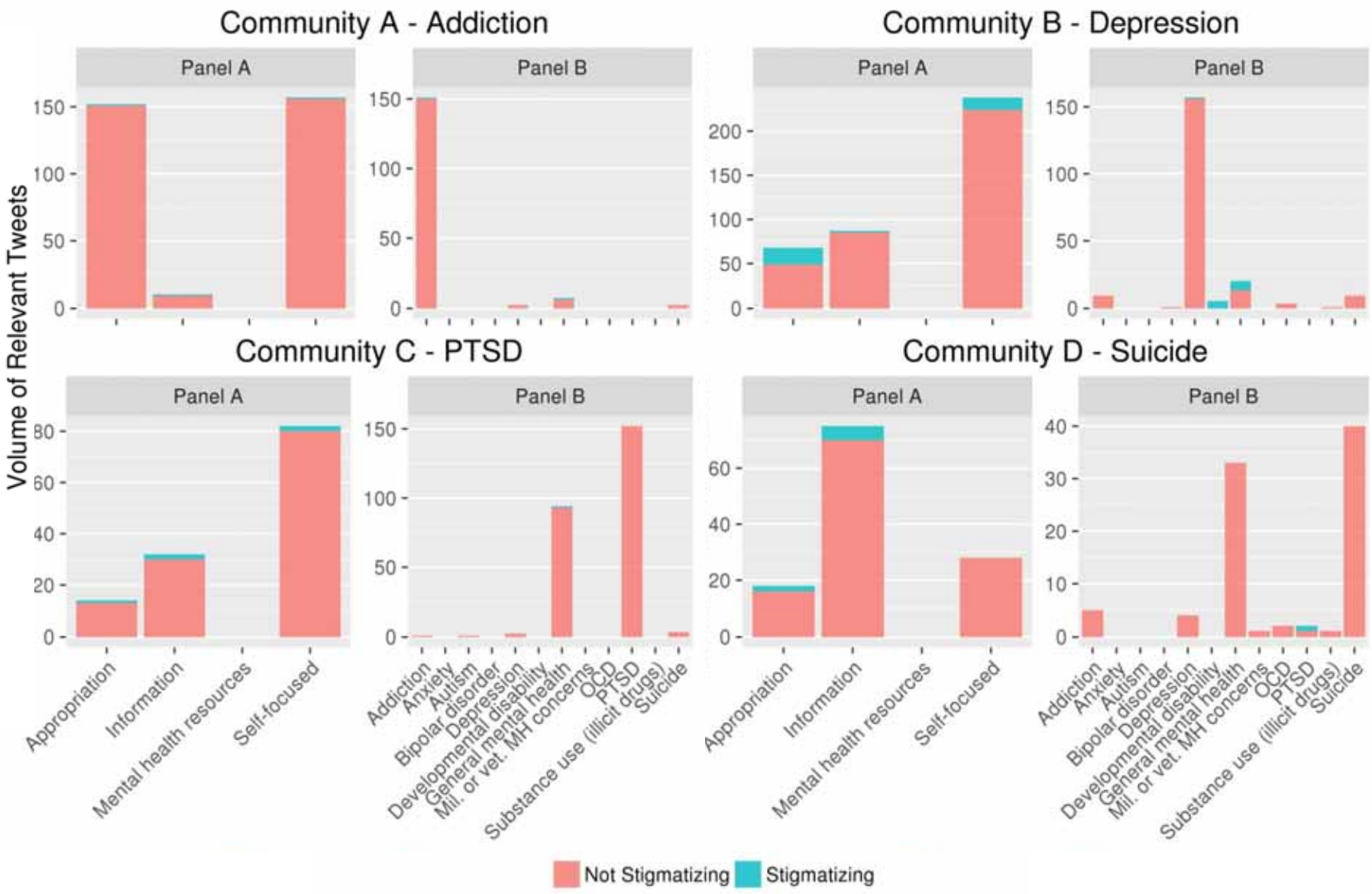
Note. Tweet types are not mutually exclusive, so the stacked bars can be higher than the total volume of Tweets in that topic.

Network community conversations about mental health vary in types and topics of Tweets

The social network analysis of mental health-relevant Tweets revealed over 126,000 communities, many of which are very small (e.g., pairs of users). Thus, I focus the analyses on the 96 communities that had at least 25 members and 150 mental health-relevant Tweets. I focused much of the interpretation on these larger communities because I wanted to ensure that there were adequate numbers of members and mental health-related activity to draw conclusions about community conversations. To provide illustrative examples of variation in mental health discourse among different communities, I identified four communities with the highest proportion of Tweets related to four mental health and substance use-related topics (i.e., addiction, depression, PTSD, and suicide) that are important concerns for service member and veteran mental health[162]. The distribution of content for those communities is in Figure 4.3. These graphs represent the distribution of all content posted by users belonging to communities with the high proportion of the characteristic of interest. Of note, the patterns of Tweet types and topics for all four communities differ from the overall patterns observed in Figure 4.1.

[162] (Bray et al., 2010)

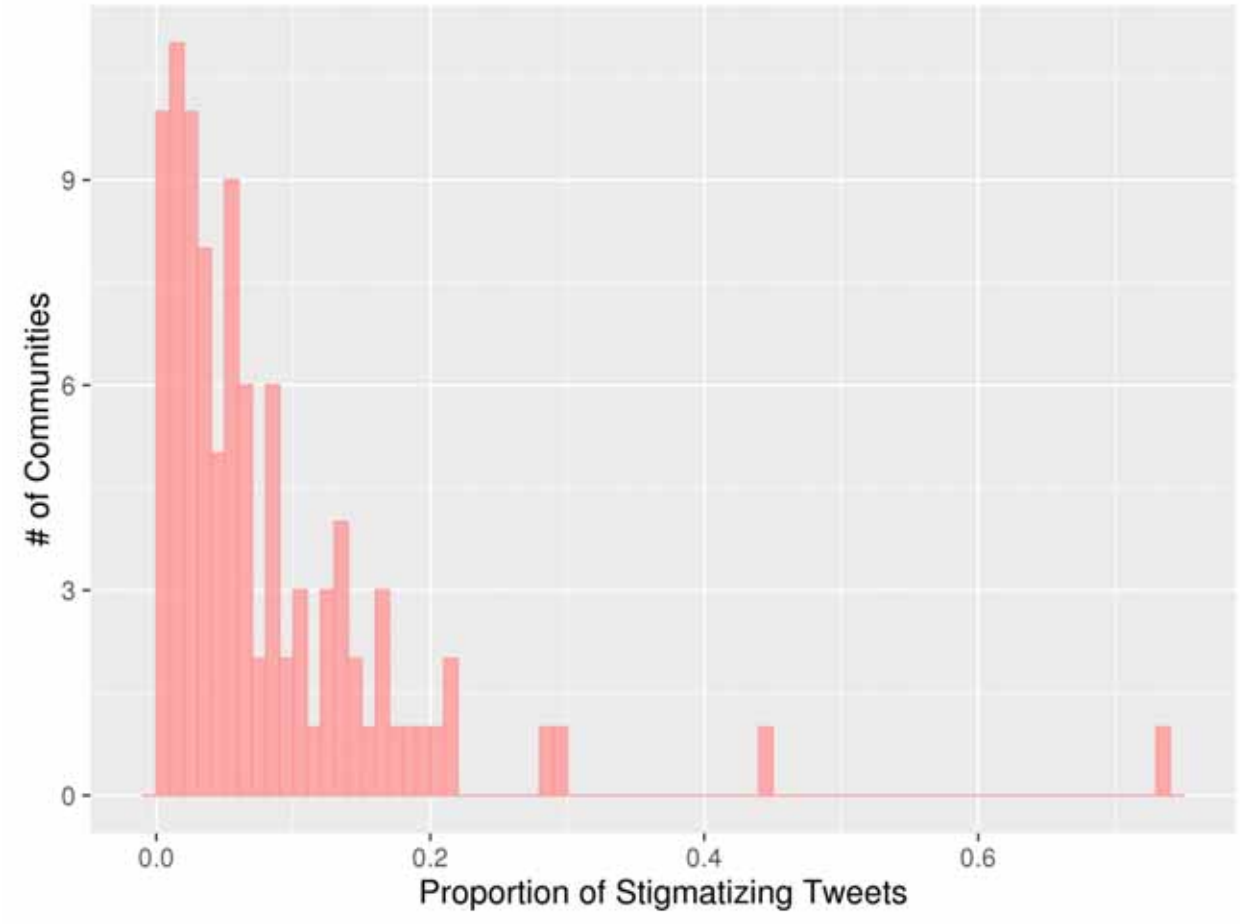
Figure 4.3 Characteristics of Tweet Types and Topic for Four Communities with High Proportions of Tweets About Addiction, Depression, PTSD, and Suicide



Among large communities with mental health conversations, 71 percent demonstrated low levels of stigma in community conversations

I also looked across the 96 larger communities to understand if volume of stigmatizing content varies across communities. Figure 4.4 shows the count of communities with different proportions of Tweets coding as being stigmatizing. For 69 of the 96 communities, about 10 percent of Tweets were stigmatizing. I manually examined the content of the two outlier communities for which 45 and 75 percent of their Tweets were stigmatizing. These two communities appear to consist largely of spam accounts that do not represent conversation among typical Twitter users[163]. I identified one additional community with high rates of stigmatizing content (30 percent of total content). Examination of this community indicated that it was largely populated by fans of pop music group One Direction. Tweet content was mostly focused on users using mental health terms when expressing strong emotions (e.g., "@Real_Liam_Payne you don't know, but you save me everyday and I don't make a suicide, thank you so much for all, Liam I love you x1178").

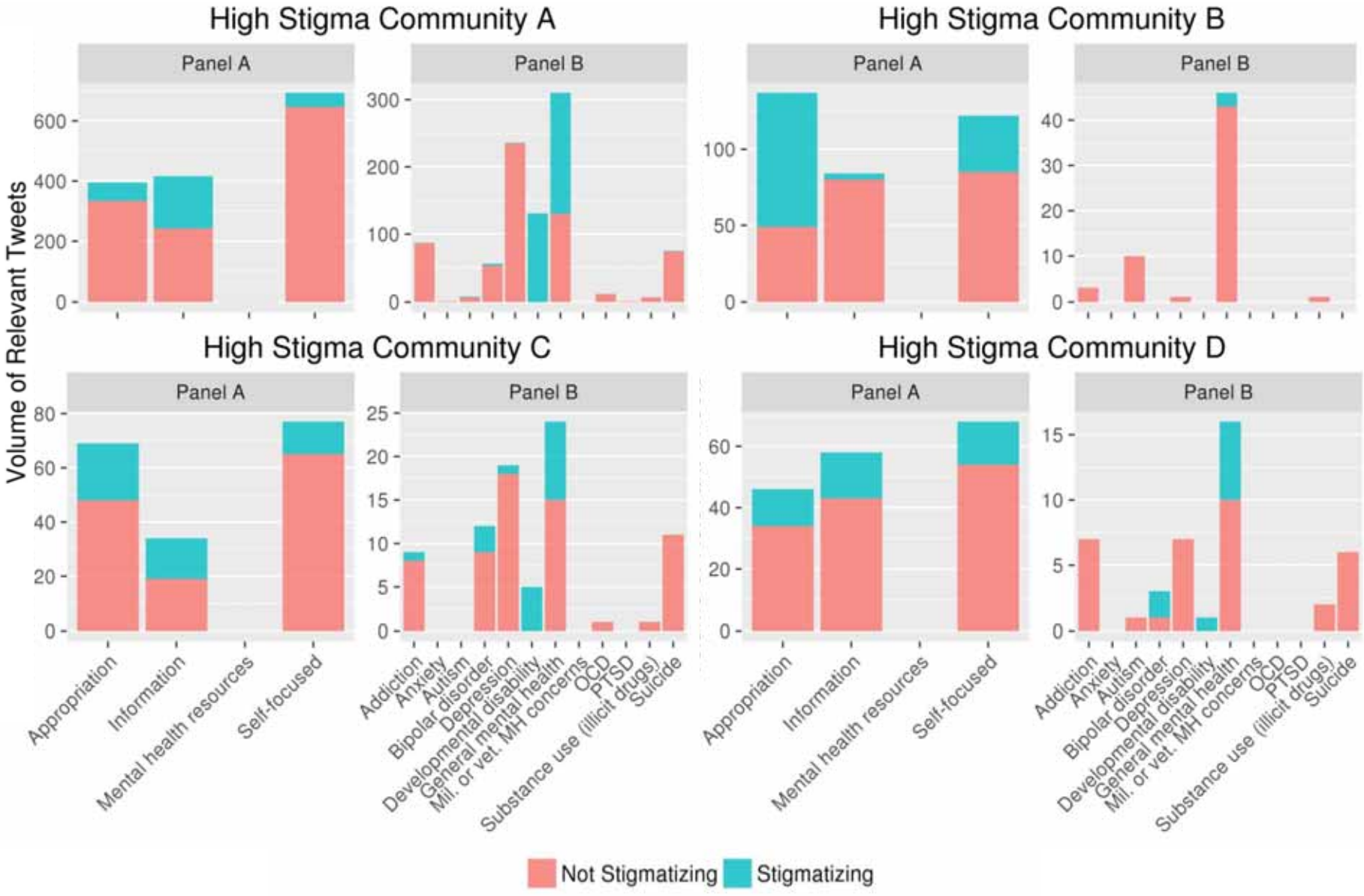
Figure 4.4 Variation in Proportion of Stigmatizing Tweets Across Communities



[163] A manual review of the content posted in these two communities showed many Twitter users that engage in "follower trading" in which individuals operate with the goal of creating accounts with a high number of followers which can then be sold or rented for profit.

I also examined the type and topics of Tweets among the four communities with the highest proportions of stigmatizing Tweets (excluding the previously mentioned communities that consisted of scam accounts) (see Figure 4.5). I chose to focus on communities where stigmatizing Tweets made up over 20 percent of the communities' Tweets as this allowed me to manually examine a manageable number of communities. Patterns of Tweet type and topics seem to largely resemble overall patterns in Figure 4.1. Communities A, C and D all have self-focused content as the largest single category of type of content, with appropriation and information individually having slightly less volume. The topic distribution for these communities was also similar to the overall topic distribution – general mental health was the single largest, with addiction, bipolar disorder, depression and suicide being the other major categories. There were some community specific differences from the overall trend – community A has larger proportion of developmental disability Tweets, community D has proportionally less depression focused Tweets – but none of these appeared to be consistent across multiple stigmatizing communities. Community B followed a different pattern of conversation. It is categorized by having appropriation as the major type of Tweet, and general mental health as almost the only topic of message posted. Given its limited amount of topic specific conversation and high amount of appropriation content it may be possible that the community conversation being detected is in fact the stigmatizing language used in part of a larger group conversation. This uncertainty is a limitation of the network construction methods I used: I did not capture the complete volume of posts by Twitter accounts, only having access to mental health specific content. It may be possible that the users that I identify as being in Community B are part of a larger unrelated group, and are a subset that happened to have used some language appropriated from mental health. Overall, the discourse among three of the four most highly stigmatizing communities did not differ significantly from the patterns seen in the overall data set. This suggests that many stigmatizing communities are not topic-focused but rather that stigma is simply part of their discussion of mental health topics.

Figure 4.5 Characteristics of Tweet Types and Topic for Four Highly Stigmatizing Communities



Chapter 5: Changes over time

Introduction

This chapter builds on the results described in the prior chapter and examines how the various characteristics of the online conversation about mental health change over time[164]. Additionally, this part of the analysis includes a review of the performance of the DoD sponsored campaigns to change attitudes towards mental health and mental health treatment. Initially the goal of studying the campaigns was to examine how the campaigns alter the online conversation, but as the campaigns proved to be too small to have an impact on the general conversation they are instead presented here as a case of study of the application of the social media analytic methods developed in this dissertation. The questions that this chapter seeks to answer are as follows:

P2. What are the longitudinal trends in content and sentiment of Twitter conversation around mental health and mental health treatment?

P2R1. Is there a change in the sentiment of the conversation? The topics? The type of content?

P2a. Do sentiment and topic changes correlate with activities of DoD public messaging social media campaigns?

P2aR1. Have there been changes in the online conversation that correlate with the DoD public messaging campaign efforts?

P2aR2. Does engagement with DoD social media campaigns vary across the communities of engaged Twitter users?

Overall, the conversation about mental health is improving in terms of stigma and awareness. The levels of stigma in the conversation are declining over time. This decline is consistent across all topics. Interest in mental health as a general topic of conversation is increasing. The public messaging campaigns studied are not large enough to impact this trend. However, the campaigns are generating engagement, and Real Warriors, which has been most active on Twitter, is generating greater engagement over time. I also see evidence that there is a community of people that are interested in the messages delivered by the campaigns.

Longitudinal trends in mental health discourse on Twitter

Tweet types and topics have changed over time with increases in informational content and discussion of general mental health and decreases in stigmatizing content. I conducted a longitudinal analysis of mental-health related Tweets from 2009 through 2016 (see Figure 5.1). To account for changing Tweet volumes over time, Tweet volumes were normalized by dividing by the total volume of mental health-relevant Tweets. I opted to analyze the proportion of mental-health relevant Tweets that were coded as each different Tweet type and topic.

As shown in Figure 5.1, Panel A, the proportions of mental-health relevant Tweets that are self-focused or that use appropriation decline over time. Informational Tweets slightly decreased

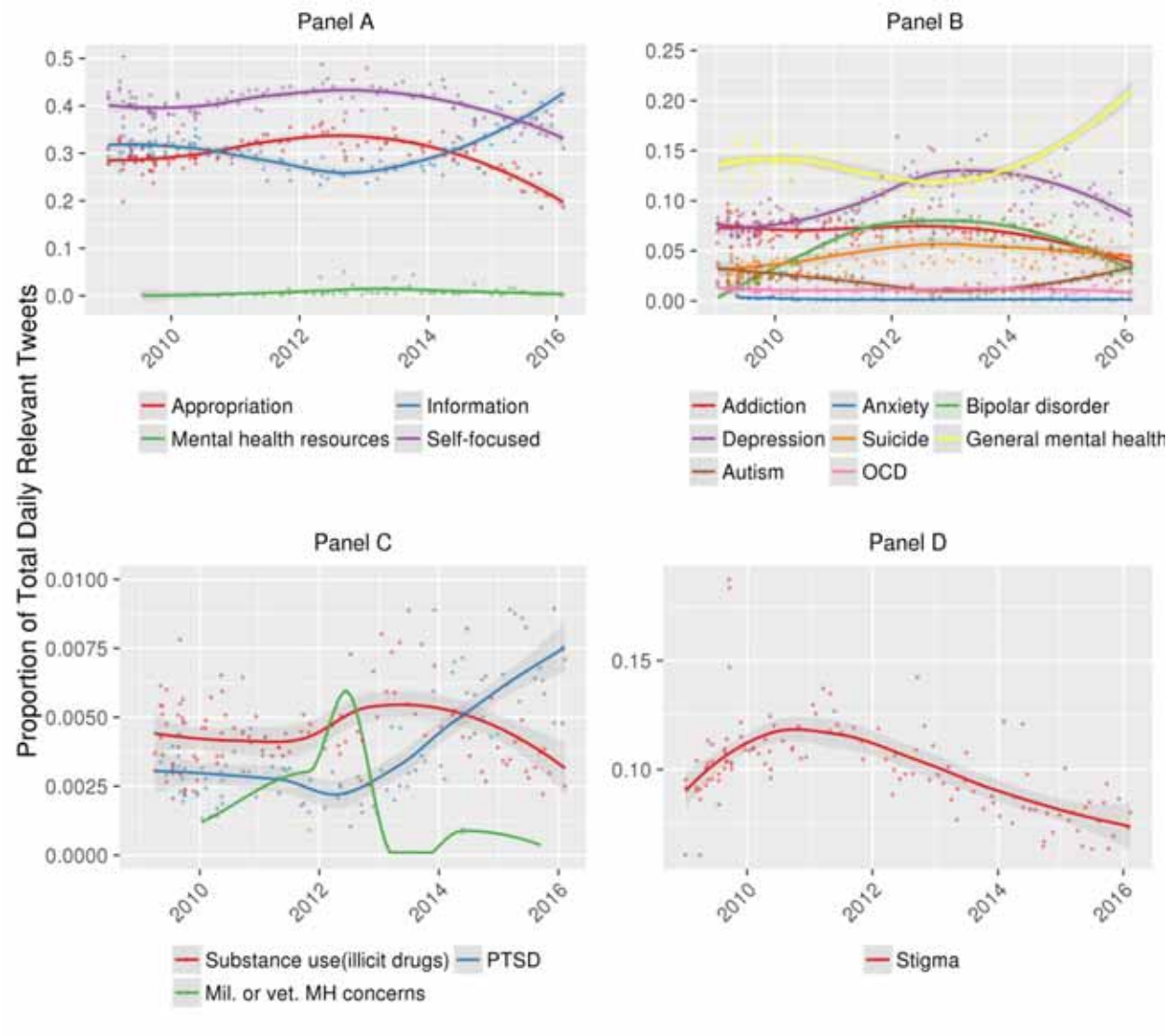
[164] All trend analysis is performed using data normalized for daily volumes of Tweets collected.

until the end of 2012 and have increased since. Despite increases in informational Tweets, the proportion of Tweets that contain mental health resources has remained steady over time.

I also see changes in Tweet topics over time (see Figure 5.1, Panels B and C). One of the most prominent trends is an increase in the volume of Tweets focusing on general mental health, which may suggest increasing engagement among Twitter users with the issue of mental health as a whole or perhaps changes in the prevalence of mental health disorders among the general population. Also, discussion of PTSD sharply increased in 2012, which may reflect the burgeoning discussion of mental health coinciding with the return of troops from Iraq and Afghanistan (e.g., the Presidential Executive Order #13625 establishing the Interagency Task Force on Military and Veterans Mental Health was issued in August 2012). Corresponding with this, I see a spike in military or veteran-related mental health discussion in mid-2012, followed by a decrease. However, caution is needed as this trend line was derived from very little data and the confidence intervals for the trend line are wider than the full range of values depicted in panel C. There is very little confidence in the specific pattern of military and veteran-related mental health discussion. Other trends of note include discussion of depression increasing until 2013 then slightly declining and discussion of bipolar disorder increasing from 2009 to 2012 before slightly declining.

The change in Tweets containing stigmatizing content over time is depicted in Figure 5.1, Panel D. The proportion of mental-health relevant Tweets that are coded as stigmatizing are declining over time, with a steady decline since a peak in 2011. In reviewing the decline of stigma seen in Figure 5.1, I analyzed the trend in stigma for each individual characteristic. The decline that is seen in the overall data set is occurring in all of the individual characteristics and at approximately similar rates. This suggests a greater awareness and acceptance for mental health issues overall and not simply greater awareness of a single common condition.

Figure 5.1 Time Trends in Tweet Volume by Tweet Type and Topic



Note. Any data point with fewer than ten Tweets and any calendar day with less than 50 Tweets total (across categories) have been excluded. Each data point has been normalized by dividing by the total volume of mental health-relevant Tweets. Trend lines were fitted using local polynomial regression (LOESS) – a non-parametric method of fitting a population curve. The shaded areas around the curve represent the 95% confidence interval of the estimate.

Campaign Twitter presence

The volume of campaign-related Twitter activity is too low to assess whether the campaign activity is affecting the overall Twitter conversation about mental health

To better understand how campaign-related Tweets were disseminated, I identified Tweets as either being posted through official channels (i.e., the Twitter accounts that the campaigns told us were used for dissemination, see Table E.9) or through any other channel (i.e., an "unofficial" channel). This allows me to distinguish between Twitter activities initiated by the campaign versus activity propagating to other users.

Table 5.1 Official Twitter Accounts

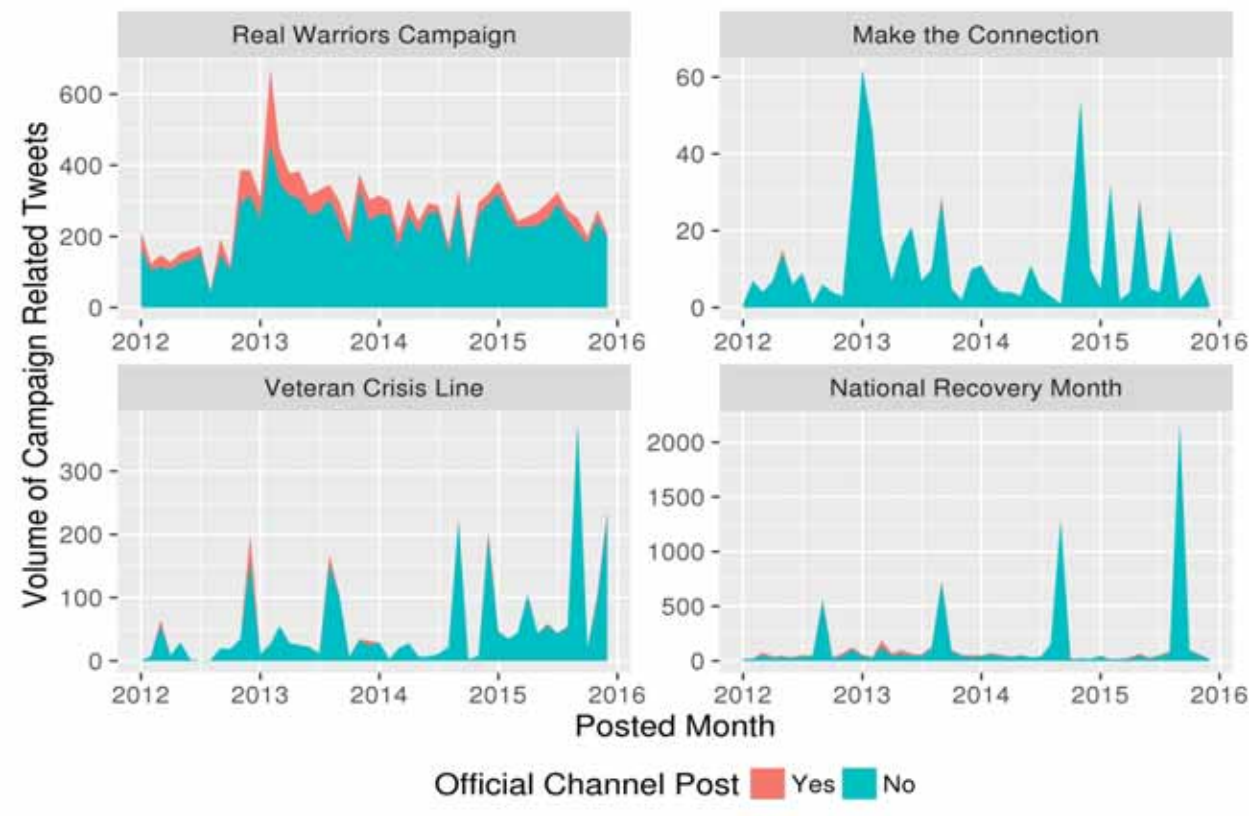
Campaigns	Associated Accounts
Real Warriors Campaign	**@RealWarriors** @DCoEPage @MilitaryHealth
Make the Connection	@VA_OEF_OIF @VAVetBenefits @DeptVetAffairs @VA_PTSD_Info @VeteransHealth
Veterans Crisis Line	@VA_OEF_OIF @VAVetBenefits @DeptVetAffairs @VA_PTSD_Info @VeteransHealth
National Recovery Month	**@RecoveryMonth** @SAMHSA

Note: Bold accounts are dedicated accounts for campaigns.

Official and unofficial campaign-related Twitter activity are shown in Figure 5.2. Most Tweets about the campaigns are from non-official channels, suggesting that campaign messages are propagating, at least to some degree, through spreading on social media. Between 47 and 670 RWC-related Tweets per month occur throughout the monitoring period, with 16 percent from the official RWC Twitter account. The search yielded only three MTC Tweets through official channels, though as noted in Chapter 1, the common use of "make the connection" as a phrase posed some challenges to identifying this campaign's Tweets. More Tweets (between 1 and 62) occur each month through unofficial channels. Between 1 and 377 VCL-related Tweets per month occur throughout the monitoring period, with 5 percent Tweeted by the official VCL Twitter account. Given NRM's focus on September as National Recovery Month, they see low volumes of Twitter activity with spikes in both official and unofficial activity corresponding with September of each year. The number of Tweets in these months range from 565 in 2012 growing to 2180 in 2015, with 1 percent coming from official channels.

Though I aimed to determine whether trends in campaign-related activity aligned with the trends observed in the previous section, the volume of campaign-related activity was not sufficient to conduct analyses. As noted earlier, there was a general increase in discussion of general mental health and informational content and a reduction in stigmatizing content during the time that these campaigns have been active. However, I cannot conclusively link these broader trends to campaign activities.

Figure 5.2 Tweet Volume by Month from Official and Unofficial Accounts



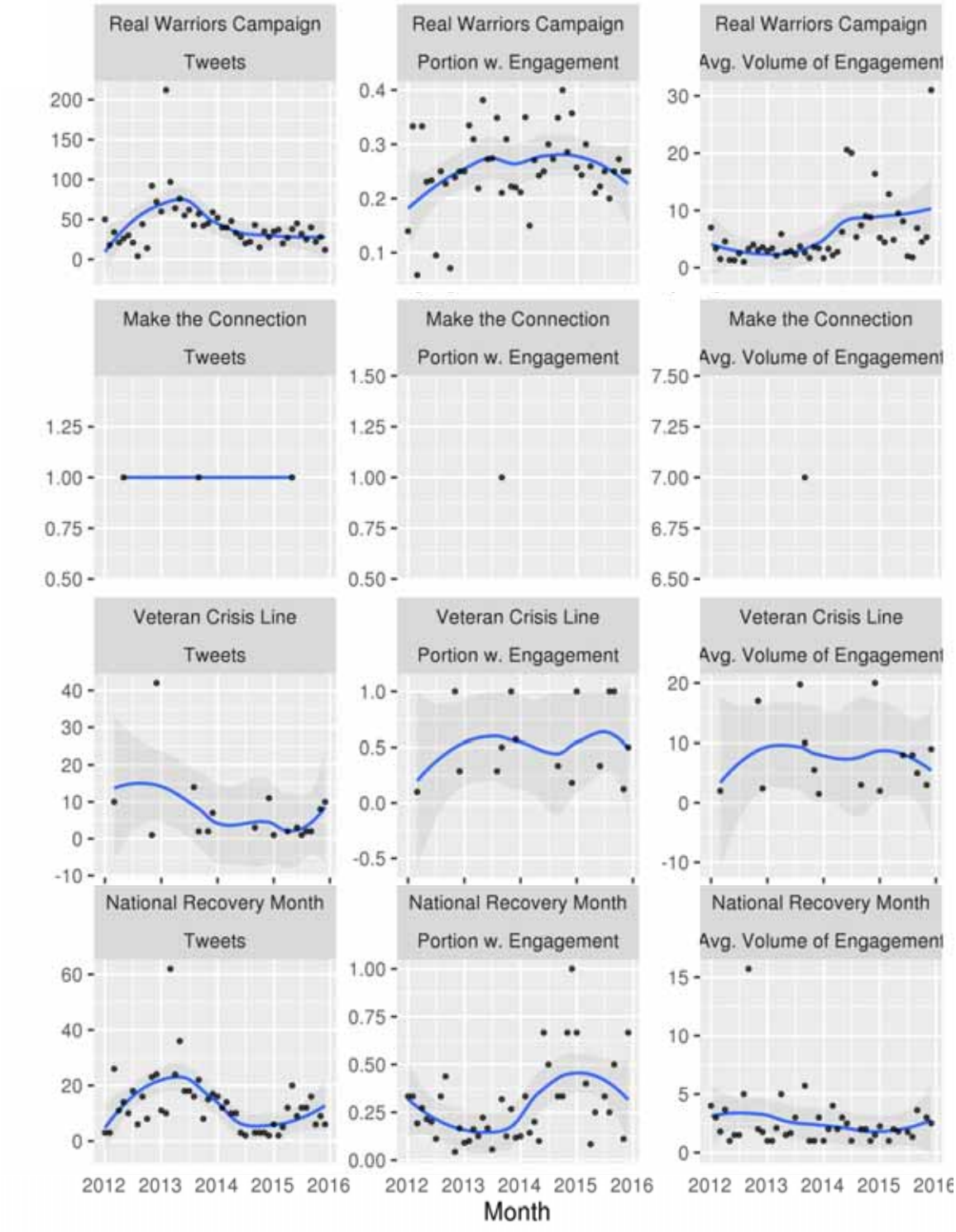
Note: This is a stacked (not overlapping) chart.

Three of the four campaigns show positive signs of engaging other Twitter users to reTweet messages

To understand how well campaigns generate interest in their messages, I considered all messages posted by official channels and identified any messages with the same content coming from a non-official channel within 30 days of the initial official post (i.e., a reTweet). I consider these reTweets to be a good measure of how much engagement the campaign materials are generating. Figure 5.3 shows volumes of messages posted by official channels, the portion of messages that receive reTweets by non-official channels, and the number of reTweets that such messages receive. For RWC, approximately 25 percent of its posts generate some engagement from non-official channels, and this volume of engagement increased over time even as the volume of official Tweets declined. Due to limited official MTC Tweets, I was unable to draw conclusions about engagement with those Tweets. VCL Tweets resulted in high volumes of engagement with 25 percent of messages receiving reTweets and reposts. NRM Tweets resulted in reTweeting of 25 percent to 50 percent of the Tweets reTweets and repost. I am not aware of any common level of retweets for public messaging campaigns, there is some older evidence that this level of retweets is in line with Twitter as a whole[165].

[165] (Suh et al., 2010)

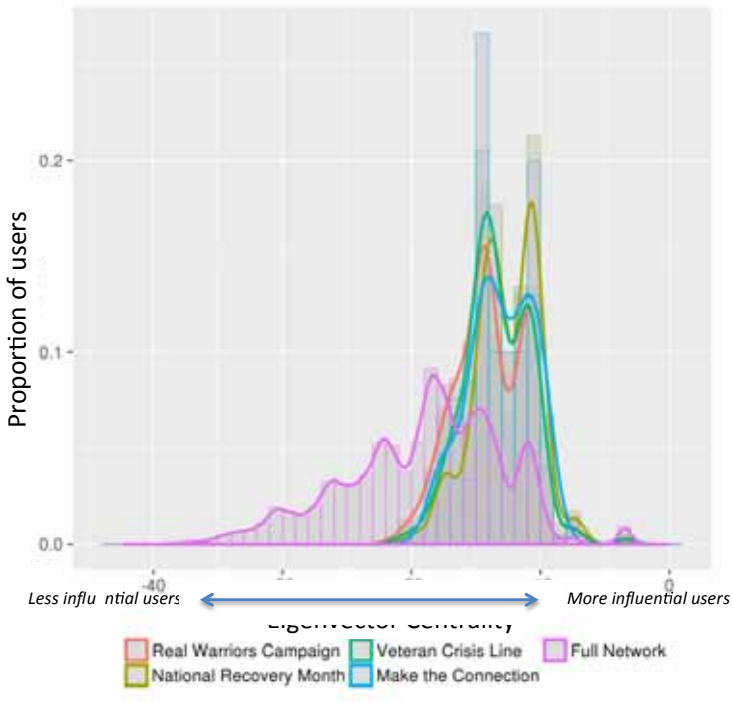
Figure 5.3 Twitter Engagement with the Campaigns



'Tweets' (column 1) shows the volume of Tweets posted by official channels related to each campaign. 'Portion with Engagement' (column 2) shows the proportion of official Tweets which generated reTweets or reposts within 30 days of initial posts. 'Average Volume of Engagement' (column 3) shows the average number of reposts/reTweets generated by an official Tweet that creates engagement.

Finally, I wanted to understand the type of Twitter users that were engaging with the military mental health campaigns. I computed eigenvector centrality (EV) as a proxy of how 'important' a user is to the network. Those with high scores were central and well connected within the conversation network and are therefore able to connect messages across conversation participants. Figure 5.4 depicts the distribution of EV scores of users who engaged the campaigns. Each of the campaigns were reaching users of similar EV levels, possibly because they were reaching the same population of users. Also, users who were engaged with the campaigns were more influential than the average user in our data set of authors who posted Tweets relevant to mental health. This result signifies that individuals that are active members of the conversation are including campaign messages within their communication. It does not mean that that the campaigns are reaching central users at a higher rate than other users, but it does indicate that the highly connected users are not ignoring the campaigns and are in fact engaging. The low volumes of official posts from three of the four campaigns combined with the kind of engagement that the campaigns are able to generate suggests that there the DoD would be able to increase the value of the campaigns by increasing volume posted by the campaigns.

Figure 5.4 Centrality of Campaign Engaged Users within Mental Health Focused Social Network by Campaign



Eigenvector centrality has been log transformed. Higher values (i.e., values closer to 0) indicate that users are more influential.

Engagement with the campaign is visible in the social network

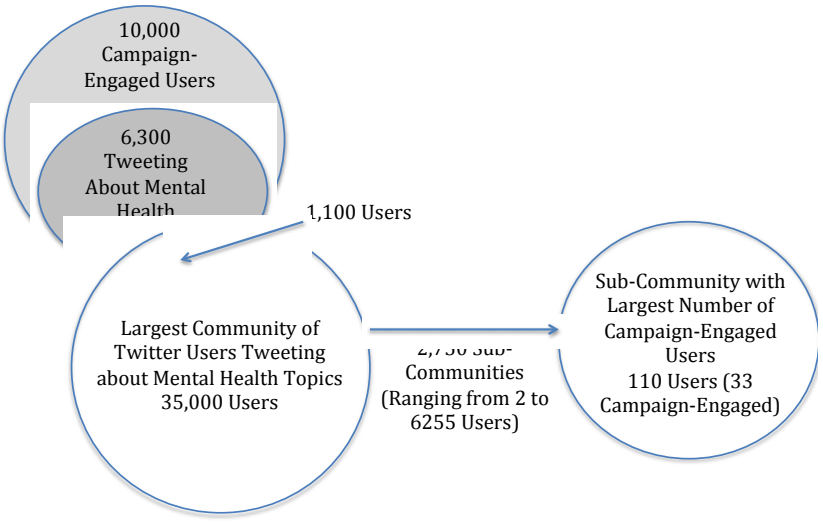
I used the social network created to understand the characteristics of communities to which campaign-engaged users (i.e., those who mention campaigns in their Tweets or reTweet campaign-related messages) belong. I then identified campaign-engaged users who were present in the social network. I found that of the 10,134 campaign-engaged users, 6,343 also were present in our network of users who made mental health-relevant Tweets[166]. I examined Tweet types and topics for campaign-engaged users present in the social network (see panels A and B in Figure 5.6). These campaign-engaged users posted primarily informational content but very little content containing mental health resources. They also posted about general mental health topics more so than any other topic.

Although I could not establish a relationship between trends in mental health discourse and the campaigns, I examined the type and topic of Tweets among the 'sub-community' of Twitter users containing the most campaign-engaged users (i.e., those who mention campaigns in their Tweets or reTweet campaign-related messages). To identify this sub-community I first identified that out of 10,134 campaign-engaged users, 6,343 were also present in the network of users Tweeting about mental health. I then examined the community membership of campaign-engaged users Tweeting about mental health, finding that 1,127 of 6,343 engaged users belong to a single community of nearly 35,000 users[167]. This group of 35,000 users was the largest community Tweeting about mental health and was the community that contained the largest number of campaign-engaged Twitter users. A community is a cluster of Twitter users that Tweet and re-Tweet each other (at least once). These users' Tweets were similar in type and content to those of campaign-engaged users (see Figure 5.6 panels A and B). Because this 35,000 member group was so large, it resembled very closely the general population in the overall network, making it a component of the overall network and likely not an actual community. Therefore, I explored sub-communities users within the body of 35,000 users who are more interconnected (i.e., five or more Tweets or re-Tweets), finding 2,698 such sub-communities ranging in size from clusters of 2 users to clusters of 6255, with 133 sub-communities of membership greater than 25 users. I identified the single sub-community containing the largest number of campaign-engaged users. This sub-community has 110 members, 36 of whom are campaign-engaged users (Figure 5.7). The process of this selection is depicted in figure 5.5. We characterize the content Tweeted by the full community in panels C and D of Figure 5.6.

[166] It was expected that not all campaign-engaged users would appear in our network. This is largely because we used a sampling approach to generate each data set, and not a census of all Tweets. In addition, the automated coding models used to code Tweets are probabilistic in nature and some false negatives are expected to occur when coding Tweets for content.

[167] Exact membership is 34,749 users but for purposes of discussion we refer to this community as 35,000 members.

Figure 5.5 Identifying a Community of the Campaign Engaged Users



I examined the type and topic of Tweets among the sub-community containing the most campaign-engaged users (see Figure 5.6, panels E and F). These sub-community members' Tweets were largely self-focused and contained informational content and mental health resources. Few of the Tweets involved appropriation. In terms of topics, the Tweets focused on mental health concerns related to depression and PTSD, along with Tweets about general mental health topics.

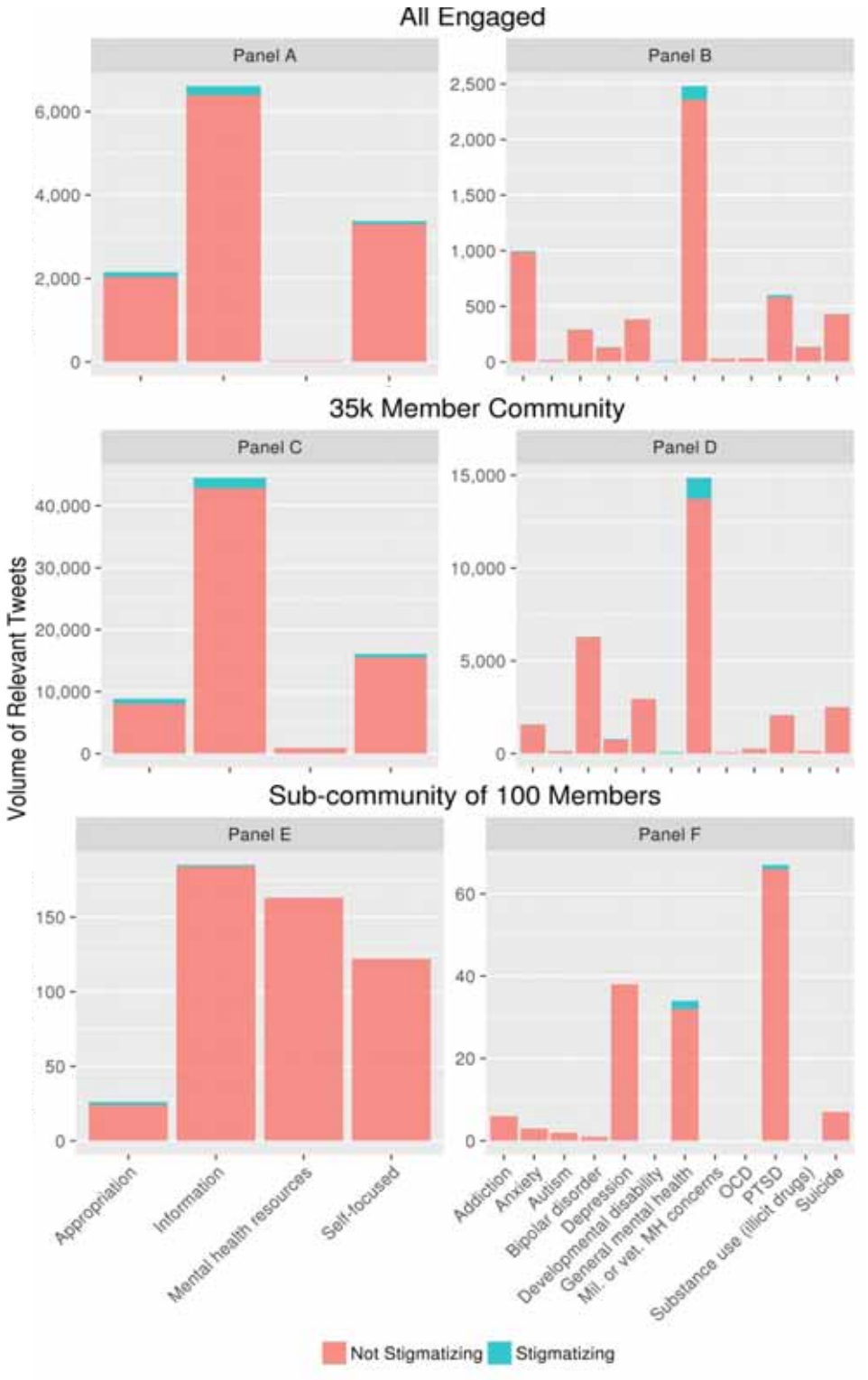
To learn more about this sub-community, I was assisted by an experienced RAND researcher who reviewed user names and their connections within the community (see Figure 5.7). The researcher visited users' Twitter pages and used their account names, photos, and brief self-entered descriptions to determine which of the following mutually exclusive categories applied:

- Official government account (n=27) – Labeled as an official account for an entity within the federal or state government (including military services and VA locations)
- Military or veteran-related account (n=29) – User self-identified as being a service member or veteran or having an interest in supporting service members or veterans
- Health-focused account (n=11) – User self-identified as having a focus on physical or mental health topics
- Other (n=43)– Users who did not fall into the previous three categories or whose accounts have been deleted or suspended.

Examining the accounts featured in the network in Figure 5.7, I see a mix of accounts of all types, with official government accounts serving as entities with the most connections to other users of diverse types[168]. This network graph suggests that although only one third of the users in the community were categorized as campaign-engaged users, there is potential for the rest of the users to see campaign-related activities as they propagate throughout the network.

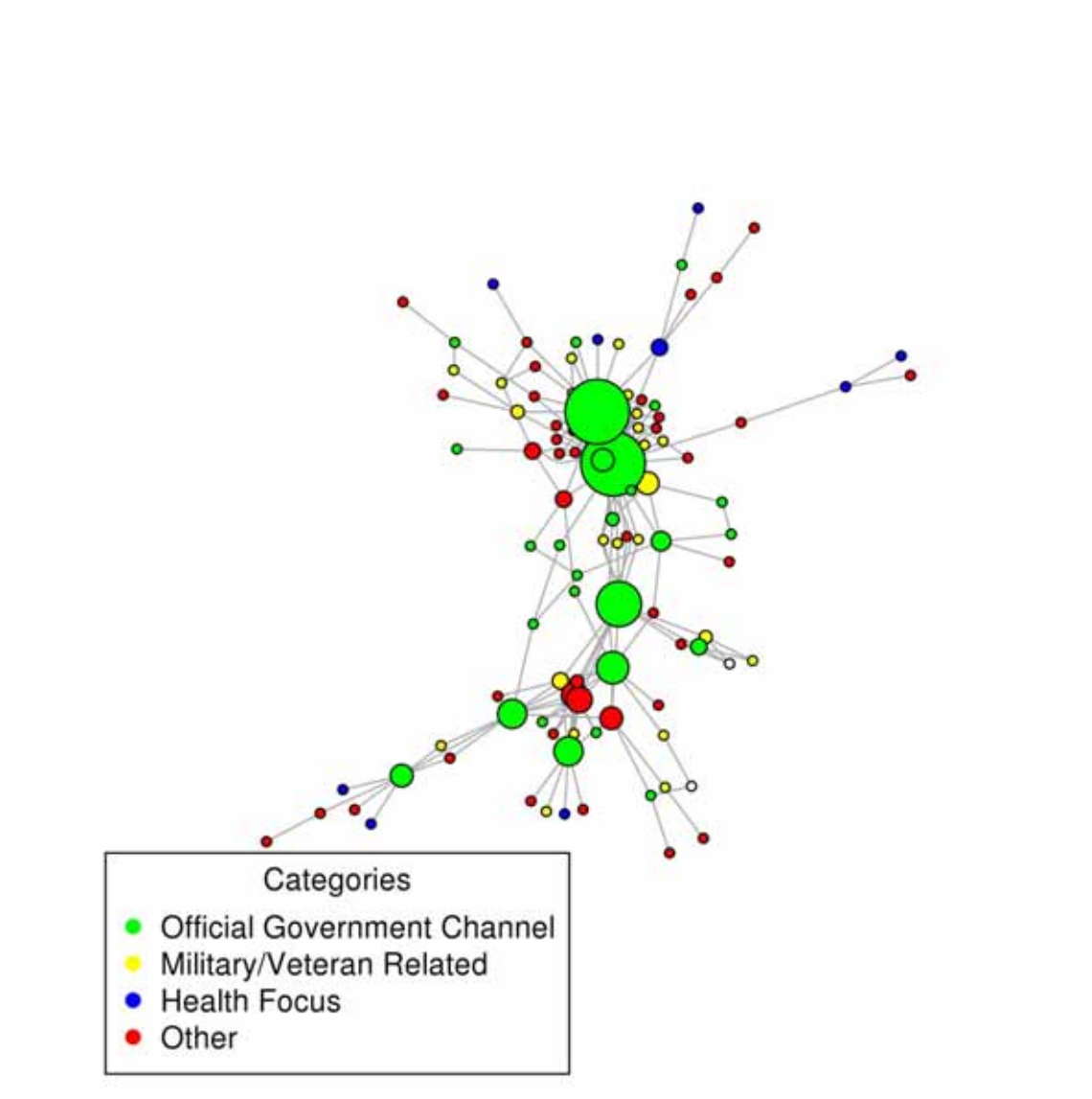
[168] While the central nodes may be able to broadcast the message, they may not be as effective at altering attitudes of the community, see Petty and Cacioppo (2012)

Figure 5.6 Centrality of Campaign Engaged Users within Mental Health Focused Social Network by Community



The graph shows the topic and stigma distribution for the population of users engaged with the campaigns, the overall community that most of those users belong to, and the sub-community that has the largest share of campaign-engaged users.

Figure 5.7 Network of Actively Engaged Users



Network of single sub-community containing the largest number of campaign-engaged users (n = 110). This graph is meant to capture the interaction between clustered active users engaged with the campaigns. The nodes are labeled according to user's Twitter handle and colored according to category of user account type designated by the research team. Nodes are scaled according to number of connections. All nodes with more than 15 connections are set to 15, and all nodes with less than 3 connections are set to 3 for the scaling step only in order to enhance readability.

Chapter 6: Comparing Dissertation Findings with the Literature

This dissertation analyzes Twitter content: exploring questions about mental health topics, stigma levels, time trends, community organization, and public messaging activity. However, the generalizability of the findings of this work to offline reality is uncertain. There are legitimate and pervasive criticisms of social media data: the demographics of individuals sampled are difficult to know precisely, online behavior may be very different from offline behavior, and there is no way to determine if the statements made online are true or accurate[169]. Researchers have attempted to answer these questions head on, in particular developing smart ways of inferring demographics: assessing location based on the social network of individual or gender based on account name and language of posts[170]. However, these approaches are in early stages and cannot fully address the concerns listed. It is uncertain if the results found in this work generalize beyond the world of Twitter conversations. To determine how my findings relate to off-line behavior I compared the results of my analysis with traditional (survey and experiment based) academic literature. I found that the social media results were broadly in line with the traditional literature and the differences that did appear were themselves interesting findings. Additionally, doing the research to compare the results showed that traditional data sources and social media data are often best for different questions and can be used together to better social science research.

This analysis was done by searching the academic literature for work that looks at questions similar to this dissertation but uses 'traditional' data to study them. The five areas of this research where I was able to find overlap with tradition literature are prevalence, stigma, type of message, community formation, and efficacy of public messaging. This search included looking for the prevalence of mental illness conditions in the U.S. population to compare to the volumes of Twitter discussion in different mental health related topics. It also included an examination of the amount of stigma associated with different topics and how these amounts change. Additionally, a search was conducted looking for information about how frequently people focus on themselves vs others when speaking about mental health and how individuals build communities around certain topics. Finally, this literature analysis explored other evaluations of public messaging campaigns that leverage Twitter.

To answer these questions, an extensive though not exhaustive literature search was performed. The strategy was to rely on indexing features of Google Scholar. This literature search used keyword queries for the areas of interest listed above — prevalence, stigma, type of message, community structure, and public messaging efficacy —and adjusted and repeated until an academic article that directly answered the question was found. That article would be used as the basis for the use of Google Scholar's 'related articles' feature, which finds 100 closely associated articles. Each resulting article was reviewed for relevance to the question. If there was an insufficient volume of relevant articles a different highly relevant article would be identified and the 'related search' repeated. This strategy was continued until either 20 relevant articles were assembled or

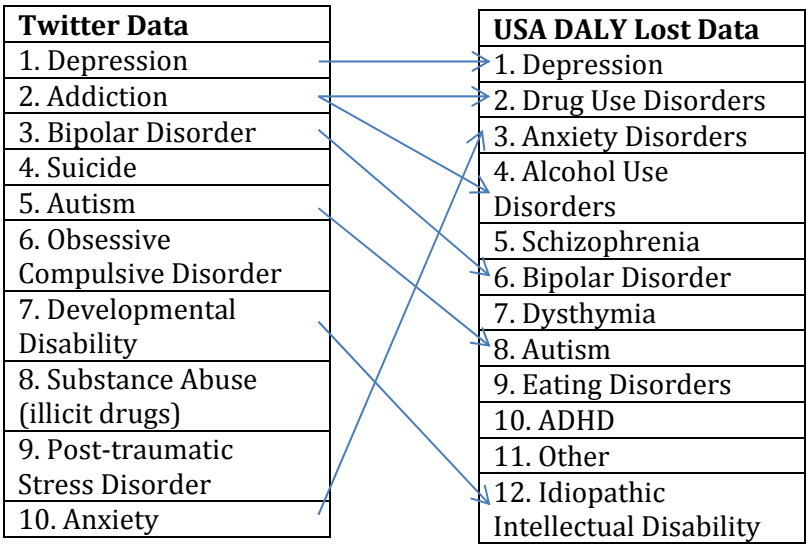
[169] (Boyd and Crawford, 2011)
[170] (Mislove et al., 2011)

the same relevant articles began to appear in repeated searches. Recent publications and meta-analyses were prioritized whenever possible.

This work is not a formal-meta analysis and does not claim to have exhausted the literature. The challenge is that often questions that naturally come from social media data are not asked with traditional data sources. I was unable to find a substantive body of literature on attitudes towards specific mental health conditions, or on whether people focus on themselves or other when it comes to mental health, and there is little in standard forms of evaluating public messaging. Academic literature relying on traditional data sources copes with small sample sizes and is focused on demonstrating the presence or absence of relationships. The large volumes of data in social media make it well suited for the study of narrower topics than traditional data sources. Traditional sources focus on how mental health is stigmatized, social media is better at asking how different conditions are stigmatized. Whenever possible, literature was identified that performed such comparative analysis and was related to the social media findings. Comparing survey and experimental literature with social media analyses cannot be done directly as measures used in the two analyses are not directly compatible. Instead, the focus of this analysis is on rank ordering. Outcome variables of interest are ranked in both sources and the relative ranking is compared across the two. The goal is to assess whether the relationship in outcome variables is consistent across the data sources.

Popularity of topics in social media tracks impact of conditions on public health

Figure 6.1 Prevalence of Mental Health Conditions



Note: Twitter Data ranking represents the relative volume of Tweets in each category of interests. USA DALY Lost Data comes from Murray, Christopher JL, et al. "The state of US health, 1990-2010: burden of diseases, injuries, and risk factors." Jama 310.6 (2013): 591-606.

To understand if the relative volumes of content for various conditions are meaningful, they are compared to the rank order of the real world health impact of various conditions. The impact of the conditions can be measured as the number of Disability Adjusted Life Years (DALYs) lost in the United States as a result of each of the conditions. DALYs are a prevalent (and at times criticized) public health metric utilized by the World Health Organization and others[171] that combines years of life lost because of premature death with the number of 'healthy' years lost due to disability. Healthy years lost is a way to capture that a year living with an illness is not equivalent to a healthy year of life. The argument for comparing Twitter volumes with DALYs is that people's attention to specific conditions ought to be broadly consistent with the impact that those conditions have on the population. This is not a certain relationship; people may be more focused on conditions with visible symptoms like schizophrenia or on publically discussed topics like drug addiction. However, as there is no survey or experimental literature that looks at how interested people are in discussing various mental health conditions, the 'burden of disease' approach is selected as a proxy. The relative impact of mental illness in the United States [172] is consistent with the relative conversation volumes for those conditions on Twitter, see Figure 6.1. For the majority of the conditions, the relative volume within the data set is consistent across the two measures (Twitter volume and DALYs). Depression, addiction, bipolar disorder, autism and intellectual disability are

[171] (Li, 2014)
[172] (Murray et al., 2013)

all in consistent order. This list spans the full range of the most prevalent disorders and illustrates that social media is consistent with survey and census data.

The taxonomy of mental health conditions varies across the two data sources. The Twitter analysis included a general addiction category, which is in line with the results in the DALY data set, but it was not possible to identify volumes of Tweets related to specific addictions to match the DALY data set. Also, schizophrenia was not a classified feature, so it is not known how it would place within the rank order set. PTSD was a condition relevant to the case study for this project, so it was classified but it was not very common. It is also not a major source of DALYs lost. It is also notable that the Global Burden of Disease (GBD) study does not include suicide as a separate category (as it is not a mental health condition and Years of Life Lost(YLL) is a more reasonable metric than DALYs) but it seems likely that suicide is a major source of DALYs lost and is appropriately high in the list of Twitter topics.

There are notable outliers. Anxiety was the least common condition in the Twitter data but is the third highest driver of DALYs lost. Anxiety was among the less frequent conditions in the training data set (though frequent enough to produce a robust model). It is possible that anxiety disorders are simply not a mental health condition that many people understand. ADHD is another outlier, and was not detectable in any significant volumes in the Twitter data sets despite efforts to classify it. It may be that, as a relatively low impact condition that is very common, it simply does not generate discussion. It is also of note that intellectual disability had a high volume in the Twitter data set. These tweets were also almost entirely stigmatizing in nature so it is likely that much of the developmental disability content that we identify is from pejorative use and does not actually reflect the prevalence of the condition as a topic of interest.

Condition specific stigma information is limited and agreement with social media is mixed

There appears to be no direct way to compare the number of Tweets that are stigmatizing in nature with the number of people that hold some stigmatizing beliefs. Stigma is typically measured with various survey instruments and vignette based questions. Most often, such analyses focus on stigma towards general mental health. While my analysis included a measure of general mental health and the associated stigma, a single characteristic of interest compared across studies does not offer confidence in the credibility of social media data. Instead of focusing on single characteristics, I aggregated the relative stigma ranking across conditions in the literature and compared them to the relative rankings of stigma within my Twitter data set. The two available comparisons are time trends and disease specific results.

There is a general consensus between the social media analysis and traditional sources that stigma is declining but still very much exists[173,174,175]. The identified literature had no studies

[173] (Pescosolido, 2013)
[174] (Schomerus et al., 2012)
[175] (Angermeyer, Matschinger and Schomerus, 2013)

published after 2011 and no analyses that included more than three time points. The social media analysis contains data at monthly intervals and begins in 2009. Thus while the overlap in time is limited, the overall trend is consistent across the two data sources.

Beyond comparing stigma across time points it is possible to compare the relative stigma of different conditions. Most of the literature related to stigma focuses on stigma in mental health in general but there is some work that compares condition specific stigma. Table 6.2 lists the ten papers identified that had comparative stigma measures. The papers did not all consider the same conditions, but collectively cover eight different conditions. The most commonly assessed conditions are alcohol addiction, depression, and schizophrenia. The least common are dementia and eating disorders.

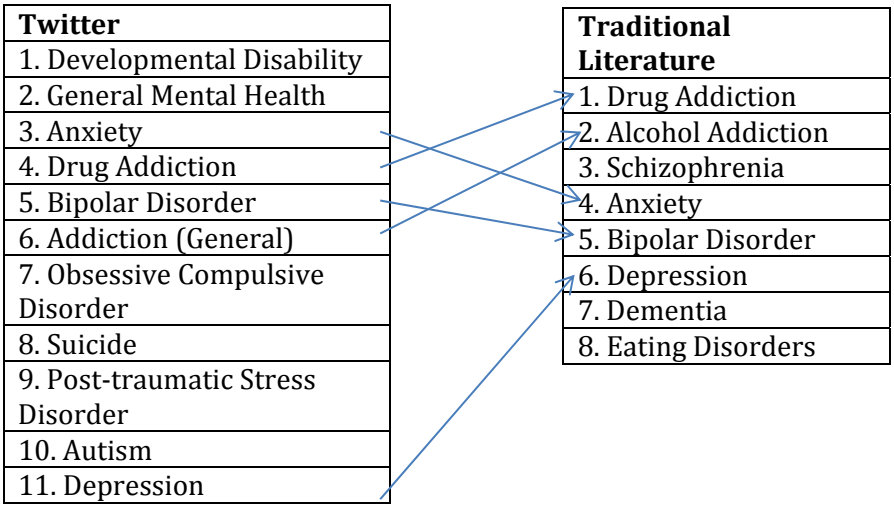
Table 6.1 Ranking of Conditions by Journal Article

	Rank Order of Conditions (1 = most stigmatized)							
Author	**Alcohol Addiction**	**Anxiety**	**Bipolar Disorder**	**Dementia**	**Depression**	**Drug Addiction**	**Eating Disorder**	**Schizophrenia**
(Angermeyer and Dietrich, 2006)	2	4			3			1
(Angermeyer, Matschinger and Schomerus, 2013)	1				3			2
(Crisp et al., 2005)	3	4		5	6	1	7	2
(Evans-Lacko et al., 2012)			2		3			1
(Martin, Pescosolido and Tuch, 2000)	2				4	1		3
(Pescosolido, 2013)	2				4	1		3
(Rao et al., 2009)	3					2		1
(Schomerus et al., 2011)	1				3			2
(Schomerus et al., 2012)					2			1
(Wood et al., 2014)		2			3			1

A casual review of the results table suggests some patterns of which conditions are more or less stigmatized, but a consensus ranking is not obvious. Additionally, the fact that different papers rank different conditions means that most voting systems (such as a single transferable vote) fail to produce consistent results. I created an approach to identify the most common ranking inspired by the work of Cook et al. (2005). The approach seeks to identify an overall ranking that would disagree with the fewest of the ten articles. In order to do this, I generated every possible ranking of the eight conditions, and compared each ranking with the ranking obtained from the literature. Note, this is only possible as the choice set is small. The number of permutations is given by n!, where n is the number of choices and the number of permutations becomes computationally

intractable very quickly. After checking all the permutations, a set of 53 permutations was found that had a minimal number of disagreements with the rankings found in the literature – all 53 rankings disagreed with three of the 10 articles identified in Table 6.1.
Then the most common item at each ranking position for these 53 best permutations was identified. This approach yielded a ranking that did not include dementia or eating disorders (they were not the most common conditions at any ranking position). As such, I added them into the final ranking as less stigmatized than depression, based on Evans-Lacko et al. (2012). The final ranking for stigma by condition according to the literature is shown in Figure 6.2, along with a comparison to the findings of this research project.

Figure 6.2 Relative Levels of Stigma



Examining the levels of stigma identified in the research project relative to the levels of stigma in the literature, we see a mixed level of agreement. Anxiety, bipolar disorder, and depression are consistently ranked across both data sources. These conditions are responsible for a large burden of disease and feature prominently in popular discourse. An agreement in the ranking of these conditions bolsters the predictive power of social media data. There is, however, disagreement between the data sets in the stigma levels attributed to addiction categories, like drug addiction, alcohol addiction, and addiction overall.

A number of factors may account for the observed differences in stigma rankings between Twitter and traditional literature. One, it may be possible that the demographics of Twitter users fundamentally alter the conversations and data obtained compared to survey and experimental studies (which strive for representative samples). Twitter is more popular among younger (18-29) internet users than older (50+) users[176], it may be that younger individuals are less concerned with the illnesses associated with substance abuse, having not witnessed the long term effects of those conditions. The Twitter users may have less experience with addiction syndromes and therefore do not stigmatize them as much. It may also be possible that the Twitter population is simply more empathetic towards addiction mechanism based on alternate demographic features.

[176] (Duggan, 2015)

A second potential reason for the disagreement may be a modeling failure. Stigma was among the more challenging aspects to model in the analysis and the results may be a reflection of modeling imperfection. Additionally, in order to create an effective model this analysis relied on the manual labeling of a single expert coder as opposed to the consensus vote of trained research assistants. If the single expert coder had a tendency to label fewer Tweets related to addiction as stigmatizing, this would be reflected in the model.

Third, it may be possible that addiction content related to substance abuse is difficult to detect, as addiction is a term that is used commonly in everyday language. This explanation is the least likely as the volume of addiction content was in-line with the general burden of disease metrics above. This make it more likely that the content itself was effectively identified. But it is still possible that some substance abuse related content was lost. Future work related to this topic would benefit from a careful analysis of addiction content.

Self-disclosure is important to individuals' mental health, but there is little literature on the kind of discourse people have about mental health

An advantage of social media as a data source is that it allows the researcher to observe how people discuss a topic outside of the narrowly defined bounds of the lab or survey. The literature search did not result in any findings from academic work using 'traditional' data characterizing how people discuss mental health. Conversely, this research analyzed not only the mental health topics that discuss but also how they discuss them: 40% of the posts are about the self, 30% contain mental health information, 30% contain appropriation, and slightly less than 1% of posts are about resource sharing. In this work, I had also hoped to classify posts that were made about an individual other than the user but, it proved impossible to create a reliable model.

Social media is unique in its ability to capture the naturally occurring conversations of people about a broad range of topics and at a large scale. Traditionally, the observation of peoples' language has been primarily the concern of sociolinguistics. Sociolinguistics developed methods to work around the observer's paradox, the need to observe how people use language while they are not being observed[177]. These methods – participant observation, interviewing, anonymous observation, and phone interviewing[178] – are not able to provide the scalability and generalizability of questions that social media can. By relying on the large volumes of consistently generated and recorded data in social media conversation, the observation of naturally occurring language can move towards greater breadth and generalizability than was possible with prior sociolinguistic methods.

Twitter can be utilized to understand the structure of natural occurring conversation while avoiding the observer paradox. A key linguistic element of the Twitter conversation is self-focused content. I found that self-focused content is the most prevalent type of Tweet related to mental

[177] (Labov, 1972)
[178] (Tagliamonte, 2013)

health. Self-focused mental health content has also been studied in more traditional academic work. There is a great deal of research into the value of self-disclosure of mental illness (communicating one's own mental health challenges) and the therapeutic benefits of discussing negative life events. There is robust research demonstrating that an individual's stigma towards mental illness is reduced by knowing someone with mental illness[179]. Additionally, there is ongoing research into health benefit of self-disclosure. Most commonly, this is studied in the context of expressive writing (a therapeutic tool where individuals write about negative experiences without social feedback). The method was first proposed by Pennebaker and Beal (1986). The most recent meta-analysis covered 146 studies[180]. The exact effect of expressive writing and self-disclosure are contested for various populations and various methods of analysis[181]. However, it is clearly a popular intervention and Twitter provides a new avenue to study and utilize it. On Twitter, self-focused content is the most popular form of post. Studying social-media self-disclosure and its role in mental health may be a natural extension of the earlier self-disclosure/expressive writing literature. Already there is some work on self-disclosure online which focuses on the privacy ramifications of such discourse.[182,183]. The academic work in self-disclosure cannot corroborate the research finding of 40% of posts being focused on the self, but it does suggest that social media analysis does offer insight into meaningful aspects of mental health discourse.

Community identification and support help to reduce stigma and self-esteem issues, but do not always improve clinical outcomes

Analyzing a social network such as Twitter allows for the analysis of community structure. This dissertation work found that there exist communities centered on mental health conversation. Additionally, these communities are heterogeneous. There are communities that are focused on particular mental health topics, have a particular way of conversing, or express common levels of stigma towards mental health. Finally, it is notable that the majority of communities have a lower level of stigma discourse than what is seen in the general Twitter conversation about mental health. These online communities are really groups of participants in the mental health conversation and do not represent communities of general social ties. Such conversations may be important in the offline world as well, and I was interested in seeing if the academic literature identifies such conversation communities and their impact off-line behavior.

The literature review finds that peer-support groups and community identification are popular topics for research. The top level finding is that peer-support is effective for reducing the feeling of self-stigma, increasing self-esteem, and agency[184,185,186]. However, improvement in clinical

[179] For an overview focused on stigma impacts (and some mental health outcome impacts) see Hyman (2008)
[180] (Frattaroli, 2006)
[181] (Ellis and Cromby, 2011)
[182] (Pavalanathan and De Choudhury, 2015)
[183] (Joinson and Paine, 2007)
[184] (Proudfoot et al., 2012)
[185] (Cook et al., 2011)
[186] (Repper and Carter, 2011)

outcomes, such as hospitalization, has not been robustly demonstrated[187]. Additionally, there is evidence that the benefit of community support declines with increased severity of symptoms[188].

The literature demonstrates the value of social support for coping with stigma[189,190]; therefore we would expect individuals who are well enough to engage socially with a networking platform such as Twitter to use the platform to engage with others who might provide such a community. It is also of note that the traditional academic literature tends to view peer-support and communities of mental health discussion as homogenous[191,192]. The social media analysis demonstrates that the type, topic, and stigma levels of various communities are very different and the impact of such groups should not be assessed uniformly. Additionally social media allows us to assess informal community organization while traditional data sources struggle to track and identify such groups outside of more formal clinical settings[193].

Public messaging campaigns underutilize Twitter; DoD performance not an outlier

To gain insight into the robustness of the analysis of the DoD campaigns conducted in Chapters 4 and 5, I review prior work on social media campaign evaluations. The DoD sponsored campaigns are an example of a public messaging campaign, a common intervention. Yet, the literature on how these campaigns perform on social media is not extensive. This is not surprising as social media as a pervasive social component is a new phenomenon, accessing social media data is challenging, and public messaging for public health has yet to develop tools and practices to utilize it. In 2011 the CDC released guidelines for social media use that offered little guidance beyond 'setting up an account' and 'keep followers engaged'[194]. Since then the space has evolved rapidly and more and more public health agencies and NGOs are hoping to leverage the network effects to communicate their message. When considering the volume of identified literature it is clear that there are few consistent and agreed upon standards for evaluating a social media campaign (on Twitter in this case).

In order to understand public messaging performance on social media performance, I utilized the framework outlined by Neiger et al. (2012). They identify five key areas where such campaigns provide value: market insights, establishing a brand and create brand awareness, disseminating critical information, expanding reach to more diverse audiences, and fostering public engagement and partnerships. I believe the DoD sponsored campaigns can be seen to provide value in all of the above elements. Neiger, et al. then go on to propose the four key performance

[187] (Lloyd-Evans et al., 2014)
[188] (Cook et al., 2012)
[189] (Quinn and Chaudoir, 2009)
[190] (Bourguignon et al., 2006)
[191] (Crabtree et al., 2010)
[192] (Goldstrom et al., 2006)
[193] Ibid.
[194] (Centers for Disease Control Prevention, 2011)

indicators for these goals: insights, exposure, reach, and engagement. Insights relate to data mining insights about attitudes of consumers and the nature of the 'market' for the intervention.. Exposure relates to the number of impressions. Reach is the number of people that are exposed to the campaign. Engagement is about how people respond to the campaign: retweeting, creating original posts, or actually taking physical actions offline in response to the campaigns. These criteria were used to outline the findings of the most recent evaluations of Twitter public messaging campaigns that could be identified in the literature.

Table 6.2 Key Performance Indicators from Recent Twitter Public Messaging Campaigns Evaluations

Citation	Title	Insights	Exposure	Reach	Engagement
(Bravo and Hoffman-Goetz, 2016)	Tweeting about prostate and testicular cancers: Do Twitter conversations and the 2013 Movember Canada campaign objectives align?	.6% of campaign associated tweets focused on the topic			
(Duke et al., 2014)	The use of social media by state tobacco control programs to promote smoking cessation: a cross-sectional study		23% update social media daily or more frequently	60% of tobacco control programs use social media, 61% of Twitter users have less than 10 followers per 100,000 adults in the state	
(Emery et al., 2014)	Are you scared yet? Evaluating fear appeal messages in Tweets about the tips campaign	17 million tobacco tweets, 77k tobacco control, 54K about ad, 245k relevant	39 million potential impressions		87% of tweets showed message acceptance, 7% showed rejection
(Gough, 2016)	Using Social Media for Public Health.. Brb [be Right Back]	There were a total of 417,678 tweet impressions			11,213 engagements and 1,211 retweets. Most engagement was humor tweets, most retweets – informative.
(Lachlan et al., 2014)	If you are quick enough, I will think about it: Information speed and trust in public health organizations	No correlation between speed of tweet posts and trusts in disaster			
(Livingston et al., 2014)	Another time point, a different story: one year effects of a social media intervention on the attitudes of young people towards mental health issues				Social media campaign reduced levels of stigma towards mental health

Citation	Title	Insights	Exposure	Reach	Engagement
(Neiger et al., 2013)	Are health behavior change interventions that use online social networks effective? A systematic review	56% of local health department tweets were about personal health, of those 58% were information, and 41% encouraged action. Of organizational tweets 52% was promotion, 35% was engagement, 13% encouraged action			
(Paek et al., 2013)	Engagement across three social media platforms: An exploratory study of a cause-related PR campaign				People engage better on platforms that they use more
(Price et al., 2015)	Improving understanding, promoting social inclusion, and fostering empowerment related to epilepsy: Epilepsy Foundation public awareness campaigns—2001 through 2013			From the posting on feeds of 457 people, reach 250,000 people	
(Ramanadhan et al., 2013)	Social media use by community-based organizations conducting health promotion: a content analysis	63% of Tweets from community based organizations are self-promotion. 21% retweets, 38% mentions	92% of CBOs posted daily as recommended		
(Thackeray et al., 2013)	Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month			1.3 million Tweets from 800k users, early spike with strong tapering, celebrities make the biggest impressions	Little behavior promotion, one way communication
(Turner-McGrievy and Tate, 2011)					Twitter support for intervention did not enhance outcome

Reviewing the various public messaging evaluations reveals a lack of consistent standards for evaluation and a lack of consistent results. Having examined the literature, it is difficult to say what the standard performance of a public messaging campaign looks like or what sort of questions an evaluator should ask of those campaigns. Below, we discuss the findings of this dissertation as they pertain to the four key performance indicators outlined in Table 6.2. The findings are compared to the literature whenever possible.

A large portion of this dissertation explored 'insights' in Twitter data for campaign evaluation – examining the 'market' for the campaigns. In analyzing the community organization of Twitter users who connect with other users about mental health, it was found that the campaigns reached better connected users than the average. Additionally, the communities that were most engaged by the campaigns were distinct from the average community with a greater focus on PTSD. The example community, which had the greatest engagement with the campaigns, contained and centered on accounts related to national security. The literature review was not able to identify any social network analysis in evaluations of public messaging campaigns on Twitter. Generally, insights are difficult to compare across evaluations as they tend to be topic specific and not about communication strategy in general.

When considering 'exposure' metrics, we have limited data on the DoD campaigns. The campaigns and other agencies associated with the campaigns posted a combined 2799 tweets over the 196 days that data were collected, for 14.3 Tweets per day on average. However, the RWC was responsible for the majority of the activity with 10.4 Tweets per day. At this volume, the campaigns overall created more exposure than the recommended 1 Tweet per day minimum[195,196]. The results on exposure performance in the literature are mixed. Only two identified evaluations included an exposure value and one finds that the majority of the evaluated public messaging campaigns do not hit the recommended threshold while a different evaluation finds that the majority do. More analysis is needed in the literature to assess the common rate of exposure for public messaging campaigns.

When analyzing the performance of the campaigns, this dissertation did not explore the 'reach' of the campaigns. This is the number of people who could have potentially seen the Tweets from the campaigns. In the literature, we see a large range of numbers for this sort of metric and typically the value is high, in the hundreds of thousands if not millions of possible impressions. The reason that this sort of metric was not included in this dissertation is twofold: the Gnip data portal that was used for data gathering does not allow for an automated large scale collection of follower lists, which is key to understanding who could potentially see the tweets, and also because this metric does not appear to be very reflective of campaign impact. Impression numbers are hard to interpret as they say nothing about how widely the message is actually read, or whether it generates engagement.

The analysis of the campaigns covered low and medium level 'engagement' as defined by Neiger et al. (2012). Low level engagement refers to the number of people who acknowledge

[195] (Duke et al., 2014)
[196] (Ramanadhan et al., 2013)

agreement with or preference for the content of the messaging campaign. Medium engagement refers to counts of people creating or sharing the content. High engagement refers to people taking action beyond the social media platforms based on the campaigns. Unfortunately, this dissertation did not obtain any data on high engagement activities. It was found that for the 2,799 Tweets posted by the campaigns over the 196 days of data collection, 20,808 Tweets that referenced the campaigns were posted by users not associated with the campaigns. Of the 27,99 posts by the campaigns and campaign associated accounts, 708 generated reTweets for a total 3232 reTweets. This is greater than the values found by Gough (2016). Beyond Gough, I was not able to identify literature focused on low and medium level engagement. The literature seems focused on deriving indications of high level engagement. This is difficult to quantify outside of narrowly defined problems, as capturing the offline behavior of high-level engagement requires both online and offline data gathering in tandem. As a result, the bulk of the articles that look at this topic provide valuable but broad ideas, like the value of celebrities in promotion and the comparative value of humorous vs. informative tweets to generate engagement.

There are examples of academic work which focuses on measuring the high levels of engagement. Bray et al. (2015) analyzed the efficacy of public information campaigns to increase use of emergency medical services in case of stroke, focusing on specific regions in Australia. Mellon et al. (2014) conducted a review of a similar intervention in Ireland focusing on time localized awareness campaigns. Livingston et al. (2014) used a multi-wave survey to capture the effects of a mental health awareness campaign specifically targeted at teenagers in British Columbia, Canada. In all three of the papers, there were very clearly defined outcome measure and a way to capture them reliably. The studies were also able to constrain the scope of potential impact. Bray and Livingston focused on geographically specific interventions while Mellon analyzed a campaign with very narrow active time. Without these constraints — clear outcome measures and specifically defined study population — it seems unlikely that a Twitter campaign could be evaluated for a high engagement.

Overall, it appears that the literature on evaluating public messaging campaigns conducted via social media has not reached maturity. There are lessons that are emerging but there does not appear to be a unifying standard that allows for comparison between campaigns. Currently, best practices are translated across campaigns in an anecdotal fashion, such as the advice offered in 'The Health Communicator's Social Media Toolkit' Centers for Disease Control Prevention (2011). Comparing the literature to this evaluation of the DoD campaigns, we find that these DoD campaigns were small (insofar as they pertain to Twitter) and had difficulty defining outcome measures. The evaluation conducted in this dissertation covered a wide swath of performance categories (as suggested in Neiger et al. (2012)). However, both designers and evaluators of social media campaigns would do well to consider the evaluation literature available before embarking on future projects.

Twitter appears to be a robust data Source but further verification is needed

The comparison of traditional literature to this Twitter analysis demonstrates consistency in several key areas. The relative volume of Twitter interest in specific topics matches the relative burden of disease of mental health conditions. Stigma associated with individual topics is consistent for a key set of conditions.

It also appears that Twitter provides novel ways to study important topics like self-disclosure and community organization. These topics are studied in the literature in normative ways: do they exist? Do they provide value? But, Twitter allows for comparative analysis: how do these things differ across people and groups? The literature focuses on the value of community and peer support for mental health conditions, Twitter allows us to examine how these communities differ. The literature is concerned with the value of people disclosing their mental health issues and concerns, while Twitter allows us to see if people chose to do so as opposed to speaking about mental health in non-personal terms.

Comparing this analysis to the literature also demonstrates places where social media analysis needs to mature. There is an inconsistency between the findings of this work and the literature on stigma associated with addiction. It is likely that more work is needed to understand how topics are separated out in the popular mind despite being interrelated in the academic setting. It may be that certain constructs, such as mental health and addiction, that can be effectively defined for subjects in a controlled setting, such as an interview, focus group or an experiment, are not viable when considering the free form conversation of social media. Additionally, there is no standard methodology or structure for the evaluation of public messaging campaigns that are focused on social media. There is an absence of standard metrics or standard questions. It appears that this evaluation of DoD campaigns covered a broad cross section of key areas of interest in evaluating social media campaigns, but the details of what exactly to measure are far from set in stone. Currently, evaluations of social media campaigns are not readily comparable.

The findings of this dissertation line up with key trends in the literature. More work is needed to understand where and how social media may differ from survey data and where social media can be a faster and cheaper source of research information. Additionally, the consistency across these two sources suggests that the findings of this analysis can be used to generate future hypotheses for experimental and survey research. Finally, by examining questions related to self-disclosure and community design, I illustrated where social media may be an essential complement to traditional research, providing a comparative analysis where traditional work is focused on normative analysis.

Chapter 7: Conclusions, Policy Implications, Limitations, and Future Work

In this chapter I discuss the main finding of this dissertation along with the relevant policy implications. The key findings are that Twitter is in fact being used by people to discuss mental health. The attitudes and norms related to mental health conversation are improving over time. People discuss these topics in self-forming groups and these groups and often have their own distinct characteristics. It also appears that public messaging campaigns and policy makers are under utilizing the potential of a platform like Twitter. There is evidence that that the results of this social media analysis correlate well with insights found in academic research which relies on survey data. I find that conducting social media analysis for policy purposes is doable, robust, and offers an opportunity for creative research.

Subsequently, I discuss the limitations of my research. The methods developed and employed in this work were not fully optimized and could have been more robust. With different time and resource constrains, the data collection strategy could have been iterated further, the training data set could have been larger, and the model further refined. After completing the work it became apparent that certain topics which are challenging to identify but are important (e.g. substance dependence and abuse) were not sufficiently emphasized in the design phase. As a result, I was unable to provide insight into certain important areas of mental health and stigma research. . Furthermore, this research does not provide meaningful information on the demographics of the people in our sample, even though the results seem to be in line with academic work that uses a representative sample. This prevents me from drawing insights on the behavior and attitudes of specific groups which may be of interest to policy makers. Social media analysis is fundamentally challenged by the fact that all the data used is proprietary and owned by technology firms. As such, data sources and data types which could have strengthened the conclusions of this research are not available. It was also found the goal of seeing the impact of public messaging campaign on the general Twitter conversation was impractical as the campaigns are too small in scale. However, fully capturing the effects of these campaigns on their proximal connections is likely not feasible given the way that Twitter data is accessed and priced.

Extensions of this work cover further exploration of the data and results obtained in these analyses as well new topics that can built upon the methods and approaches demonstrated here. A proposed idea for extending the work that was already done is looking more carefully at the various individual communities to see how they discuss specific topics and how these communities grow and change over time. Another potential extension is conducting linguistic analysis of various mental health Tweets to see how emotions, like fear and hope, associate with mental health topics. Finally, there is always an opportunity to explore other classification mechanism, like regression trees, to identify certain features more accurately. Possible new research projects that may build upon what was learned in this dissertation include a demographic analysis of Twitter users, a very narrowly defined evaluation of policy interventions, and development of a flexible tool to monitor the Twitter feed at various time points (proactively and retroactively).

The chapter concludes with a discussion of the overarching goals of this dissertation, a summary of the completed work, and proposed future research. With this work I hope to improve

the tools of social science for using internet behavior data and inform policy makers as to the relevance of social media for public policy. The completed work covers data parsing methods, measurement questions of classifying the conversation, examining time trends, and parsing the social ties of Twitter users. The findings cover why and how Twitter, social media, and internet behavior data is of use to social scientists and policy makers. Suggestions for further work include a refinement of the methods developed in this dissertation as well as deeper analyses of areas like demographics and communities.

Findings and policy relevance

Twitter discourse contains mental health discourse

Finding: This dissertation finds that people discuss mental health on Twitter. A keyword search of Twitter, followed by a filtering of the content for relevance, identified over 20,000 Tweets per day related to mental health. Much of this volume comes from distinct authors, which indicates a large collection of people using Twitter as a platform to engage with mental health topics. The conversation about mental health is diverse. This research identified a dozen different topics being discussed. The popularity of topics on Twitter was in line with the impact that various mental health conditions have on the overall population. This suggests that Twitter may be reflective of individuals' concerns for the various issues.

Policy relevance: The reality that people use Twitter to discuss mental health indicates that any policy intervention that seeks to connect with people on the topic of mental health should consider social media as part of its implementation. The conversation is happening on Twitter (and likely other platforms), and if a policy intervention, such as a public messaging campaign, seeks to engage with people's concerns about mental health, that intervention has to be present where the conversation is happening, including social media. Additionally, the finding that conversation about mental health is happening online provides an opportunity to expand policy research. Designing and measuring interventions, tracking public health, understanding public perceptions related to mental health can all be done with a cheap, large, available data source.

The reality that Twitter (and likely other social networks) is a platform for people to engage with topics of public welfare like mental health has implications for the policies that organization like Twitter may adopt and how they may be regulated. Currently, it is difficult for someone working in the public good to collect and understand Twitter data. The company could be much more transparent and allow for non-profit institutions to use the data more cheaply and more comprehensively. For example, I was unable to get sufficient data to fully map the social network of users and there is no opportunity for me as a researcher to request specific access. Additionally, the utility that Twitter (and platforms like it) provides for the public good suggests that the way it is regulated may have to be different from other technology companies. The idea of Twitter being regulated as a public monopoly has been suggested given the continued financial difficulties of the platform[197]. My finding of people using the platform to discuss and engage with matters of public

[197] (Bonasio, 2016)

concern suggests that developing a way to ensure that the platform remains pro-public use may be important.

People are getting better at discussing mental health

Finding: The online conversation is becoming increasingly respectful and aware of mental health challenges. There is a decline in the portion of mental health Tweets that are stigmatizing. This decrease can be seen consistently across all conditions. There is also a decline in appropriation related content and self-focused Tweets. This suggests that people are less cavalier about using mental health terms to intensify non-mental health topics and are perhaps less likely to self-diagnose various conditions in response to emotional stresses. This all indicates that people are more aware of the value of mental health language and are more judicious in its use. Additionally, there is an increase in the prevalence of mental health as a topic of conversation. Rather than speaking about diseases, people are discussing mental health generally. All of this suggests that the longitudinal trends indicate that people's attitudes and respect towards mental health issues is improving.

Policy relevance: The evidence of changing attitudes towards mental health suggests avenues and scope for policy intervention. The fact that attitudes are changeable suggests that policy interventions may be able to influence people's attitudes towards a variety of mental health topics. Since we observe people communicating more about mental health, public service announcements aimed at informing the public may have potential to inform interested individuals. The decline in self-referential discussion of mental health suggests that people are less flippant about mental health conditions and that interventions to increase individuals' awareness of their own mental health may be effective. Finally, the reality that these attitudes are changing, that people are demonstrating greater awareness of mental health concerns, and are increasingly discussing mental health as topic in its own right, all suggests that there is an appetite for messaging on reducing stigma and increasing understanding towards mental health. It may be that now is a time that individuals are receptive to these kinds of interventions to shape and influence public perception of mental health.

Mental health communities exist and are heterogeneous

Finding: In analyzing the content posted on Twitter this analysis found that, while most of the posts are one-off posts meant to be read by indeterminate individuals, a number of posts reference other accounts and such interconnections generates communities of associated posters. These communities may range from a handful of individuals to thousands of people. These communities generally contain discourse that is low stigma but are otherwise very heterogeneous. There are communities that are dominated by very different topics. There are also communities of people that are directly connected to policy specific topics like veterans' mental health and military health issues. In this research, we found that some communities are centered on large government accounts (as in the case of veterans' mental health issues) while others are built around topics of

mental health or shared pop-culture interest. In understanding communities online, each one has to be considered individually.

Policy relevance: The existence of online communities is highly relevant for designing and targeting public health interventions. Individuals who reach out to others who share their concerns are demonstrating a high degree of engagement with a topic. Members of such communities are more likely to be highly engaged and motivated individuals than an individual who makes a single stand-alone post on the topic. Therefore communities focused on mental health conversation contain individuals who are likely to engage with targeted public messages. A readymade network allows for an amplification of public program messages, as a community engaged with a topic is able to provide the sort of repeated dissemination that is required for content to leverage modern social network structure (i.e., 'go viral'). Further, a community structure is more conducive to creating behavior change. Changing beliefs or established behavior patterns requires repeated exposure and is made more effective when the reinforcement comes from multiple points in an individual's social network[198]. Additionally, the ability to detect communities allows policy makers to test for the existence of an audience when designing interventions. Finally, the ability to study individual communities allows for tailored policy interventions. In this work, the most stigmatizing community in the sample was made up of fans of the band 'One Direction' because of their misuse of mental health language. This suggests that partnering with pop personalities like members of 'One Direction' could be an effective way to increase compassion and awareness around mental health conditions. Other communities may have other structural elements and understanding them individually can maximize the impact of public health interventions.

Public messaging campaigns are performing well on Twitter but need greater volumes of content

Findings: The DoD public messaging campaigns, whose Twitter activity was used as a case study in this research, appear to be connecting with active users in the space and are generating engagement, however the volume of Tweets coming from the campaigns is low. It was found that among the four campaigns studied, only one actively managed a Twitter presence, posting at least 1 message per day. However, it was also evident that there was a good deal of engagement with the content that was posted, indicating an appetite for this information. Additionally, the campaign that did post frequently seemed to get greater engagement on more posts as time went on, indicating either a growing level of skill in leveraging the Twitter platform or greater trust from its audience. Finally, a review of published evaluation of social media public health related messaging campaigns illustrated the lack of maturity in the space. There are few established norms or techniques in how to conduct a public messaging campaign in social media.

Policy relevance: This dissertation finds that Twitter (and likely other platforms) is underutilized by entities wishing to enable positive social change or improve public health despite an apparent public appetite for the content. Campaigns that seek to influence attitudes and behaviors are not taking advantage of ready-made communities and audiences on communication platforms like

[198] (Centola, 2010)

Twitter. A policy maker would do well to explore the ways in which a policy intervention could maximize the reach of social networks like Twitter. Additionally, there is little in the way of the gold standard or playbook for using social media. Taking advantage of the opportunities offered by platforms like Twitter will require additional experimentation and learning.

Social media data correlates with traditional research data but may lend itself to different questions

Finding: Comparing the findings of this dissertation with literature that asks similar questions using traditional survey and experimental data suggests consensus on several dimensions between the two data sources. The online conversation is focused on various mental health topics in proportion to the impact that those conditions have on population health. The degree to which various topics are stigmatized on Twitter is mostly consistent with the findings in the literature. The trend towards declining stigma is supported both by survey and Twitter analysis. Self-disclosure is often studied as a powerful therapeutic tool for mental health issues and self-focused content is the most common type of content on Twitter. Additionally, the traditional literature contains work on the role and impact of community and support around mental health illness, and this dissertation demonstrates the presence of such groups.

However, social media data often suggests different questions from the ones that are explored in the literature using traditional sources. Most of the traditional literature focuses on the presence or absence of stigma in general, and the specific kinds of stigma. With social media data, it is much more natural to ask what is stigmatized and how, since the presence of stigma is easy to demonstrate. Similarly, traditional literature focuses on the presence or absence of support communities for individuals, while social media naturally lends itself to asking how the various communities compare. In general, traditional literature seeks to demonstrate the existence or absence of a thing, while social media with its large volumes of data seeks to differentiate qualities.

Policy relevance: The fact that Twitter analysis closely tracks the findings of traditional research methods suggests that social media can be used to understand policy questions. This work shows that the conversation that happens on Twitter in regard to mental health is reflective of how that conversation is understood in the offline world. This means that additional policy questions like gauging constituent interest in various mental health topics, disease surveillance, help seeking behavior, and many other questions related to mental health policy can be explored using social media. While this work only demonstrates the value of Twitter for mental health policy questions, it suggests that social media may be useful for other topics as well. Exploring whether social media can be useful for policy questions related to security, justice, general health, education, childhood development and many others will require an analysis similar to the one conducted in this dissertation. But the work already done indicates that exploring the possibilities of social media for such questions is a worthwhile use of resources.

Methods for analyzing text and social media are available, robust, and reproducible

Finding: This dissertation demonstrates the process to classify and analyze large volumes of unstructured text. This method, which includes a keyword search strategy, robust qualitative coding, and machine learning, is effective and is seen in other high quality research[199]. In developing this work, I found that it is possible to create a rigorous approach to generating a keyword search strategy (by iteratively drawing samples based on keywords, identifying frequent terms, and expanding the search strategy using those). Qualitative coding is an established method in research that directly translates to this work and requires no additional innovation. The classification side is already robust and can be done with simple and powerful tools, or with complex, cutting edge, and rapidly evolving tools. Finally, the resources needed to scale such analyses are readily available, and cloud computing allows for cheap, on demand resources to tackle very large problems. The workflow for collecting and interpreting large volumes of human interaction through social media is viable. These data are meaningful, plentiful, and practicable.

Policy relevance: The use of social media for analysis allows for a new approach to answering policy questions. Instead of single time point snapshots of a survey, social media is always happening and always being recorded. This means that policy questions can be considered longitudinally and continuously. The relative ease with which social media evaluation can be repeated suggests that updates of analyses can be performed on demand. The fact that social media conversation is occurring outside of research settings means an elimination of many of the sources of bias that arise from respondent and researcher[200]. By identifying specific communities that are focused on specific issues, policy makers can more easily engage with stakeholders, and design interventions with greater precision. This can be useful for both individuals affecting national scale policy as well as those seeking targeted local intervention. For large policy questions, social media allows for a broad reach and rapid collection of data. For targeted local intervention, it allows for cheap, repeatable analysis. The variation in available computational, analytic, and machine learning methods means that organizations working in the public good have the ability to implement cheap and easy computational tools for clear questions or invest resources into developing sophisticated tools to better serve their populations. The efficacy of these analytic methods suggests that there is space for more innovation in policy research.

Limitations

Insufficient time and resource for methodological refinement

The methods used in this dissertation were not iterated and refined to furthest extent possible. The search strategy that was developed used an iterative approach of sampling the data and then using the sample to further improve the search strategy. This was done twice but there could have been value in going further in the process. Additionally, given greater resources, this project could have benefited from a larger data set. Here, the limitation was cost in addition to time.

[199] (Emery et al., 2014)

[200] The use of Twitter data minimizes seven of the nine human sources of research biases enumerated in Sarniak (2015)

A greater volume of Tweets could have allowed for more detailed analysis of less frequently occurring topics like military and veteran health issues. Also, with greater time and resources, the performance of this project would have likely improved with a greater training data set tagged by human coders. Spending more time with the coders, iterating the coding schema, tagging more Tweets all would have likely increased the accuracy of the models. There are opportunities to further the approach used for processing of text. I based my work on considering individual words as predictive variables; it is possible to use groups of words as predictors, potentially increasing performance. It is also possible to use specialized software to detect idiomatic content of text further enhancing the robustness of the analysis. Finally, modeling is something that can absorb all the time available, and there are many modeling approaches that could have been tested in the hopes of increasing accuracy, developing working models of additional topics, and having greater insight into hard to define categories like stigma.

Omitted some topics of interest

After completing the project, it became apparent that certain topics of interest were not explored sufficiently because of data limitations, and would have benefited from greater emphasis in the design phase of the project. A comparison of the results of this analysis to the traditional literature revealed a strong focus on schizophrenia and addiction in academic research. These topics were included in the social media analysis but proved either rare in conversation (schizophrenia) or challenging to delineate from colloquial usage (addiction). While these topics were not central to the main research questions, after completing this work it appears that investing additional resources into better capturing these topics would have improved the overall product. There were also other features of interest and there simply was not enough data to analyze them (mental health Tweets about other people, information sharing, ADHD, and several others). With knowledge of those challenges, and additional time and resources, it may have been beneficial to design a data gathering strategy particularly focused on ensuring sufficient sample sizes for these features.

Insights not differentiated by demographics

This analysis does not attempt to randomly sample Twitter or the general population. Part of the initial motivation for this research was to understand how well Twitter reflects off-line behavior and how well it represents the general population. However, given the broader questions that had to be asked in this research, i.e. is mental health discussed online? how do we capture it?, there was not enough space to derive a precise understanding of demographics and representation. This analysis did not pursue the imputation methods necessary to analyze who the people discussing mental health are, or if they represent Twitter in general or the population in general. The results of this analysis do correlate with the results found in the studies using representative national surveys, but the precise demographics of our population are unknown. Getting a clear sense of demographics could be an interesting separate project for future work.

Some Relevant Data Is Not Accessible

This project used 15 million Tweets posted across 100 days, over 100 gigabytes of data. However, performing social media analyses in a research setting is limited by the proprietary nature of the data. Twitter is a public platform where users post nearly all content with the intent

for it to be publically available. As such, it can be gathered and analyzed. However, there are relevant pieces of data that Twitter prevents outside entities from gathering. There is, for example, no systematic way to gain follower lists. This information would have allowed for a more a robust network analysis by allowing for the complete mapping of people's social networks. Instead, this research was only able to use instances of direct communication between people, and ignored all other mutual ties. Additionally, Twitter is neither the only nor the biggest social media platform. However, the data sets of most companies like Facebook, Instagram, Pinterest, and LinkedIn (which are all larger than Twitter) are not publically or commercially available. This means that there is a large volume of diverse activity which may touch upon mental health issues that are not accessible to researchers. . Additionally, there is activity on platforms like Tumblr, Reddit, and other personal blogs which may contain relevant mental health discourse that were beyond the scope of this dissertation.

Campaign effects are small and hard to capture

Part of the motivation for this work was to see how public messaging campaigns on Twitter are able to shift the conversation about a topic. However, it was found that the public messaging campaigns studied are small and their effects could not be measured within the broader conversation. This project was limited by not being able to gather data on specific conversations. It would have been preferable to gather all content that was posted in association with the campaigns over their entire existence, identify anyone who engaged with them, identify all network connections of those individuals that posted something related to the campaigns, perhaps identify the connections of the connections, then gather all posts by accounts within this network. Then, having constructed a full mapping of the conversation surrounding the campaigns, it would have been possible to detect if campaign activity has an impact on the individuals that the campaigns have a chance to reach. This approach was not feasible for several reasons. There are no tools to gather follower lists. The cost structure for Twitter data is such that a far reaching search like the one described would require a colossal budget. Finally, the technical challenges of crawling through the data to gather all of the information would have been time prohibitive. As a result, this dissertation finds that the campaigns are too small to affect the overall Twitter conversation about mental health but cannot say if there was a change in a specific portion of the conversation.

Future Work

Richer measurement of mental health discourse

As discussed above, there was not sufficient time and resource to explore all questions that arose from this work. Three large areas of extending this work stand out. First, identifying the demographic characteristics of social media users is a challenging but important direction for social media research. While this study showed that social media produces results comparable to representative surveys, a detailed analysis of who is in the data would still be important. Accurately identify the demographic features of social media users would allow for more specific questions and questions about specific groups (e.g. inner-city youth). Capturing demographic characteristics could be done by developing tools for imputing demographics from metadata, language, and

connections. Or, by narrowly defining a topic that has well understood results for specific populations. For example, a study could be done to look for indicators of exercise behavior among the Twitter population and then compare the results to survey data on the same question. By systematically going through topics in this fashion it could be understood if and where social media suffers from sampling bias.

The data source for this research is Twitter text data. For the purposes of this work, text was simply tagged in a binary fashion – presence or absence of a particular kind of topic. However, there is opportunity to apply linguistic processing tools to the now filtered and categorized text. It may be of interest to see how emotional features vary across various topics. For example, it is possible to see if words associated with fear are more present in addiction Tweets or in depression Tweets. Similarly, different topics could be analyzed for language associated with hope. This linguistic analysis could even extend to identification of syntax and metaphor. Perhaps there is a dominant metaphor (e.g. 'winning the war against …') for specific parts of mental health conversation. The fact that text is already grouped into topics makes linguistic analysis a natural extension.

Finally, there is ample room to continue to explore, tune, and improve modeling approaches. Machine learning with text data is not a mature field and as this dissertation demonstrates each outcome variable is unique and may be classified with a unique approach. This dissertation made some inroads in comparing SVM with NN and basic logistic regression. However, there is a large class of regression tree based models that were not considered. There is also many additional iterations of logistic, SVM, and NN models that could be tested. Given additional resources it could be a worthwhile extension to continue to explore the various trade-offs of different approaches to classifying text data. The framework built in this dissertation would allow for such a comparison to be extended effectively.

Deeper and broader community analysis, richer longitudinal analysis

Social media analysis is a young field with many opportunities. This work demonstrated a methodology to capture and understand policy-relevant discourse. Many projects could be developed based on this work, but three ideas in particular stood out as potentially interesting.

This dissertation successfully demonstrated the existence of diverse communities of Twitter users discussing mental health topics. Given the unique nature of each community there is an opportunity to examine the content posted within each of these communities and better characterize ongoing conversations. There may be interesting questions about the timelines of these communities (how they grow, how they change) that were not explored. Additionally the topics that I used to define these communities may be discussed very differently in various communities and manually examining various communities could generate additional hypotheses into how people engage with the issues of mental health.

Additionally, it was shown in this dissertation that measuring the impact of a small public messaging campaign is not feasible. Most public messaging campaigns are too small to perceptibly

impact the overall discourse. However, as discussed in the 'limitations' section above, it may be possible to capture the impact of a campaign by capturing all the Tweets that are posted by individuals that may be within several degrees of connection of the campaigns, forming the full social network of engaged individuals as opposed to simply a network of conversation participants as I have done. This would allow the measurement of shifts in attitude and topics in the total conversation, identifying how the messages spread across the full map of personal interactions of Twitter users. This would be incredibly costly, but by partnering with Twitter, or simply getting custom access to data it is possible to make detailed analyses of campaign impact. This would require finding a list of people who engage with the campaigns, and then getting the friends of those individuals and perhaps the friends of the friends. Then by looking at the total activity of all people that the campaigns might hope to reach it would be possible to see if and how the campaigns affected the actions of those individuals.

Lastly, it may be of interest to better utilize the time-stamp data inherent in every Tweet. Several times during this research it became apparent that tracking Twitter sentiment over time and relating it to specific events would be a highly interesting exercise. For example, it would be notable to see how the tragic school shooting at Newtown, CT in 2012 affected people's attitudes towards mental health. Additionally developing a method to examine the sentiment on Twitter for a particular snapshot could allow for a stand-alone tool. It would be valuable to develop an approach that could quickly gather data after an event (e.g., natural disaster), use pre-existing models to sort and classify the data, and then provide insight to policy makers about popular interest and sentiment. This could enable better messaging from government officials after a large scale event, as well as suggest areas for possible policy intervention.

Final thoughts

The goal of this dissertation is two-fold: to improve social science tools that use the large amounts of messy data that people create as they use the internet for their own purposes, and to inform policy makers about the relevance and insights of social media regarding public attitudes towards mental health. Developing tools and methods to understand the records of human behavior may allow social science in general and policy analysis specifically to gain unprecedented insights into human behavior, social structure, and the influence of public policy. It may be possible to design smarter, cheaper, and more timely policy interventions. This analysis of social media may also be beneficial as it offers an indirect way of dealing with a sensitive topic such as mental health without requiring face-to-face interviewing. It may also offer a more direct way for policy makers to reach the people who are most invested in mental health issues.

Figure 7.1 Summary of Dissertation Process, Findings, Implications, and Future Directions

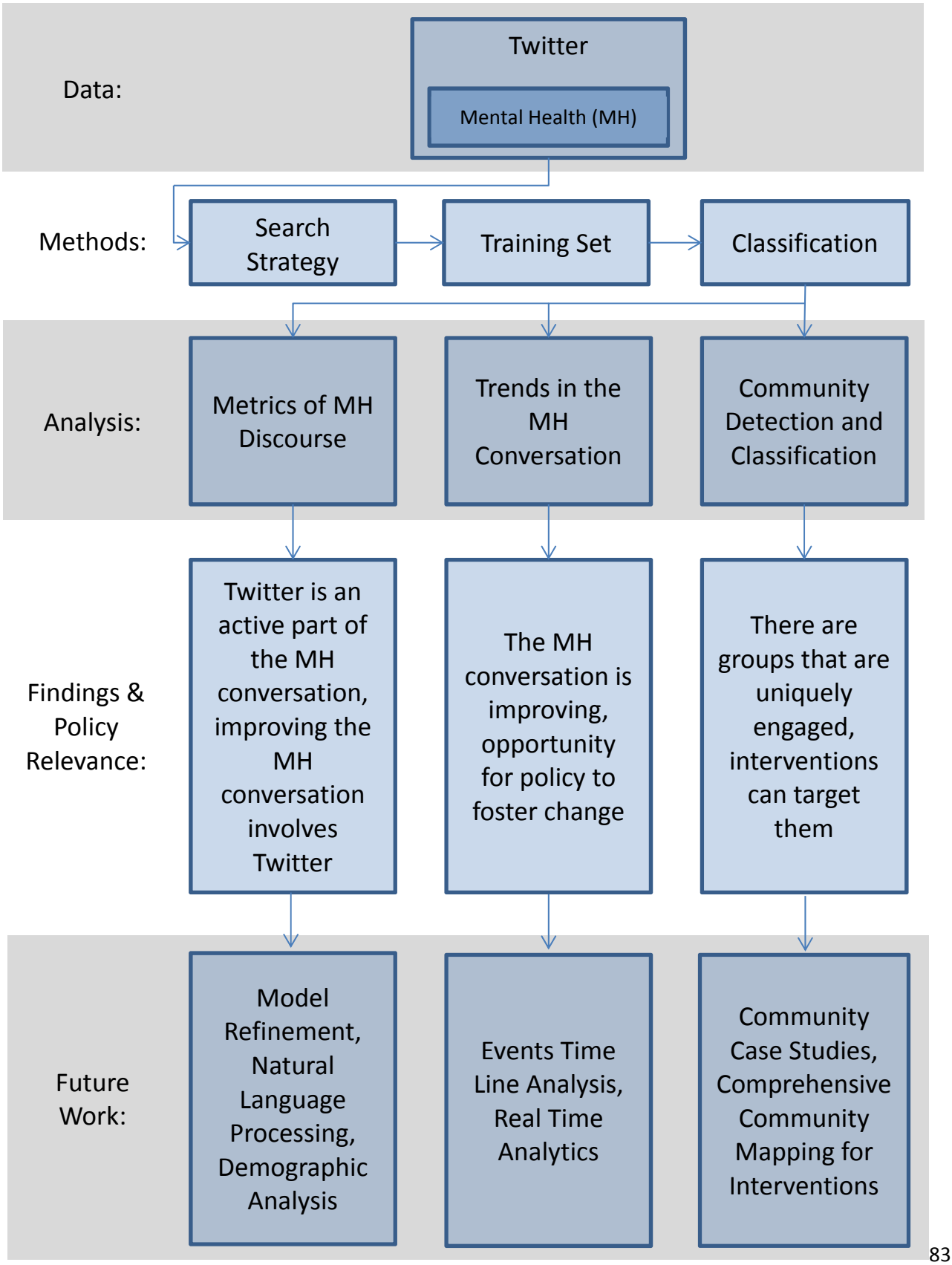


Figure 7.1 contains a summary map of this research. This dissertation took the large unstructured data source of Twitter and searched for the content that is relevant to mental health. This required developing a robust search strategy to identify potentially relevant content. Then, the analysis required selecting a section of the data, and manually determining what relevant Tweets look like and what categories they fall into. Finally, automated methods were constructed to replicate human coded examples. From this methodological approach, I was able to create metrics which describe the Twitter conversation about mental health, analyze time trends in the conversation, and identify communities that are having their own conversation about mental health.

From the analysis I can conclude that there is in fact an active conversation about mental health online. It spans a large number of users and many diverse topics. Combined with findings of prior literature, it appears likely that Twitter (and likely other social media platforms) play an important part in how people relate to issues of mental health. As such. a policy maker that is interested in improving perceptions, attitudes, and actions related to mental health would do well to include social media as a source of insights and as a target for intervention. The results of the Twitter analysis also show the nature of the conversation is improving over time with people misappropriating mental health language less and with stigmatizing content becoming a smaller share of the conversation. This suggests that there is a trend towards progress and it may be beneficial to design policy interventions with the goal of facilitating and nurturing this change. Finally, this analysis discovered that there are groups of individuals which form communities that communicate about mental health topics. Members of these groups converse with each other and the groups have their distinct features. The ability to identify these communities can offer a policy maker with a receptive audience for a policy intervention as well as an active and engaged constituency for analysis and development of interventions.

I conclude with suggestions of future work that can advance the large goals of bettering the ability of social science to understand online behavior data and to further the policy relevance of social media data. There is room to better the methodologies and tools for working with data, for example improving modeling performance or developing a tool that can be applied without modification to various time points of data. There are also opportunities to explore social media data more deeply – examining the characteristics and sentiment of the classified text data, conducting community case studies, and developing a greater understanding of the demographics of the online data sets which do not come with representative sampling. There is ample opportunity to have a large impact on future analyses of internet data.

Bibliography

Acosta, Joie , Amariah Becker, Jennifer L. Cerully, Michael P. Fisher, Laurie Martin, and Raffaele Vardavas, *A Literature Review on Mental Health Stigma Reduction*, unpublished: RAND, 2013.

Acosta, Joie, Amariah Becker, Jennifer L Cerully, Michael P Fisher, Laurie T Martin, Raffaele Vardavas, Mary Ellen Slaughter, and Terry L Schell, *Mental Health Stigma in the Military*: Rand Corporation, 2014.

Ajzen, Icek, "The theory of planned behavior," *Organizational behavior and human decision processes,* Vol. 50, No. 2, 1991, pp. 179-211.

———, "Nature and operation of attitudes," *Annual review of psychology,* Vol. 52, No. 1, 2001, pp. 27-58.

AlSumait, Loulwah, Daniel Barbará, and Carlotta Domeniconi, "On-line lda: Adaptive topic models for mining text streams with applications to topic detection and tracking," *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, 2008, pp. 3-12.

American Psychiatric Association, *DSM 5*: American Psychiatric Association, 2013.

Angermeyer, Matthias C, and Sandra Dietrich, "Public beliefs about and attitudes towards people with mental illness: a review of population studies," *Acta Psychiatrica Scandinavica,* Vol. 113, No. 3, 2006, pp. 163-179.

Angermeyer, Matthias, Herbert Matschinger, and Georg Schomerus, "Attitudes towards psychiatric treatment and people with mental illness: changes over two decades," *The British Journal of Psychiatry,* Vol. 203, No. 2, 2013, pp. 146-151. http://bjp.rcpsych.org/content/bjprcpsych/203/2/146.full.pdf

Aphinyanaphongs, Yin, Bisakha Ray, Alexander Statnikov, and Paul Krebs, "Text classification for automatic detection of alcohol use-related tweets: A feasibility study," *Information Reuse and Integration (IRI), 2014 IEEE 15th International Conference on*, 2014, pp. 93-97.

Archontaki, Despina, Gary J Lewis, and Timothy C Bates, "Genetic influences on psychological well-being: A nationally representative twin study," *Journal of Personality,* Vol. 81, No. 2, 2013, pp. 221-230.

Bakshy, Eytan, Jake M Hofman, Winter A Mason, and Duncan J Watts, "Everyone's an influencer: quantifying influence on twitter," *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011a, pp. 65-74.

———, "Identifying influencers on twitter," *Fourth ACM International Conference on Web Seach and Data Mining (WSDM)*, 2011b.

Barke, Antonia, Seth Nyarko, and Dorothee Klecha, "The stigma of mental illness in Southern Ghana: attitudes of the urban population and patients' views," *Social psychiatry and psychiatric epidemiology,* Vol. 46, No. 11, 2011, pp. 1191-1202.

Barry, Colleen L, Emma E McGinty, Jon S Vernick, and Daniel W Webster, "After Newtown—public opinion on gun policy and mental illness," *New England Journal of Medicine,* Vol. 368, No. 12, 2013, pp. 1077-1081.

Bathje, Geoff, and John Pryor, "The relationships of public and self-stigma to seeking mental health services," *Journal of Mental Health Counseling,* Vol. 33, No. 2, 2011, pp. 161-176.

Beldie, Alina, Johan A Den Boer, Cecilia Brain, Eric Constant, Maria Luisa Figueira, Igor Filipcic, Benoît Gillain, Miro Jakovljevic, Marek Jarema, and Daniela Jelenova, "Fighting stigma of mental illness in midsize European countries," *Social psychiatry and psychiatric epidemiology,* Vol. 47, No. 1, 2012, pp. 1-38.

Benetoli, Arcelio, Timothy F Chen, and Parisa Aslani, "The use of social media in pharmacy practice and education," *Research in Social and Administrative Pharmacy,* Vol. 11, No. 1, 2015, pp. 1-46.

Benevenuto, Fabricio, Gabriel Magno, Tiago Rodrigues, and Virgilio Almeida, "Detecting spammers on twitter," *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)*, 2010, p. 12.

Bishop, Christopher M, "Pattern recognition," *Machine Learning,* Vol. 128, 2006.

Blei, David M, Andrew Y Ng, and Michael I Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research,* Vol. 3, 2003, pp. 993-1022.

Boardman, Felicity, Frances Griffiths, Renata Kokanovic, Maria Potiriadis, Christopher Dowrick, and Jane Gunn, "Resilience as a response to the stigma of depression: A mixed methods analysis," *Journal of affective disorders,* Vol. 135, No. 1, 2011, pp. 267-276.

Bonacich, Phillip, "Factoring and weighting approaches to status scores and clique identification," *Journal of Mathematical Sociology,* Vol. 2, No. 1, 1972, pp. 113-120.

Bonasio, Alice, "SAVING THE SWAN SONG:

Twitter is failing as a public company, but there are other ways to keep this bird aloft," 2016. As of 1/19/2017:

https://qz.com/868522/buying-twitter-to-save-the-social-media-publication-we-should-regard-it-as-a-public-service-not-a-private-company/

Bourguignon, David, Eleonore Seron, Vincent Yzerbyt, and Ginette Herman, "Perceived group and personal discrimination: differential effects on personal self-esteem," *European Journal of Social Psychology,* Vol. 36, No. 5, 2006, pp. 773-789.

Boyd, Danah, *It's Complicated: the social lives of networked teens*: Yale University Press, 2014.

Boyd, Danah, and Kate Crawford, "Six provocations for big data," *A decade in internet time: Symposium on the dynamics of the internet and society*, 2011, pp. 1-17.

Bradburn, Norman M, "The structure of psychological well-being," 1969.

Bravo, Caroline A, and Laurie Hoffman-Goetz, "Tweeting about prostate and testicular cancers: Do Twitter conversations and the 2013 Movember Canada campaign objectives align?," *Journal of Cancer Education,* Vol. 31, No. 2, 2016, pp. 236-243.

Bray, Janet E, Lahn Straney, Bill Barger, and Judith Finn, "Effect of public awareness campaigns on calls to ambulance across Australia," *Stroke,* Vol. 46, No. 5, 2015, pp. 1377-1380.

Bray, Robert M, Michael R Pemberton, Marian E Lane, Laurel L Hourani, Mark J Mattiko, and Lorraine A Babeu, "Substance use and mental health trends among US military active duty personnel: key findings from the 2008 DoD Health Behavior Survey," *Military medicine,* Vol. 175, No. 6, 2010, pp. 390-399.

Centers for Disease Control and Prevention, Substance Abuse and Mental Health Services Administration, National Association of County Behavioral Health & Developmental Disability Directors, National Institute of Mental Health, and The Carter Center Mental Health Program, *Attitudes Toward Mental Illness: Results from the*

Behavioral Risk Factor Surveillance System, Atlanta, GA, Centers for Disease Control and Prevention, 2012.

Centers for Disease Control and Prevention: Program Performance and Evaluation Office, "Mental Health," 2013. As of 12/16/2016:

https://www.cdc.gov/mentalhealth/basics.htm

Centers for Disease Control Prevention, "The health communicator's social media toolkit," *Atlanta, GA*, 2011.

Centola, Damon, "The spread of behavior in an online social network experiment," *Science,* Vol. 329, No. 5996, 2010, pp. 1194-1197.

Chatfield, Akemi, and Uuf Brajawidagda, "Twitter tsunami early warning network: a social network analysis of Twitter information flows," 2012.

Chen, Liangzhe, KSM Tozammel Hossain, Patrick Butler, Naren Ramakrishnan, and B Aditya Prakash, "Syndromic surveillance of Flu on Twitter using weakly supervised temporal topic models," *Data Mining and Knowledge Discovery*, 2015, pp. 1-30.

Cheong, France, and Christopher Cheong, "Social Media Data Mining: A Social Network Analysis Of Tweets During The 2010-2011 Australian Floods," *PACIS,* Vol. 11, 2011, pp. 46-46.

Cole-Lewis, Heather, Arun Varghese, Amy Sanders, Mary Schwarz, Jillian Pugatch, and Erik Augustson, "Assessing Electronic Cigarette-Related Tweets for Sentiment and Content Using Supervised Machine Learning," *Journal of medical Internet research,* Vol. 17, No. 8, 2015, p. e208.

Cook, Judith A, Mary Ellen Copeland, Jessica A Jonikas, Marie M Hamilton, Lisa A Razzano, Dennis D Grey, Carol B Floyd, Walter B Hudson, Rachel T Macfarlane, and Tina M Carter, "Results of a randomized controlled trial of mental illness self-management using Wellness Recovery Action Planning," *Schizophrenia Bulletin*, 2011, p. sbr012.

Cook, Judith A, Pamela Steigman, Sue Pickett, Sita Diehl, Anthony Fox, Patricia Shipley, Rachel MacFarlane, Dennis D Grey, and Jane K Burke-Miller, "Randomized controlled trial of peer-led recovery education using Building Recovery of Individual Dreams and Goals through Education and Support (BRIDGES)," *Schizophrenia research,* Vol. 136, No. 1, 2012, pp. 36-42.

Cook, Wade D, Boaz Golany, Moshe Kress, Michal Penn, and Tal Raviv, "Optimal allocation of proposals to reviewers to facilitate effective ranking," *Management Science,* Vol. 51, No. 4, 2005, pp. 655-661.

Corley, Courtney D, Diane J Cook, Armin R Mikler, and Karan P Singh, "Text and structural data mining of influenza mentions in web and social media," *International journal of environmental research and public health,* Vol. 7, No. 2, 2010, pp. 596-615.
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2872292/pdf/ijerph-07-00596.pdf

Corrigan, Patrick, "How stigma interferes with mental health care," *American psychologist,* Vol. 59, No. 7, 2004, p. 614.

Corrigan, Patrick W, Amy Kerr, and Lissa Knudsen, "The stigma of mental illness: explanatory models and methods for change," *Applied and Preventive Psychology,* Vol. 11, No. 3, 2005, pp. 179-190.

Corrigan, Patrick W, and John R O'Shaughnessy, "Changing mental illness stigma as it exists in the real world," *Australian Psychologist,* Vol. 42, No. 2, 2007, pp. 90-97.

Crabtree, Jason W, S Alexander Haslam, Tom Postmes, and Catherine Haslam, "Mental Health Support Groups, Stigma, and Self-Esteem: Positive and Negative Implications of Group Identification," *Journal of Social Issues,* Vol. 66, No. 3, 2010, pp. 553-569.

Crisp, Arthur, Michael Gelder, Eileen Goddard, and Howard Meltzer, "Stigmatization of people with mental illnesses: a follow-up study within the Changing Minds campaign of the Royal College of Psychiatrists," *World psychiatry,* Vol. 4, No. 2, 2005, p. 106.

Csardi, Gabor, and Tamas Nepusz, "The igraph software package for complex network research," *InterJournal, Complex Systems,* Vol. 1695, No. 5, 2006, pp. 1-9.

Davey, Graham CL, "Mental health and stigma," *Psychology Today*, 2013.

De Choudhury, Munmun, Scott Counts, Eric J Horvitz, and Aaron Hoff, "Characterizing and predicting postpartum depression from shared facebook data," *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 2014, pp. 626-638.

De Choudhury, Munmun, and Sushovan De, "Mental health discourse on reddit: Self-disclosure, social support, and anonymity," *Proc. ICWSM, AAAI*, 2014.

De Choudhury, Munmun, Sanket S Sharma, and Emre Kıcıman, "Characterizing Dietary Choices, Nutrition, and Language in Food Deserts via Social Media," 2016.

Diener, Ed, Randy J Larsen, Steven Levine, and Robert A Emmons, "Intensity and frequency: Dimensions underlying positive and negative affect," *Journal of personality and social psychology,* Vol. 48, No. 5, 1985, p. 1253.

Diener, Ed, Eunkook M Suh, Richard E Lucas, and Heidi L Smith, "Subjective well-being: Three decades of progress," *Psychological bulletin,* Vol. 125, No. 2, 1999, pp. 276-302.

Douglas, Nick "Twitter blows up at SXSW Conference," 2007. As of 12/9/2015:

http://gawker.com/243634/twitter-blows-up-at-sxsw-conference

Duggan, Maeve, *Mobile Messaging and Social Media – 2015*: Pew Research Center, 2015. http://www.pewinternet.org/2015/08/19/mobile-messaging-and-social-media-2015/

Duke, Jennifer C, Heather Hansen, Annice E Kim, Laurel Curry, and Jane Allen, "The use of social media by state tobacco control programs to promote smoking cessation: a cross-sectional study," *Journal of medical Internet research,* Vol. 16, No. 7, 2014, p. e169.

Ediger, David, Kui Jiang, Jason Riedy, David Bader, Courtney Corley, Rob Farber, and William N Reynolds, "Massive social network analysis: Mining twitter for social good," *Parallel Processing (ICPP), 2010 39th International Conference on*, 2010, pp. 583-593.

Eichstaedt, Johannes C, Hansen Andrew Schwartz, Margaret L Kern, Gregory Park, Darwin R Labarthe, Raina M Merchant, Sneha Jha, Megha Agrawal, Lukasz A Dziurzynski, and Maarten Sap, "Psychological language on twitter predicts county-level heart disease mortality," *Psychological science,* Vol. 26, No. 2, 2015, pp. 159-169.

Ellis, Darren, and John Cromby, "Emotional inhibition: A discourse analysis of disclosure," *Psychology & health,* Vol. 27, No. 5, 2011, pp. 515-532.

Emery, Sherry L, Glen Szczypka, Eulalia P Abril, Yoonsang Kim, and Lisa Vera, "Are you scared yet? Evaluating fear appeal messages in tweets about the tips campaign," *Journal of Communication,* Vol. 64, No. 2, 2014, pp. 278-295.

Epstein, Seymour, "Aggregation and beyond: Some basic issues on the prediction of behavior," *Journal of Personality,* Vol. 51, No. 3, 1983, pp. 360-392.

Evans-Lacko, Sara, Elaine Brohan, Ramin Mojtabai, and Graham Thornicroft, "Association between public views of mental illness and self-stigma among individuals with mental illness in 14 European countries," *Psychological medicine,* Vol. 42, No. 8, 2012, p. 1741. http://journals.cambridge.org/download.php?file=%2FPSM%2FPSM42_08%2FS0033291711002558a.pdf&code=b1acc1e40d8839b8559d97c60a9e146d

Farley, BWAC, and W Clark, "Simulation of self-organizing systems by digital computer," *Transactions of the IRE Professional Group on Information Theory,* Vol. 4, No. 4, 1954, pp. 76-84.

Feinerer, Ingo, "Introduction to the tm Package Text Mining in R," 2015.

Frattaroli, Joanne, "Experimental disclosure and its moderators: a meta-analysis," *Psychological bulletin,* Vol. 132, No. 6, 2006, p. 823.

Freelon, Deen, "On the interpretation of digital trace data in communication and social computing research," *Journal of Broadcasting & Electronic Media,* Vol. 58, No. 1, 2014, pp. 59-75.

Freeman, Linton C, "Centrality in social networks conceptual clarification," *Social Networks,* Vol. 1, No. 3, 1979, pp. 215-239.

Fritsch, Stefan, Frauke Guenther, and Maintainer Frauke Guenther, "Package 'neuralnet'," *Training of Neural Network,* Vol. 1, 2012.

Frommer, Dan, and Kamelia Angelova, "CHART OF THE DAY: Twitter Raises Cash Pile As Traffic Growth Slows," September 24, 2009, 2009. As of November 30, 2015:

http://www.businessinsider.com/chart-of-the-day-twitter-worldwide-uniques-2009-9

Galuba, Wojciech, Karl Aberer, Dipanjan Chakraborty, Zoran Despotovic, and Wolfgang Kellerer, "Outtweeting the twitterers-predicting information cascades in microblogs," *Proceedings of the 3rd conference on Online social networks*, 2010, p. 3âAS3.

Gamer, Matthias, Jim Lemon, Ian Fellows, and Puspendra Singh, "irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84," *Internet resource: http://cran.r-project. org/package= irr](Verified April 10, 2013)*, 2012.

Goffman, Erving, *Stigma: Notes on the management of spoiled identity*: Prentice-Hall, 1963.

Goldstrom, Ingrid D, Jean Campbell, Joseph A Rogers, David B Lambert, Beatrice Blacklow, Marilyn J Henderson, and Ronald W Manderscheid, "National estimates for mental health mutual support groups, self-help organizations, and consumer-operated services," *Administration and Policy in Mental Health and Mental Health Services Research,* Vol. 33, No. 1, 2006, pp. 92-103.

González-Bailón, Sandra, Ning Wang, Alejandro Rivero, Javier Borge-Holthoefer, and Yamir Moreno, "Assessing the bias in samples of large online networks," *Social Networks,* Vol. 38, 7//, 2014, pp. 16-27. http://www.sciencedirect.com/science/article/pii/S0378873314000057

Gough, Aisling, "Using Social Media for Public Health.. Brb [be Right Back]," *2016 National Conference on Health Communication, Marketing, and Media (August 23-25)*, 2016.

Graves, Alex, and Jürgen Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," *Advances in neural information processing systems*, 2009, pp. 545-552.

Guimera, Roger, Stefano Mossa, Adrian Turtschi, and LA Nunes Amaral, "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles," *Proceedings of the National Academy of Sciences,* Vol. 102, No. 22, 2005, pp. 7794-7799.

Hallgren, Kevin A., "Computing inter-rater reliability for observational data: an overview and tutorial," *Tutorials in quantitative methods for physchology,* Vol. 8.1, No. 23, 2012.

Hawkins, Douglas M, "The problem of overfitting," *Journal of chemical information and computer sciences,* Vol. 44, No. 1, 2004, pp. 1-12.

Hecht-Nielsen, Robert, "Theory of the backpropagation neural network," *Neural Networks, 1989. IJCNN., International Joint Conference on*, 1989, pp. 593-605.

Hoge, Charles W, Carl A Castro, Stephen C Messer, Dennis McGurk, Dave I Cotting, and Robert L Koffman, "Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care," *New England Journal of Medicine,* Vol. 351, No. 1, 2004, pp. 13-22.

Hong, Liangjie, and Brian D Davison, "Empirical study of topic modeling in twitter," *Proceedings of the First Workshop on Social Media Analytics*, 2010, pp. 80-88.

Hornik, Kurt, "Approximation capabilities of multilayer feedforward networks," *Neural networks,* Vol. 4, No. 2, 1991, pp. 251-257.

Huang, Jin, and Charles X Ling, "Using AUC and accuracy in evaluating learning algorithms," *IEEE Transactions on knowledge and Data Engineering,* Vol. 17, No. 3, 2005, pp. 299-310.

Hui, Cindy, Yulia Tyshchuk, William A Wallace, Malik Magdon-Ismail, and Mark Goldberg, "Information cascades in social media in response to a crisis: a preliminary model and a case study," *Proceedings of the 21st international conference companion on World Wide Web*, 2012, pp. 653-656.

Hyman, Iris, *Self-disclosure and its impact on individuals who receive mental health services*: US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, 2008.

Java, Akshay, Xiaodan Song, Tim Finin, and Belle Tseng, "Why we twitter: understanding microblogging usage and communities," *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, 2007, pp. 56-65.

Jeong, Hawoong, Sean P Mason, A-L Barabási, and Zoltan N Oltvai, "Lethality and centrality in protein networks," *Nature,* Vol. 411, No. 6833, 2001, pp. 41-42.

Ji, Xiang, Soon Ae Chun, Zhi Wei, and James Geller, "Twitter sentiment classification for measuring public health concerns," *Social Network Analysis and Mining,* Vol. 5, No. 1, 2015, pp. 1-25.

Joinson, Adam N, and Carina B Paine, "Self-disclosure, privacy and the Internet," *The Oxford handbook of Internet psychology*, 2007, pp. 237-252.

Jonnalagadda, Siddhartha, Ryan Peeler, and Philip Topham, "Discovering opinion leaders for medical topics using news articles," *J. Biomedical Semantics,* Vol. 3, 2012, p. 2.

Kassam, Aliya, Nick Glozier, Morven Leese, Joanne Loughran, and Graham Thornicroft, "A controlled trial of mental illness related stigma training for medical students," *BMC medical education,* Vol. 11, No. 1, 2011, p. 51.

Kessler, Ronald C, Wai Tat Chiu, Olga Demler, and Ellen E Walters, "Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication," *Archives of general psychiatry,* Vol. 62, No. 6, 2005, pp. 617-627.

Keyes, Corey Lee M, "Social well-being," *Social psychology quarterly*, 1998, pp. 121-140.

Kim, Elsa, Sam Gilbert, Michael J Edwards, and Erhardt Graeff, "Detecting sadness in 140 characters: Sentiment analysis and mourning Michael Jackson on Twitter," *Web Ecology,* Vol. 3, 2009, pp. 1-15.

Kohavi, Ron, and Foster Provost, "Glossary of terms," *Machine Learning,* Vol. 30, No. 2-3, 1998, pp. 271-274.

Koschade, Stuart, "A social network analysis of Jemaah Islamiyah: The applications to counterterrorism and intelligence," *Studies in Conflict & Terrorism,* Vol. 29, No. 6, 2006, pp. 559-575.

Kushin, Matthew James, and Masahiro Yamamoto, "Did social media really matter? College students' use of online media and political decision making in the 2008 election," *Mass Communication and Society,* Vol. 13, No. 5, 2010, pp. 608-630.

Labov, William, *Sociolinguistic patterns*: University of Pennsylvania Press, 1972.

Lachlan, Kenneth A, Patric R Spence, Autumn Edwards, Katie M Reno, and Chad Edwards, "If you are quick enough, I will think about it: Information speed and trust in public health organizations," *Computers in Human Behavior,* Vol. 33, 2014, pp. 377-380.

Lakeman, Richard, P McGowan, L MacGabhann, M Parkinson, M Redmond, Ingrid Sibitz, C Stevenson, and J Walsh, "A qualitative study exploring experiences of discrimination associated with mental-health problems in Ireland," *Epidemiology and psychiatric sciences,* Vol. 21, No. 03, 2012, pp. 271-279.

Lee, Kathy, Diana Palsetia, Ramanathan Narayanan, Md Mostofa Ali Patwary, Ankit Agrawal, and Alok Choudhary, "Twitter trending topic classification," *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, 2011, pp. 251-258.

Lerman, Kristina, and Rumi Ghosh, "Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks," *ICWSM,* Vol. 10, 2010, pp. 90-97.

Li, Veronica, "The Rise, Critique and Persistence of the DALY in Global Health," *The Journal of Global Health*, 2014.

Liljeros, Fredrik, Christofer R Edling, Luis A Nunes Amaral, H Eugene Stanley, and Yvonne Åberg, "The web of human sexual contacts," *Nature,* Vol. 411, No. 6840, 2001, pp. 907-908.

Lincoln, Tania M, Elisabeth Arens, Cornelia Berger, and Winfried Rief, "Can antistigma campaigns be improved? A test of the impact of biogenetic vs psychosocial causal explanations on implicit and explicit attitudes to schizophrenia," *Schizophrenia Bulletin,* Vol. 34, No. 5, 2008, pp. 984-994.

Link, Bruce G, Francis T Cullen, Elmer Struening, Patrick E Shrout, and Bruce P Dohrenwend, "A modified labeling theory approach to mental disorders: An empirical assessment," *American Sociological Review*, 1989, pp. 400-423.

Liu, Bing, and Lei Zhang, "A survey of opinion mining and sentiment analysis," *Mining text data*: Springer, 2012, pp. 415-463.

Livingston, James D, and Jennifer E Boyd, "Correlates and consequences of internalized stigma for people living with mental illness: A systematic review and meta-analysis," *Social science & medicine,* Vol. 71, No. 12, 2010, pp. 2150-2161.

Livingston, James D, Michelle Cianfrone, Kimberley Korf-Uzan, and Connie Coniglio, "Another time point, a different story: one year effects of a social media intervention on the attitudes of young people towards mental health issues," *Social psychiatry and psychiatric epidemiology,* Vol. 49, No. 6, 2014, pp. 985-990.

Lloyd-Evans, Brynmor, Evan Mayo-Wilson, Bronwyn Harrison, Hannah Istead, Ellie Brown, Stephen Pilling, Sonia Johnson, and Tim Kendall, "A systematic review and meta-analysis of randomised controlled trials of peer support for people with severe mental illness," *BMC psychiatry,* Vol. 14, No. 1, 2014, p. 1.

Lobo, Jorge M, Alberto Jiménez-Valverde, and Raimundo Real, "AUC: a misleading measure of the performance of predictive distribution models," *Global ecology and Biogeography,* Vol. 17, No. 2, 2008, pp. 145-151.

Mainka, Agnes, Steve Hartmann, Wolgang G Stock, and Isabella Peters, "Government and social media: a case study of 31 informational world cities," *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, 2014, pp. 1715-1724.

Martin, Jack K, Bernice A Pescosolido, and Steven A Tuch, "Of fear and loathing: the role of'disturbing behavior,'labels, and causal attributions in shaping public attitudes toward people with mental illness," *Journal of Health and Social Behavior*, 2000, pp. 208-223.

McFarland, Lynn A., and Robert E. Ployhart, "Social media: A contextual framework to guide research and practice," *Journal of Applied Psychology,* Vol. 100, No. 6, 2015, pp. 1653-1677. http://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=2015-24693-001&site=ehost-live

lynn.mcfarland@moore.sc.edu

http://psycnet.apa.org/journals/apl/100/6/1653/

Mejova, Yelena, Hamed Haddadi, Anastasios Noulas, and Ingmar Weber, "# FoodPorn: Obesity Patterns in Culinary Interactions," *Proceedings of the 5th International Conference on Digital Health 2015*, 2015, pp. 51-58.

Mellon, Lisa, Anne Hickey, Frank Doyle, Eamon Dolan, and David Williams, "Can a media campaign change health service use in a population with stroke symptoms? Examination of the first Irish stroke awareness campaign," *Emergency Medicine Journal,* Vol. 31, No. 7, 2014, pp. 536-540.

Mill, John Stuart, *Utilitarianism*: Longmans, Green and Company, 1901.

Milliken, Charles S, Jennifer L Auchterlonie, and Charles W Hoge, "Longitudinal assessment of mental health problems among active and reserve component soldiers returning from the Iraq war," *Jama,* Vol. 298, No. 18, 2007, pp. 2141-2148.

Mislove, Alan, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J Niels Rosenquist, "Understanding the Demographics of Twitter Users," *ICWSM,* Vol. 11, 2011, p. 5th.

Mohammad, Saif M, Svetlana Kiritchenko, and Xiaodan Zhu, "NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets," *Second Joint Conference on Lexical and Computational Semantics (* SEM)*, 2013, pp. 321-327.

Mojtabai, Ramin, "Americans' attitudes toward mental health treatment seeking: 1990–2003," *Psychiatric services,* Vol. 58, No. 5, 2007, pp. 642-651.

Moyer, Christopher, "How Google's AlphaGo beat a Go world champion," *The Atlantic, March,* Vol. 28, 2016.

Murray, Christopher JL, Jerry Abraham, Mohammed K Ali, Miriam Alvarado, Charles Atkinson, Larry M Baddour, David H Bartels, Emelia J Benjamin, Kavi Bhalla, and Gretchen Birbeck, "The state of US health, 1990-2010: burden of diseases, injuries, and risk factors," *Jama,* Vol. 310, No. 6, 2013, pp. 591-606.

Murray, CJL, and AD Lopez, *he Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries and Risk Factors in 1990 and Projected to 2020*, Geneva, Switzerland: World Health Organization, 1996.

Neiger, Brad L, Rosemary Thackeray, Scott H Burton, Callie R Thackeray, and Jennifer H Reese, "Use of twitter among local health departments: an analysis of information sharing, engagement, and action," *Journal of medical Internet research,* Vol. 15, No. 8, 2013, p. e177.

Neiger, Brad L, Rosemary Thackeray, Sarah A Van Wagenen, Carl L Hanson, Joshua H West, Michael D Barnes, and Michael C Fagen, "Use of social media in health promotion purposes, key performance indicators, and evaluation metrics," *Health promotion practice,* Vol. 13, No. 2, 2012, pp. 159-164. http://hpp.sagepub.com/content/13/2/159.full.pdf

Obama, B, "Improving access to mental health services for veterans, service members, and military families. Executive Order issued August 31, 2012 by President Barack Obama," 2014.

Oreskovic, Alexei, "Here's another area where Twitter appears to have stalled: tweets per day," *Business Insider*, June 15, 2015, 2015. http://www.businessinsider.com/twitter-tweets-per-day-appears-to-have-stalled-2015-6

Paek, Hye-Jin, Thomas Hove, Yumi Jung, and Richard T Cole, "Engagement across three social media platforms: An exploratory study of a cause-related PR campaign," *Public Relations Review,* Vol. 39, No. 5, 2013, pp. 526-533.

Panahi, Sirous, Jason Watson, and Helen Partridge, "Social media and physicians: Exploring the benefits and challenges," *Health informatics journal*, 2014, p. 1460458214540907.

Pang, Bo, and Lillian Lee, "Opinion mining and sentiment analysis," *Foundations and trends in information retrieval,* Vol. 2, No. 1-2, 2008, pp. 1-135.

Park, Minsu, David W McDonald, and Meeyoung Cha, "Perception Differences between the Depressed and Non-Depressed Users in Twitter," 2013.

Paul, Michael J, and Mark Dredze, "You are what you Tweet: Analyzing Twitter for public health," *ICWSM*, 2011, pp. 265-272.

———, "A model for mining public health topics from Twitter," *Health,* Vol. 11, 2012, pp. 16-16.

Pavalanathan, Umashanthi, and Munmun De Choudhury, "Identity Management and Mental Health Discourse in Social Media," *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 315-321.

Pawar, Kishori K, Pukhraj P Shrishrimal, and RR Deshmukh, "Twitter Sentiment Analysis: A Review."

Pennebaker, James W, and Sandra K Beall, "Confronting a traumatic event: toward an understanding of inhibition and disease," *Journal of abnormal psychology,* Vol. 95, No. 3, 1986, p. 274.

Pescosolido, Bernice A, "The Public Stigma of Mental Illness What Do We Think; What Do We Know; What Can We Prove?," *Journal of Health and Social Behavior,* Vol. 54, No. 1, 2013, pp. 1-21. http://hsb.sagepub.com/content/54/1/1.full.pdf

Petty, Richard, and John T Cacioppo, *Communication and persuasion: Central and peripheral routes to attitude change*: Springer Science & Business Media, 2012.

Pons, Pascal, and Matthieu Latapy, "Computing communities in large networks using random walks," *International Symposium on Computer and Information Sciences*, 2005, pp. 284-293.

Price, P, R Kobau, J Buelow, J Austin, and K Lowenberg, "Improving understanding, promoting social inclusion, and fostering empowerment related to epilepsy: Epilepsy Foundation public awareness campaigns—2001 through 2013," *Epilepsy & Behavior,* Vol. 44, 2015, pp. 239-244.

Proudfoot, Judith, Gordon Parker, Vijaya Manicavasagar, Dusan Hadzi-Pavlovic, Alexis Whitton, Jennifer Nicholas, Meg Smith, and Rowan Burckhardt, "Effects of adjunctive peer support on perceptions of illness control and understanding in an online psychoeducation program for bipolar disorder: a randomised controlled trial," *Journal of affective disorders,* Vol. 142, No. 1, 2012, pp. 98-105.

Quinn, Diane M, and Stephenie R Chaudoir, "Living with a concealable stigmatized identity: the impact of anticipated stigma, centrality, salience, and cultural stigma on psychological distress and health," *Journal of personality and social psychology,* Vol. 97, No. 4, 2009, p. 634. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4511710/pdf/nihms706276.pdf

Ramanadhan, Shoba, Samuel R Mendez, Megan Rao, and Kasisomayajula Viswanath, "Social media use by community-based organizations conducting health promotion: a content analysis," *BMC Public Health,* Vol. 13, No. 1, 2013, p. 1.

Rao, H, H Mahadevappa, P Pillay, M Sessay, A Abraham, and J Luty, "A study of stigmatized attitudes towards people with mental health problems among health professionals," *Journal of psychiatric and mental health nursing,* Vol. 16, No. 3, 2009, pp. 279-284. http://onlinelibrary.wiley.com/store/10.1111/j.1365-2850.2008.01369.x/asset/j.1365-2850.2008.01369.x.pdf?v=1&t=irdynry6&s=600ff7c094af29dc7f61e6fc461f7884077541dc

Reavley, Nicola J, and Anthony F Jorm, "Stigmatizing attitudes towards people with mental disorders: findings from an Australian National Survey of Mental Health Literacy and Stigma," *Australian and New Zealand Journal of Psychiatry,* Vol. 45, No. 12, 2011, pp. 1086-1093.

Repper, Julie, and Tim Carter, "A review of the literature on peer support in mental health services," *Journal of Mental Health,* Vol. 20, No. 4, 2011, pp. 392-411.

Ripley, Brian, and W Venables, "nnet: Feed-forward neural networks and multinomial log-linear models," *R package version,* Vol. 7, No. 5, 2011.

Romero, Daniel M, Brendan Meeder, and Jon Kleinberg, "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter," *Proceedings of the 20th international conference on World wide web*, 2011, pp. 695-704.

Roy, Durga, Jacob Taylor, Christine C Cheston, Tabor E Flickinger, and Margaret S Chisolm, "Social media: portrait of an emerging tool in medical education," *Academic Psychiatry*, 2014, pp. 1-5.

Ruiz, Susan L, and Lee Stadtlander, "Social Media as Support for Partners of Veterans With Posttraumatic Stress Disorder," *Journal of Social, Behavioral, and Health Sciences,* Vol. 9, No. 1, 2015, p. 1.

Ruths, Derek, and Jürgen Pfeffer, "Social media for large studies of behavior," *Science,* Vol. 346, No. 6213, 2014, pp. 1063-1064. http://www.sciencemag.org/content/346/6213/1063.full.pdf

Ryff, Carol D, "Happiness is everything, or is it? Explorations on the meaning of psychological well-being," *Journal of personality and social psychology,* Vol. 57, No. 6, 1989, p. 1069.

Ryff, Carol D, and Corey Lee M Keyes, "The structure of psychological well-being revisited," *Journal of personality and social psychology,* Vol. 69, No. 4, 1995, p. 719.

Salathé, Marcel, and Shashank Khandelwal, "Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control," *PLoS Comput Biol,* Vol. 7, No. 10, 2011.

Sarniak, Rebecca, "9 types of research bias and how to avoid them," *Quirk's Marketing Research Review*, 2015. As of 12/30/2016:

http://www.quirks.com/articles/9-types-of-research-bias-and-how-to-avoid-them

Schnittker, Jason, "An uncertain revolution: Why the rise of a genetic model of mental illness has not increased tolerance," *Social science & medicine,* Vol. 67, No. 9, 2008, pp. 1370-1381.

Schomerus, G, C Schwahn, A Holzinger, PW Corrigan, HJ Grabe, MG Carta, and MC Angermeyer, "Evolution of public attitudes about mental illness: a systematic review and meta-analysis," *Acta Psychiatrica Scandinavica,* Vol. 125, No. 6, 2012, pp. 440-452.

Schomerus, Georg, Michael Lucht, Anita Holzinger, Herbert Matschinger, Mauro G Carta, and Matthias C Angermeyer, "The stigma of alcohol dependence compared with other mental disorders: a review of population studies," *Alcohol and Alcoholism,* Vol. 46, No. 2, 2011, pp. 105-112. http://alcalc.oxfordjournals.org/content/alcalc/46/2/105.full.pdf

Schulze, B, M Richter-Werling, H Matschinger, and MC Angermeyer, "Crazy? So what! Effects of a school project on students' attitudes towards people with schizophrenia," *Acta Psychiatrica Scandinavica,* Vol. 107, No. 2, 2003, pp. 142-150.

Spagnolo, Amy B, Ann A Murphy, and Lue Ann Librera, "Reducing stigma by meeting and learning from people with mental illness," *Psychiatric rehabilitation journal,* Vol. 31, No. 3, 2008, p. 186.

Sriram, Bharath, Dave Fuhry, Engin Demir, Hakan Ferhatosmanoglu, and Murat Demirbas, "Short text classification in twitter to improve information filtering," *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, 2010, pp. 841-842.

Suh, B., L Hong, P. Pirolli, and E.H. Chi, "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network," paper presented at 2010 IEEE Second Internation Confrence on Social Computing, 2010.

Tagliamonte, Sali A, *Comparative sociolinguistics*: Wiley Online Library, 2013.

Tausczik, Yla R, and James W Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of language and social psychology,* Vol. 29, No. 1, 2010, pp. 24-54.

Tess, Paul A, "The role of social media in higher education classes (real and virtual)–A literature review," *Computers in Human Behavior,* Vol. 29, No. 5, 2013, pp. A60-A68.

Thackeray, Rosemary, Scott H Burton, Christophe Giraud-Carrier, Stephen Rollins, and Catherine R Draper, "Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month," *BMC cancer,* Vol. 13, No. 1, 2013, p. 1.

Tremayne, Mark, "Anatomy of protest in the digital era: A network analysis of Twitter and Occupy Wall Street," *Social Movement Studies,* Vol. 13, No. 1, 2014, pp. 110-126.

Turner-McGrievy, Gabrielle, and Deborah Tate, "Tweets, Apps, and Pods: Results of the 6-month Mobile Pounds Off Digitally (Mobile POD) randomized weight-loss intervention among adults," *Journal of medical Internet research,* Vol. 13, No. 4, 2011, p. e120.

US Department of Health and Human Services, "Mental health: A report of the Surgeon General," 1999.

US Department of Health and Human Services *President's New Freedom Commission on Mental Health (2003)*, Rockville, MD, Services, US Department of Health and Human, 2003.

Vogt, Dawne, "Mental health-related beliefs as a barrier to service use for military personnel and veterans: a review," *Psychiatric services,* Vol. 62, No. 2, 2011, pp. 135-142.

Wang, Shiliang, Michael J Paul, and Mark Dredze, "Social Media as a Sensor of Air Quality and Public Response in China," *Journal of medical Internet research,* Vol. 17, No. 3, 2015.

Wood, Amy L, and Otto F Wahl, "Evaluating the effectiveness of a consumer-provided mental health recovery education presentation," *Psychiatric rehabilitation journal,* Vol. 30, No. 1, 2006, p. 46.

Wood, Lisa, Michele Birtel, Sarah Alsawy, Melissa Pyle, and Anthony Morrison, "Public perceptions of stigma towards people with schizophrenia, depression, and anxiety," *Psychiatry research,* Vol. 220, No. 1, 2014, pp. 604-608.

World Health Organization, "Mental Health: a state of well-being," 2014. As of 12/16/2016:

http://www.who.int/features/factfiles/mental_health/en/

Yang, Shuang-Hong, Alek Kolcz, Andy Schlaikjer, and Pankaj Gupta, "Large-scale high-precision topic modeling on twitter," *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 1907-1916.

Yanos, Philip T, David Roe, and Paul H Lysaker, "The impact of illness identity on recovery from severe mental illness," *American journal of psychiatric rehabilitation,* Vol. 13, No. 2, 2010, pp. 73-93.

Zhang, Ley, Riddhiman Ghosh, Mohamed Dekhil, Meichun Hsu, and Bing Liu, "Combining lexiconbased and learning-based methods for twitter sentiment analysis," *HP Laboratories, Technical Report HPL-2011,* Vol. 89, 2011.