Mining 'Big' Social Data to Learn about Mental Health

Opportunities to Inform Public Policy with Text Analyses

AMOD-5410H: Big Data

March 6th, 2018

Nicholas Hopewell

0496633

Mining 'Big' Social Data to Learn About Mental Health

How do we come to know about and understand mental health? Traditionally, knowledge discovery in research regarding mental health has come from analyses done on survey data with participant samples. Our knowledge has been largely limited to asking questions about mental health and recording responses across a relatively small number of possible selections. Clinical trials are another classic methodology. While such trials have seen extensive use in research, they tend to purposefully exclude individuals with psychiatric comorbidities or multiple medical conditions in order to study a single disorder or condition (Kisely, 2017). Although these methodologies have been most practical in the past, there exists a significant amount of untapped information from social media outlets, which may hold key insights for progressing knowledge and awareness of mental health. This wealth of information can provide enormous opportunities in developing public health policy in Canada and around the world.

Considering how popular social media has become, as well as its potential to broadcast information and ideas to wide audiences, it serves as a highly engaged channel of public opinion. The power of social media is that it has transformed citizens from consumers of information to producers of information. In a time where massive numbers of people share their daily lives and conversations on social media, policy makers ought to use these platforms as vehicles to monitor public opinion about important topics, including mental health issues. How issues arise, gain attention, are influenced, and change over time can be understood through opinion mining of social media data. Learning from large-scale information sharing is an increasingly relevant topic across many domains. Twitter, for example, has been used to predict election outcomes (Tumasjan, Sprenger, Sandner, & Welpe, 2011), forecast stock market prices with 86.7% accuracy (Bollen, Mao, & Zeng, 2011), understand the rise of protest networks (Tremayne, 2014), learn about abuse of prescription drugs (Hanson et al., 2013), and trace trends of public opinion about nuclear power (Kim & Kim, 2014), vaccinations (Salathé & Khandelwal, 2011), and tobacco marketing (Cobb et al., 2013). Such analyses are particularly interesting because they present clear policy implications. Despite the potential value of Twitter data, until recently, research on public mental health has not utilized it. Most of the available data remains a largely untapped resource in mental health research (Sinnenberg et al., 2017). If technical challenges associated with the processing and modelling of very large data can be overcome, social media analytics has the potential to act as a strategic tool to develop and deploy health policy.

Researching large-scale social data may offer an entirely new opportunity for social scientists and policy makers. As tech giants collect increasingly large amounts of information on individuals, leveraging this data may propel social science research to new heights in this era of 'big data'. These data are completely free of experimental and observational limitations as it comes directly from the mind of the individual in their own words. One trade-off when working with such data is that it is often high dimensional and completely unstructured. The nature of such data requires much more processing and cleaning to analyze efficiently. Additionally, due to sample size and because careful procedures and standards are not rigorously applied to massive social media samples, these samples may suffer from statistical heterogeneity (a classical limitation of inferential analyses) and so the application of less-traditional, more robust analytical techniques is often a necessity (Wang, 2017; Gandomi & Haider, 2015). Although such analyses require a new level of technical competence and 'data munging' skills common amongst computer scientists but less common amongst social scientists, gaining these technical skills may open new doors in research.

With the widespread prevalence of social media usage, very recent research suggests that social media analysis can provide important information into mental health-related areas where data previously did not exist (Ueda, Mori & Sawada, 2017). The goal of this paper is to highlight the potential insights offered from big data research using social media data about widespread beliefs and attitudes towards mental health. Such insights gained from social media analyses can, and should be, used to inform public health policy. Furthermore, it should become clear that proper and ethical usage of large-scale social media data could lead to robust analyses, which thrive where traditional analyses are of little use. This paper will focus on data gathered from Twitter, because Twitter offers an open API which allows individuals and researchers easy access to user data (in the form of tweets) which can efficiently be extracted and analyzed.

Mental Health in Canada and the Use of Social Media

Mental health and social media usage are prevalent and pressing topics in Canada, especially amongst young citizens. Every year, approximately 1 in 5 Canadians experience a mental health problem (Smetanin et al. 2011). Furthermore, about 70% of mental problems have their onset during childhood or adolescence (Government of Canada, 2006). Young people aged 15 to 24 are more likely to experience mental health illness than any other age group in Canada (Pearson, Janz & Ali, 2013). Although similar statistics are not as readily available for Canadian citizens, a 2016 PEW research poll found that 88% of Americans aged 18-29 actively used Facebook and 36% actively used Twitter (Pew Research Center, 2016). Sources suggest that the frequency of social media use by young people is steady increasing. An American study of digital access amongst young adults found that in 2014, 89.42% of young adults reported regular engagement with at least one social media platform, a number which rose to 97.5% in 2016 (Villanti et al., 2017). It is apparent that a meaningful demographic overlap exists between ages of most frequent social media use and ages of the most common occurrences of mental illnesses. Young people are both the most active social media users and the most likely to experience a mental illness.

Social media activity may have a strong influence on the development of beliefs and attitudes towards mental illness. Past research has shown that an individual’s attitude toward mental health and illness is a result of an interaction between all parts of their personal background (Ajzen, 2001). Traditionally, this background would be comprised of interactions with family, close personal relationships, friendships, and mass media. A breadth of past research has shown that mass media shapes negative beliefs and opinions about the mentally ill, as well as those who provide mental health services. The media often portrays the mentally ill as potentially dangerous and undesirable and those who treat the mentally ill as unprofessional and untrustworthy (Review: Smith, 2015). In modern times, social media contributes significantly to a person’s experiences and background, with greater numbers of people relying on social media for news. Therefore, it is likely that social media has a significant impact on the way people understand and talk about mental health related issues. Research that investigates this linkage may be key to revealing how social media can combat stigma and improve help-seeking behaviours in the mentally ill.

Although attitudes towards mental health appear to be improving, with 57% of Canadians surveyed in 2015 believing that stigma surrounding mental illness had reduced over the past five years (Bell Canada, 2015), stigma still appears to be a significant barrier. In 2016, 40% of people surveyed reported experiencing anxiety and depression but did not seek professional help (Shoppers Run for Women Poll, 2016). Furthermore, 39% of Ontario workers have reported that they would not inform their managers if they were experiencing mental illness (Dewa, 2014).

This is perhaps because stigma is a key deterrent to seeking help. Research has consistently found stigma to be a leading barrier to looking for and participating in help and care regarding mental illness. People who experience mental health stigma feel less inclined to seek professional help (Clement et al., 2013; Corrigan, Druss, & Perlick, 2014). Stigma towards individuals experiencing mental health problems continues to be prevalent. Some commonly held beliefs are that these individuals are in some way responsible for their illness - perhaps due to some inherent weakness or laziness - and that they are incompetent and a burden on society (Lakeman et al., 2012; Beldie, 2012). Mental health professionals face similar stigma. These associations are important to research because they allow sentiments regarding mental illness and those who suffer from mental illness to be classified as positive or negative. Such classifications may be used to track developments in attitudes over time and predict profiles of individuals who are not likely to seek help.

Stigma regarding mental health may be direct, such as the refusal to work or do business with an individual who suffers from mental illness, or indirect, such as extending assumptions about an individual's limitations beyond what is accurate or reasonable. Attitudes and beliefs regarding mental health and illness influence how individuals interact face-to-face and online with individuals experiencing mental illness. Stigma research suggests that people with mental illness experience social exclusion, bullying and harassment from peers, as well as feelings of isolation and detachment (Elkington et al., 2012; Moses, 2010). Especially common amongst young people, mental health stigma is often internalized, altering an individual's sense of self-worth and belonging in social groups and society (Kranke & Floersch, 2009). Victims of stigma also experience serious overarching consequences in terms of receiving unequal access to community involvement, appropriate health care, employment opportunities, criminal justice, and education (Knaak, Mantler, & Szeto, 2017; Livingston, 2013). Approximately half of Canadians who have experienced major depression will receive adequate care (Patten et al., 2016). Furthermore, mental health care in Ontario is reportedly underfunded by about $1.5 billion, with only one third of Canadians aged 15 or older who require mental health care stating that their needs are adequately met (Brien et al., 2015; Sunderland & Findla, 2013).

There are clear values in mental health research and improving popular perceptions of mental illnesses. Firstly, improvements in widespread opinions about individuals living with mental illness has been shown to result in greater use of mental health services and overall improvements in public health (Corrigan, 2004). The use of mental health services should be a key talking point, as approximately 4,000 suicides occur in Canada every year (Statistics Canada, 2017). Furthermore, the economic impact of mental illness costs Canada tens of billions of dollars per year, accounted for by the health care costs and loss of national productivity (Center for Addition and Mental Health, 2017). Furthermore, individuals with mental illness account for a disproportionally large share of health care costs in Canada, incurring 30% more health care costs than other patients (De Oliveira et al., 2016). Across the world, mental illness is the fifth largest single contributing factor to the global burden of disease (Ferrari et al., 2014).

As the expansion of computer architecture and support for efficiently processing big data continues, the demand for mental health research to leverage social media data is clear. Improving information systems for mental health, including significantly increasing capacities for public health monitoring was a key goal of the 2013-2020 World Health Organization's Comprehensive Mental Health Action Plan (Mental health action plan 2013-2020). Developments in machine learning, as well as natural language processing, introduce expansive opportunities in public health improvement for social scientists. One key point is that this research area is inherently interdisciplinary; newly evolving domains at the intersection of social/behavioural science (mostly psychology and clinical psychology) and computer science await those who wish to study mental health through acquiring the technical expertise needed to process and analyze big data. Although past research has explored online peer-to-peer support over social media (E.g., Naslund, Aschbrenner, Marsch, & Bartels, 2016), the value of large-scale analyses of general social media interactions and trends in order to understand mental health has recently become clearer.

Big Data, Social Media Analytics, and Public Health

As previously mentioned, social media has become a well-established source of data in a wide-range of areas, with Twitter data being utilized for its ease of access. Twitter provides insight into people's raw and unfiltered opinions in a way that is impossible in an experiment or clinical setting and unrealistic using survey data. These data come from the users themselves (e.g., age, gender, ethnicity, number of followers, number following, number of total tweets) as well as posts (e.g., number of retweets, hashtags used, links to other sites; as well as language and sentiments). One might argue that this naturalistic access to public opinion and belief is a more accurate representation of what social scientists desire to study and highly indicative of emotional wellbeing. Twitter offers natural observation of human behavior and discourse without imposing the biases of traditional research. Twitter is a particularly exceptional resource for social science research as it has stored and provided open access to the attitudes and opinions of millions of people since its launch in 2006. Because every public Tweet is searchable and scrapable, research on how beliefs and attitudes change over time, or after notable events, is very feasible with access to Twitter data. Twitter data also offers unique opportunities to explore which individuals or movements are shaping and progressing conversations online. Though the research opportunities with Twitter and other social media data are vast, social media data introduces new biases and limitations discussed at length in a later section of this paper.

One of the key distinguishing factors which contributes to the strength of social media analytics for understanding mental health are their feasibility. These analyses can be carried where more traditional analyses would require too much time, coordination, and funding, or would be almost impossible to conduct. Take the example of a health researcher who desires to study the impact of publically known and discussed tragedy, such as a school shooting or hate crime, on public opinions about mental health and safety. A traditional approach to understanding these phenomena is limited to studying the impact of the event sometime after it had occurred on a specific sample, which may or may not be a desirable sample for the research. It is apparent that this approach has a scoping issue; it is impossible to understand opinion, beliefs, and discourse leading up to and even *during* the event. This missing information may have led to conclusions and insights which could augment understandings about mental health as it is known and discussed by the public. Although this example highlights a potential limitation of traditional analyses, this limitation is true for most traditional mental health analyses. Traditional mental health research typically does not have access to the time and resources to study longitudinal effects of important issues regarding mental health. Although these issues cannot be understood in such a way via traditional methodologies, social media allows this type of analysis and provides a wealth readily available information to expand research into the future. As such, the value of social media analytics to social scientists and policy makers is vast.

A growing body of research has explored how social media can be used to learn about mental health. Research indicates that individuals suffering from mental illness may be more inclined, or at least *as* inclined, to rely on social media activity for their social interactions, compared to other people. Social media may provide these individuals an outlet to overcome the social isolation they experience while avoiding the social stigmas regarding mental illness (Ruiz Lee, 2015; Naslund, Aschbrenner & Bartels, 2016). Social media also offers an opportunity to communicate with others and establish meaningful and empathetic social connections. Researchers have found that non-depressed Twitter users view the platform primarily as a means of sharing and consuming information, while depressed users see the platform as a tool for spreading social awareness and providing emotional interaction (Park, McDonald, & Cha, 2013).

In addition, a growing body of research is exploring whether a person's mental state can be accurately be predicted based on their social media activity and postings (E.g. De Choudhury, Gamon, Counts, & Horvitz, 2013; Park, McDonald, & Cha, 2013). One study suggests that references to depression or drinking posted on social media are correlated with self-reports of depression amongst college students (Moreno, Christakis, Egan, Brockman, & Becker, 2011). Another study, which analyzed the social media postings of new mothers, concluded that reduced social interaction on Facebook was a strong predictor of postpartum depression (Choudhury, Counts, Horvitz, & Hoff, 2014). Recent research has used social media analytics to understand shifts in suicidal ideations and use these shifts as markers to predict and identify individuals who are likely to engage in suicidal ideation in the future (Choudhury, Coppersmith, Dredze, Kiciman, & Kumar, 2016).

To understand suicide trends in Japan, one study integrated Twitter activities with traditional media and statistical analyses. Researchers found that increased numbers of suicides in the Japanese public followed the suicides of well-known individuals and celebrities only when these suicides were discussed widely on Twitter. Crucially though, when no significant conversation on Twitter surrounded the death of a celebrity or well-known individual, the number of suicides in the Japanese public did not increase, even if these suicides were covered by traditional media and news outlets (Ueda, Mori, Matsubayashi, & Sawada, 2017). The researchers concluded that trends on social media may lead people to engage in social learning of behaviours which their online communities see as normal and acceptable, a process which likely accounted for the increased suicide trends following Twitter coverage of celebrity suicides. This study in particular highlights the influence of social media on mass cultural and behavioural trends, and the importance of leveraging social media to learn about mental health.

Gruebner et al. (2017) provides an overview of a few interesting new directions in social media mining for mental health research. One of these directions involves geospatial analyses of social media data with geographical location tags. An example would be to understand how the local conditions such as regional poverty and access to parks affect people's mental health, and how changes in local conditions may predict mental health outcomes overtime. Geospatial modelling of such relationships could lead to the ability to make predictions about which areas, during which times, are most at risk for mental health crises. Such analyses could inform public health policy in terms of providing knowledge about where and when aid and social service funding is most necessary in order to allocate resources in the most effective way. Researchers also state that social media analyses following occurrences such as terrorist attacks could study the social support connections people develop and whether these supports predict mental health outcomes and recovery to a similar extent compared to real social supports. Additionally, opportunities to construct in-depth social profiles of individuals, resembling social media case studies, may allow researchers to gain a more intimate understanding of the precursors and social symptoms of mental illnesses. These data could be integrated with traditional sources of information taken from individual's community or social circle to create holistic representation of the individuals experience and environment. As new research is published, optimal methodologies for analyzing social media to understand mental health should develop.

Appropriate Techniques for Social Media Analyses

Appropriate techniques to study mental health using social media data come from well-known research papers in this new and developing area. These papers are respected for their research design quality and possible implications for public health. Zaydman (2017) provides a very good summary of these types of analyses and how they can apply to Twitter data. These analyses, covered below, include machine learning/supervised classification, topic modelling, sentiment analysis, and network and cascade analysis.

Machine Learning/Supervised Classification

Machine learning involves the application of induction algorithms that can derive rules about known data to predict the state of similar unknown data (Kohavi, & Foster, 1998). Machine learning has two overarching categories: supervised and unsupervised learning. Supervised machine learning occurs when an induction algorithm is provided a set of data, comprised of numerous attributes, as well as a nominal outcome variable with two or more levels (non-continuous). This outcome variable contains class labels associated with each record of the data set. The induction algorithm learns a relationship between each attribute in the data set and the outcome variable that allows the class label of new records to be predicted (Pang-Ning, Steinbach, & Kumar, 2005). For instance, if the data set contained records of different animals and the attributes of the data set where the different characteristics of each animal (such as lays eggs, has hair, is nocturnal, etc.), these class labels might include the labels 'warm-blooded' and 'cold-blooded'. In this example, the induction algorithm would parse through the data and determine which characteristics were associated with warm and cold-blooded animals, and use these as rules for classification. When new animals were entered into the data set that did not have accompanying class labels, the classifier could use the rules induced from the previous data (where class labels were known) to map the new animals onto appropriate class labels automatically. On the other hand, unsupervised machine learning methods do not involve class labels and try to uncover knowledge about the underlying structure or relationships within the data (Pang-Ning, Steinbach, & Kumar, 2005).

Supervised machine learning with social media data is useful for many different applications and allows researchers to predict public health issues based on discussions that occur outside of traditional public health spaces. Supervised machine learning algorithms have been applied to Twitter data to categorize postings from users about e-cigarettes (Cole-Lewis et al., 2015), classify postings about alcohol consumption (Yin, Ray, Statnikov, & Krebs, 2014), and much more. Adopting a machine learning approach to understanding public health can aid social scientists in the development of methods for automatically generating epidemiological information from rapidly changing social media networks, providing immediate access to information in real time. Researchers suggest that insights gained from such tools should be leveraged during the development of governmental communication strategies and policy in order to better influence public health behaviours with more accurate and relevant information (Cole-Lewis et al., 2015). Machine learning allows public health research to scale to expansive new heights; researchers can now rapidly gain information about the behaviours and attitudes of people on a very large scale. Learning from such a large number of people in such a small amount of time is impossible or impractical with non-automated methods.

Topic Modelling

Blei (2012) provides a useful explanation and overview of topic modelling. As knowledge and information continues to expand and stored digitally, discovering what we set out to find becomes an increasingly difficult task to manage. Topic modelling provides a way to automatically organize and understand large amounts of information, making knowledge discovery a much more manageable task. Topic modelling is an unsupervised machine learning algorithm which finds groups of themes within text (referred to as topics). This process involves grouping words that are related and separating words that are not related. Because topic modelling is an unsupervised learning technique, the text does not require prior thematic labels. Topic modelling affords the ability to search through texts for common themes, whether broad or specific, and understand how these topics develop over time. How these common themes relate and change are modelled as statistical probabilities. Although this method is applicable many types of documents, it may be challenging to apply to Twitter data because tweets are often too small to allow the derivation of an underlying structure (the goal of unsupervised approaches).

One topic modelling technique that is applicable to social media analyses is Latent Direchlet allocation (LDA). Introduced by Blei, Ng, and Jordan (2003), LDA is based on a generative process - an imaginary random process by which documents of text are assumed to have arose. The purpose of LDA is to generate a set of probabilistic topics from text, each having a list of associate terms. The idea is that each document in the text share the same topics, but these topics are represented to a greater or lesser extent (proportionally) in the different documents. Thus, topics can be thought of as probability distributions of terms. For instance, the topic 'jazz' would have associated term such as 'band' or 'music' with high probabilities. These topics are assumed to exist before the document is generated. Words in each document are generated first by randomly choosing a distribution of topics, then a single topic (with associated words) is randomly chosen from the topic distribution, and finally, a random associated word is chosen from the topic. This process repeats until a text is generated from the data (for every word in the document). This analysis provides a very powerful solution for gaining knowledge about large volumes of textual data.

The topic structures of documents LDA tries to find are latent; these structures 'lie hidden' in the text data and must be inferred through the document itself. If the generative process is imagined as probability distributions of topics and terms which generate a document, then inferring the topic structure of a known document can be thought of as the opposite process. In making inferences about topic structures, LDA explores which topic structure most likely generated the known document. LDA learns the topic structure of a document automatically with no prior information about its subject matter. The result of LDA is a document of words that are now annotated based on the topic structure, allowing researchers to organize and get important information from the document quickly. The resulting topic structure derived via LDA should resemble the actual qualitative meaning of the document (I.e., derived topics should match actual topics).

LDA is particularly useful for Twitter analyses because it may allow researchers to arrive at hypotheses about the themes of conversations over tweets. As the subject matter of a large number of tweets cannot feasibly be determined manually, implementing the LDA algorithm becomes a necessity for big data analyses of text. LDA models of social media data have been used to study a variety of health issues such as flue surveillance (Chen et al., 2015) and how anonymous social media platforms such as Reddit may offer support for those who suffer from highly stigmatized mental illnesses (De Choudhury & Sushovan, 2014). Furthermore, LDA has been used to study homophily on Twitter, a concept referring to the idea that people develop homogenous social networks. Researchers found that Twitter users reinforce homophily by following other users who post about and interact with topics that they both find interesting (Weng, Lim, Jiang, & He, 2010). LDA with Twitter data has also been adapted to improve public health monitoring systems, providing a scalable method to maximize information gain that is reliable when compared to governmental public health records (Paul & Dredze, 2011).

Sentiment Analysis

Sentiment analysis - also referred to as opinion mining - attempts to provide a reproducible and reliable method of extracting and quantifying the underlying affective meaning of text to allow judgments to be made about its otherwise subjective properties (Pang & Lillian, 2008). These subjective properties most commonly include opinions, beliefs, attitudes, and feelings expressed over text. Such properties are difficult to decipher reliably with human judgement alone, as much of the accompanying social information (expression, gesture, tone, etc.) is lost when messages are communicated via text. Sentiment analysis derives from natural language processing and computational linguistics, and relies on the notion that an individual's use of language (including written language) typically reflects their emotional state. Sentiment analysis could be used to determine the attitudes and emotional reaction of individuals to a product launch using text taken from an online forum, for example. An interesting sentiment analysis conducted by Eichstaedt et al. (2015) correlated the use of hateful words on Twitter to community rates of heart disease. Similar analyses can be conducted to understand people's reactions and attitudes towards topics and postings on social media platforms. A sentiment analysis of social media data may look at trends in anger words surrounding Trump tweets, trends in sad and empathic words surrounding a notable death or tragedy, as well as trends in positive words surrounding social media campaigns.

As for most types of analyses, sentiment analyses can be conducted in multiple ways, all sharing the goal of studying affective states through text. Lexical-based sentiment analysis depends on a sentiment lexicon - a dictionary of known sentimental terms that reflect positive, negative, neutral sentimental features. This dictionary is referenced to match terms in the text to known terms in the lexicon, each with associated values. The process of matching terms allows the values of groups of terms, such as sentences, to be summed, and an overall positive or negative (or some encoded emotion) score to be attached to the sentence. Dictionary matching is not limited to single terms, but also groupings of terms, referred to as n-grams. Lexicon-based sentiment analysis can also involve more complex dictionaries that associate terms with parts of speech (nouns, verbs, adjectives, etc.). These more complex dictionaries allow values attributed to terms to change depending on which part of speech the term occupies. As an example, the term ‘well’ can be used in different ways that may have impact on the emotional content of a message. The following sentences all use well as different parts of speech: My day is going well (adverb); All is well in the world (adjective); Well, at least someone has a brain (interjection); There is a well outside the school (noun); Tears well up in my eyes (verb). Without making the distinction between parts of speech, the emotional value of a sentence or other grouping of terms may be inaccurate.

Machine learning sentiment analysis of text involves implementing supervised classification to give emotional context to text. This is done by first diving the data into a subset where sentimental attributes are manually encoded by the researcher. Some text is easy to assign a sentimental category to, while others may be ambiguous and difficult to categorize. A classifier is then induced from this subset of encoded data and applied to the non-encoded data to quantify the performance of the classifier. As long as the sentimental features are adequately represented in the encoded subset used to train the classifier, the classifier will learn the rules required to predict unseen records associated with each feature correctly. Machine learning sentiment analysis is particularly applicable to social media data due to the unique use of language online and how language changes over social media. For instance, new non-formal terms often arise in social media, such as abbreviations or entirely new words, often with seemingly no direct connection to the meaning of the word. In addition, some words are intentionally misspelled or misused over social media by certain groups of people. This poses a major problem for lexical-based analysis that would likely have no way to match most of these terms. On the other hand, supervised machine learning can learn these patterns and sort through the data in an effective way without referencing a dictionary. Lexical and machine learning approaches can be applied, gaining the benefits of both and possibly resulting in a better model.

Network and Cascade Analysis

Network analysis has obvious applicability to studying social networks on platforms like Twitter. Network analysis attempts to understand relationships between individual entities by applying graph theory. The background of this theory is extensive and not within the scope of the current paper. Network analysis has been used to study the spread of infections (Christley et al., 2005), extremist groups (Koschade, 2006), the movements of first responders (Erikson et al., 2014), and much more. When applied to social media data, network analyses can model social interactions and geographic distribution of users. Learning about key actors of online communities and social hot spots are possible with these analyses.

Network analyses provide interesting details about the structure of a network, such as degree centrality of a node in the network. Centrality refers to the number of direct connections to a particular node in the network. Betweeness centrality refers to the number of times a node acts as a connection between the shortest paths of two other nodes. This measure attempts to quantify the level of control of one node on the communication between other nodes. When looking at social media, this would measure how one person controls or facilitates communication between two other people. Closeness centrality reflects the mean distance of the shortest paths of a particular node to all other nodes in the network. The more central a node is the more close it is on average to the other nodes of the network. As average distance is being measured, closeness only makes sense to calculate in fully connected networks with connections coming from multiple directions – single direction networks excluded. Finally, eigenvalue centrality quantifies the influence of a node within a network by measuring the number of highly connected nodes to the node (Freeman, 1979; Borgatti, 2005). This measure provides an understanding of how far a tweet sent from one user could travel through the network. When modelling social media interactions, these measures allow potential influencers - those who drive discourse online - to be recognized.

Perhaps the most important information gained by network analyses is how information moves through a network of nodes (in this case, social media users). Twitter is the ideal platform for these analyses for many reasons. Firstly, trends can be easily located and followed by searching for a key word or phrase preceded by a hashtag (#). Twitter also allows one directional connections between users in the form of followers, and bidirectional connections in the form of mutual friends. Analyzing the direction of social media interactions allows researchers to discover how information is shared and received, as well as what information is visible to which users. Twitter also provides information on direct communication and interaction between individual users. Twitter allows users to tweet directly 'at' other users using the '@' symbol, as well as 'like' (indicate agreement) and 'retweet' (share) their postings. As users like and share posts, this allows the posts to be seen by an individual's friends and followers. This system allows networks to model how one user may pass information to another user or to a group, and how that information is propagated forward to new users and audiences. Although the scope of these discussions are well within the capability of network analyses, the meaning of the network and its parts may require some qualitative interpretation by the researcher. For instance, when one user retweets another user, there is no absolute meaning to the action. A retweet may mean the user finds the post interesting or agreeable and desires to share it, or that the user disagrees with the post and wants to make their opinion known. Overall, although the undercurrents of social interactions can be modelled through network analysis, the meaning may need to be inferred or speculated.

The precise method used for analyzing how information moved through a social network is referred to as cascade analysis. This type of analysis has potentially large implications for public health policy and is particularly relevant in today's world. Cascade analyses allow researchers to forecast what information is most probable to spread throughout a social network and pick out false and destructive trends of information that may inform potential prevention efforts (Hui et al., 2012). As previously stated, social media data may be used to determine when and where crucial resources can be allocated most efficiently. In an era of big data, information is a crucial resource now more than ever. Cascade analysis may allow researchers to determine which networks require accurate emergency information and help the spread of such information (Hui et al., 2012). The result of such interventions should be healthier public discourse. Many ethical considerations must be accounted for when efforts are made to alter or shape public discourse; it is important for researchers to consider the potential consequences of these interventions. Cascade analyses with Twitter data are usually conducted by creating a network of individuals, based on followers or shared content, and searching for instances where certain content appears chronologically. This content may be a specific tweet, a certain video or image, or an off-site link.

Past research in cascade analyses have unearthed some important general properties of social networks. Lerman and Ghosh (2010) tracked the number of likes and retweets across many Twitter stories (representing the size of the cascade), and monitored how they grew in early stages of visibility. The researchers found that the rate a story spreads depends on the density and size of the network. More dense social networks spread information faster early on compared to more sparse networks. The size of the network determines if this pace will continue and how long it will take the cascade to slow down, however. A dense but smaller network may share a story very quickly but the rate of sharing will slow down rapidly as non-connected users see the story. A larger network may sustain a healthy rate of information sharing for longer periods as more connected users exist to see the story. Not surprisingly, individuals with a high number of followers are most likely to start cascades, but attempting to predict which users with average numbers of followers will start cascades may be unreliable (Bakshy, Hofman, Mason, & Watts, 2011). Furthermore, the type of story shared on Twitter may have great influence on the ways the story is spread across networks. Researchers have found that topics that produce persistent hashtags that commonly reappear to users are more likely become adopted and spread. These topics were usually controversial or political, and, through repeated exposure, these stories became more widely shared (Romero, Meeder, &Kleinberg, 2011).

Limitations and Ethical Challenges of Social Media Analytics

Although the opportunities for mental health research presented by social media analyses are vast and have great implications for public health, these analyses come with unique challenges, limitations, and ethical considerations. Firstly, naturally occurring biases accompanying social media data may bring some to argue these data are not representative of the general population. People who live near or frequently come to areas with publically accessible wireless internet access, such as downtown areas, would likely be overrepresented in the data compared to people outside of these areas. More specifically, data may be biased towards very specific locations, such as cafes and airports, where public wireless connections are widely used. This introduces a potential bias. Perhaps there are underlying similarities between people who are more likely to be found at cafes and airports that has influence on the data and limits the inferential scope of the analysis. Even more generally, individuals who do not engage in mobile social media activities might show different posting patterns compared to those who do. This could also introduce a representation bias in the data overall or at certain times of the day.

Activity on social media would likely show inherent trends within different age groups and ethnicities including when, where, and how frequently they post. This would introduce important relationships to the data, which would need to be investigated and controlled for to arrive at valid conclusions. Similarly, when looking to explore any phenomenon or relationship using social media data, it should be considered that the data may be biased across genders, ethnicities, and socio-economic status depending on the topic of interest and the location of the users. Furthermore, people from different areas, countries, age groups, and ethnicities may use one social media platform more frequently than another, biasing representations of these people depending on which platform is studied.

Biases introduced from age and socio-economic status of social media users in particular present vital ethical considerations that should be addressed. If the goal of social media analyses are to make public health policy more well informed, people of different ages and socio-economic status must be adequately represented. People of lower socio-economic status may engage in social media less due to limited access to technology compared to others. Additionally, recent research has found that there are significant differences in age, gender, and socio-economic status of Twitter users who choose to allow their tweets to be geotagged versus those who do not (Sloan & Morgan, 2015). This has significant implications for any social media research that studies relationships involving geographic location, as the available data may have immediate biases that must be addressed.

Additionally, demographic biases in social media data might be much more difficult to address when compared to data obtained using traditional research designs and sampling methods. Because careful sampling procedures are not used, one can only verify the demographics of individuals to a certain level of accuracy. A user may choose not to include his or her age or location, or may provide false information about themselves. Because users may be opposed to providing their demographic information on social media, the extent of this limitation could be significant. If adequate methodologies to overcome these limitations are not developed, the consideration of how these limitations impact the validity and scope of research findings must be adequately discussed.

In addition to potential biases, many ethical considerations involving privacy and consent are relevant when analyzing social media data. Firstly, although social media postings may be publically available, it may not always be ethical to use these posts depending on the research topic. Given the sensitivity of mental health and implications of mental health research, this consideration should be a primary concern for researchers. Considering mental health research often involves vulnerable individuals, such ethical considerations are far-reaching and these considerations should be no less extensive for social media analyses of mental health. Some might even argue that these considerations are even more important, as individuals have not formally agreed to participate in the research. It is crucial that researchers consider whether their research may be intrusive or damaging to individuals involved.

Eysenbach and Till (2001) stress that a key consideration on the part of the researcher is the extent to which users perceive their online posts or interactions as more public or more private. Social media posters do not expect to one day be part of research and may be opposed to the idea. The researchers mention that it is important for social media researchers to determine sufficient ways to protect the confidentiality of users involved in their analyses and conclude whether or not formal consent from the users is required or appropriate and be able to justify why it is not required. Typically, informed consent is not required in any situation where an individual's information is publically available. That being said, it can be argued that publically available information taken from social media is qualitatively different from other publically available information (such as news articles or online biographies). This may be the case because an individual may be seeking public visibility when they allow their information to be public in some ways but not in others. Researchers state that individuals who post on social media may not desire public visibility, and thus the simple distinction between public and private communities may not be adequate to address ethical challenges. This distinction may be even more relevant when considering mental health-related posts on social media. It is perhaps best to err on the side of caution and consider to what extent individuals perceive their posts to be publically visible and the implications of working with such data. Eysenbach and Till (2001) provide some guidelines for measuring the extent to which postings may be perceived as more public or more private, including whether or not the platform is subscription based, the size of the social community in which the posting was made, and the scope of the intended audience.

Additionally, focus groups conducted with Twitter users have found that even when users are accepting of the use of their social data for research purposes, many do not fully realize the extent of their digital footprint (Conway & O'Connor, 2016). Some users do not realize that deleting posts or reaching a scroll limit where posts become invisible to other users does not mean that the data is unavailable to researchers. Furthermore, some Twitter users reported feeling uncomfortable when their tweets reached audiences beyond those which they intended to reach and were unaware of the power of sophisticated algorithms in terms of findings and organizing their tweets via pattern recognition. Ultimately, some users had a false sense of the level of anonymity of their posts, not realizing how quickly their past posts could be found and extracted. These users, who are less in touch with the capabilities of modern technologies, are more likely to fail to protect their data and leave a larger digital footprint for researchers to discover. By consistently failing to protect their social media postings, these individuals may not be aware that they are consenting to their posts being used research purposes. The researchers discuss how this raises ethical concerns for demographics of people who are statistically less likely to be technologically literate, such as older women. These individuals may not understand that failing to protect their information is a form of public consent.

Another issue that arises from the use of social media data is whether there are significant differences between how people behave online versus in person, and whether those differences have meaningful implications for answering key research questions. If the answer to these questions is yes, then another concern is raised: is it ethical to learn from a group of people based entirely on their online activities and use insights gained to inform public policy? Furthermore, it people may not feel as like their online activities represent their true opinions and beliefs, and might disagree with using such information to build profiles of their tendencies and needs. Because social media methodologies are in their infancy when it comes to mental health research, questions about the validity of the insights gained may be a point of conversation. Researchers should address how and why social media data in particular has enabled findings that represent a population. Perhaps future research could integrate social media data with more traditional data sources specifically to address the concern of differences in behaviour online and offline.

Finally, researchers should address why they chose to research mental health with social media data, as opposed to other methodologies. Although an increasing number of social and behavioural science researchers are realizing the value of social media data, research in the area should not be done for the sake of following innovative trends. Innovative methods alone do not result in quality research that can progress our understanding of mental health. To develop this area with good research, mental health questions must be explored in an ethical manner. Considering the great potential benefits of social media research, ethical considerations should act as a challenge and not as a barrier for researchers. Ultimately, the onus is on the researcher to make the potential benefits of using big social media data clear and explain how these benefits outweigh potential privacy and confidentiality concerns.

Despite privacy and confidentiality concerns during social media analyses, this does not mean such research is inherently flawed. Utilizing large-scale social media data consisting of millions or tens of millions of users means that personal identification of specific users is extremely unlikely. The most significant ethical challenges in these analyses do not involve the handling of individual records of users, but rather the insights researchers gain from groups of individuals. Issues would be more likely to arise from how the researchers discuss and portray groups of people and their needs/desires. Without careful ethical consideration, groups may be portrayed in ways that researchers did not intend or that the groups in question disagree with. These situations may cause privacy issues to bubble to the surface and result in potential conflicts between research and public experience. Considering that mental health research is at a high risk of resulting in unintended consequences that can be damaging to groups involved, great ethical attention should be given to these studies.

Conclusion

Mental health is an important public health topic in Canada, and researchers will play an important role in advancing knowledge about public health-related opinions and behaviours. These insights will in turn generate better-informed policy decisions. With advancements in data storage and processing capabilities, researchers have access to increasingly large amounts of open data that can and should be leveraged across all applicable domains. In the information age, we can research mental health using social media platforms like Twitter that provide access to unfiltered longitudinal data that cannot be gained from traditional research methodologies. Opportunities for social scientists to study public health with large-scale social media data are plentiful. Social media provides a medium that allows for quick and scalable analyses of public health matters, enabling researchers to uncover trends in public attitudes and opinions that may be used to inform more relevant and accurate governmental strategies. Efforts in social media analyses of big data may produce effective ways to identify issues that are important to public health and spread quality of life information to citizens effectively. Though these opportunities abound, social media analytics also presents ethical challenges that should be considered by researchers while best practices are still being developed. Ethical concerns are especially relevant to mental health research, as the potential for unexpected or unintended consequences to be damaging is serious.

References:

Ajzen, I. (2001). Nature and operation of attitudes. Annual review of psychology, 52, 1. 27:58.

Bakshy, E., & Hofman, J.M., Mason, W. & Watts, D, J. (2011). Everyone's an Influencer: Quantifying Influence on Twitter. Proceedings of the 4th ACM International Conference on Web Search and Data Mining, WSDM 2011. 65-74. 10.1145/1935826.1935845.

Beldie, A., Den Boer, J.A., Brain, C., Constant, E., Figueira, M.L., Filipcic, I., Gillain, B., Jakovljevic, M., Jarema, M., & Jelenova, D. (2012). Fighting stigma of mental illness in midsize European countries. Social psychiatry and psychiatric epidemiology. 47(1): 1-38.

Bell Canada (2015). Bell Let’s Talk: The first 5 years (2010-2015). Retrieved from http://letstalk.bell.ca/letstalkprogressreport

Blei, D. M. (2012). Probabilistic Topic Models. Communications of the ACM.55(4). 77-84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of Machine Learning Research, 3, 993–1022.

Bollen. J., Mao, H., Zeng, X. (2011). Twitter mood predicts the stock market. J Comput Sci 2:1-8.

Borgatti, S.P. (2005). Centrality and network flow. Social Networks. 27(1), 55-71.

Brien et al. (2015). Taking Stock: A report on the quality of mental health and addictions services in Ontario. An HQO/ICES Report. Toronto: Health Quality Ontario and the Institute for Clinical Evaluative Sciences

Center for Addition and Mental Health (2017). Mental Illness and Addictions: Facts and Statistics. Retrieved from: http://www.camh.ca

Chen, L., Hossain, KSM, T., Butler, P., Ramakrishnan, N., & Prakash, A.B. (2015). Syndromic surveillance of Flu on Twitter using weakly supervised temporal topic models. Data Mining and Knowledge Discovery. 1-30.

Christley, R.M., Pinchbeck, G.L., Bowers, R.G., Clancy, D., French, N.P., et al. (2005) Infection in social networks: using network analysis to identify high-risk individuals. Am J Epidemiol 162: 1024–1031.RM.

Cobb, N., Chapman, W., Conway, M., Emery, S.L., Hernandez, T.M., Myslín, M., & Zhu, S. (2013). Using Twitter to Examine Smoking Behavior and Perceptions of Emerging Tobacco Products. Journal of medical Internet research.

Clement, S., Schauman, O., Graham, T., Maggioni, F., Evans-Lacko, S., Bezborodovs, N., . . . Thornicroft, G. (2013). What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Manuscript submitted for publication.

Cole-Lewis, H., Varghese, A., Sanders, A., Schwarz, M., Pugatch, J., Augustson, E. (2015). Assessing Electronic Cigarette-Related Tweets for Sentiment and Content Using Supervised Machine Learning. J Med Internet Res.

Conway, M. & O'Connor, D. (2016). Social media, big data, and mental health: current advances and ethical implications. Curr Opin Psychol. 9:77–82. doi: 10.1016/j.copsyc.2016.01.004.

Corrigan, P. (2004). How stigma interferes with mental health care. American Psychologist. *59*(7), 614-625. http://dx.doi.org/10.1037/0003-066X.59.7.614

Corrigan, P.W., Druss, B.G., Perlick, D. (2014). The impact of mental illness stigma on seeking and participating in mental health care. *Psychological Science in the Public Interest*.15:37–70. https://doi.org/10.1177/1529100614531398.

De Choudhury, M., Counts, S., Horvitz, E., & Hoff, A. (2014). Characterizing and predicting postpartum depression from shared facebook data. *CSCW*.

De Choudhury, M., Coppersmith, G., Dredze, M., Kiciman, E., & Kumar, M. (2016). Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. Proceedings of the SIGCHI conference on human factors in computing systems. CHI Conference. 2098-2110.

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting Depression via Social Media. In Proc. ICWSM 2013.

De Choudhury, M., & Sushovan D. (2014). Mental health discourse on reddit: Self-disclosure, social support, and anonymity. Proc. ICWSM, AAAI.

De Oliveira et al. (2016). Patients with high mental health costs incur over 30% more costs than other high-cost patients. Health Affairs. 35: 36-43.

Dewa (2014). Worker attitudes towards mental health problems and disclosure. International Journal of Occupational and Environmental Medicine. 5: 175-86.

Eichstaedt, J.C., Hansen A.S., Margaret L.K., Gregory P., Darwin R.L., Raina M.M., Sneha J., Megha A., Lukasz A. D., and Maarten S. (2015). Psychological language on twitter predicts county-level heart disease mortality. Psychological science. 26(2). 159-169.

Elkington, K., Hackler, D., McKinnon, K., Borges, C., Wright, E. R., & Wainberg, M. L. (2012). Perceived mental illness stigma among youth in psychiatric outpatient treatment. Journal of Adolescent Research, 2, 290e317.

Erickson, J. Chastain, K., Patton, E.W., Fry, Z., Yan, R., McCusker, James., & Mcguinness, D. (2014). Identifying first responder communities using social network analysis. CEUR Workshop Proceedings. 1272. 425:428.

Eysenbach G, Till J., E (2001). Ethical issues in qualitative research on internet communities. BMJ. 23:1103-1105.

Ferrari AJ, Norman RE, Freedman G, Baxter AJ, Pirkis JE, Harris MG, Page A, Carnahan E, Degenhardt L, Vos T. (2014). Whiteford HA: The burden attributable to mental and substance use disorders as risk factors for suicide: findings from the global burden of disease study 2010. PLOS ON, 9:e91936.

Freeman, L.C. (1979). Centrality in networks: I. Conceptual clarification. Social Networks, 1. 215-239

Gandomi, A., & Haider M. (2015). Beyond the hype: big data concepts, methods, and analytics. Int. J. Inf. Manage., 35 (2) (2015), pp. 137-144.

Government of Canada (2006). The human face of mental health and mental illness in Canada. Ottawa: Minister of Public Works and Government Services Canada.

Gruebner et al. (2017). Big data opportunities for social behavioral and mental health research Soc. Sci. Med. 189: 167

Hanson, C. L., Burton, S. H., Giraud-Carrier, C., West, J. H., Barnes, M. D., & Hansen, B. (2013). Tweaking and Tweeting: Exploring Twitter for Nonmedical Use of a Psychostimulant Drug (Adderall) Among College Students. Journal of Medical Internet Research. 15(4), e62. http://doi.org/10.2196/jmir.2503

Hui, C., Tyshchuk, Y., Wallace, W.A., Magdon-Ismail, M., & Goldberg, M. (2012). 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. 653-656.

Kim, D., Kim, J. (2014). Public opinion sensing and trend analysis on social media: a study on nuclear power on Twitter. Int J Multimedia Ubiquitious Comput. 9:373-384.

Kisely, S. (2017). On Adjusting for Life's Confounding: Harnessing Big Data to Answer Big Problems. Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie. *62*(3), 182–185. http://doi.org/10.1177/0706743716660710

Knaak, S., Mantler, E., & Szeto, A. (2017). Mental illnessrelated stigma in healthcare: Barriers to access and care and evidence-based solutions. Healthcare Management Forum, 30, 111-116. doi:10.1177/0840470416679413.

Kohavi, R., & Foster, P. (1998). .Glossary of terms. Machine Learning. 30(2-3), 271- 274.

Koschade, S. (2006). A social network analysis of Jemaah Islamiyah: The applications to counterterrorism and intelligence. Studies in Conflict & Terrorism, 29:6, 559-575.

Kranke, D., & Floersch, J. (2009). Mental Health Stigma among Adolescents: Implications for School Social Workers. *School Social Work Journal, 34*(1), 28-42.

Lakeman, R., McGowan, P., MacGabhann, L., Parkinson, M., Redmond, M., Sibitz, I., Stevenson, C., & Walsh, J. (2012). A qualitative study exploring experiences of discrimination associated with mental-health problems in Ireland. Epidemiology and psychiatric sciences. 21(03): 271-279.

Livingston, J. (2013). Mental Illness-Related Structure Stigma: The Downward Spiral of Systemic Exclusion Final Report. Mental Health Commission of Canada.

Moreno, M. A., Christakis, D. A., Egan, K. G., Brockman, L. N., & Becker, T. (2011). Associations between displayed alcohol references on Facebook and problem drinking among college students. Archives of Pediatrics & Adolescent Medicine. 166(2):157-63.

Moses, T. (2010). Being treated differently: stigma experiences with family, peers and school staff among adolescents with mental health disorders. Social Science & Medicine, 70, 985e993.

Naslund, J. A., Aschbrenner, K. A., & Bartels, S. J. (2016). How people with serious mental illness use smartphones, mobile apps, and social media. *Psychiatric Rehabilitation Journal, 39*(4), 364-367. http://dx.doi.org/10.1037/prj0000207 Retrieved from http://web2.trentu.ca:2048/login?url=https://search.proquest.com/docview/1797262442?accountid=14391

Naslund, J. A., Aschbrenner, K. A., Marsch, L. A., & Bartels, S. J. (2016). The future of mental health care: Peer-to-peer support and social media. *Epidemiology and Psychiatric Sciences, 25*(2), 113-122. http://dx.doi.org/10.1017/S2045796015001067

Pang, B., & Lillian, L. (2008). Opinion mining and sentiment analysis. Foundations and trends in information retrieval. 2(1-2), 1-135.

Pang-Ning, T., Steinbach, M., & Kumar, V. (2005). Introduction to Data Mining, (First Edition). Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Park, M., McDonald D., & Cha, M. (2013). Perception Differences between the Depressed and Non-depressed Users in Twitter. In Proc. ICWSM 2013.

Patten et al. (2016). Major depression in Canada: what has changed over the past 10 years? *Canadian Journal of Psychiatry*, 61: 80-85. "Potentially adequate treatment" defined as "taking an antidepressant or 6 or more visits to a health professional for mental health reasons."

Paul, M., & Dredze, M. 2011. A model for mining public health topics from twitter. Technical report, Johns Hopkins University.

Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter. WWW. 695-704. 695-704. 10.1145/1963405.1963503.

Ruiz, S. L, and Lee S. (2015). Social Media as Support for Partners of Veterans With Posttraumatic Stress Disorde. Journal of Social, Behavioral, and Health Sciences. 9(1):1

Salathé, M., & Khandelwal, S. (2011). Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control. *PLoS Computational Biology*, *7*(10), e1002199. http://doi.org/10.1371/journal.pcbi.1002199

Shoppers LOVE. YOU. Run for Women Poll (2016). Online survey conducted by Environics Research.

Sinnenberg, L., Buttenheim, A.M., Padrez, K., Mancheno, C., Ungar, L., Merchant, R.M. (2017). Twitter as a tool for health research: a systematic review. Am. J. Public Health 107, 143. http://dx.doi.org/10.2105/AJPH.2016.303512a

Sloan, L. & Morgan, J., 2015. Who tweets with their Location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. PLoS One 10, e0142209. http://dx.doi.org/10.1371/ journal.pone.0142209.

Smetanin et al. (2011). The life and economic impact of major mental illnesses in Canada: 2011-2041. Prepared for the Mental Health Commission of Canada. Toronto: RiskAnalytica.

Smith, B. (2015). Mental Illness Stigma in the Media. The Review: A Journal of Undergraduate Student Research. 16: 50-63.

Statistics Canada (2017). Deaths and mortality rate, by selected grouped causes, age group and sex, Canada, 2014. CANSIM 102-0551.

Sunderland & Findlay (2013). Perceived need for mental health care in Canada: Results from the 2012 Canadian Community Health Survey – Mental Health. Statistics Canada Catalogue no.82-003-X.

Tremayne, M. (2014). Anatomy of protest in the digital era: A network analysis of Twitter and Occupy Wall Street. Social Movement Studies. 13:1, 110-126.

Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M. (2011). Election forecasts with Twitter: how 140 characters reflect the political landscape. Soc Sci Comput Rev. 29:402-418.

Pearson, Janz & Ali (2013). Health at a glance: Mental and substance use disorders in Canada. Statistics Canada Catalogue no. 82-624-X.

Pew Research Center. (2016). Social Media Update 2016*:* Facebook usage and engagement is on the rise, while adoption of other platforms holds steady. Retrieved from http://www.pewinternet.org

Ueda, M., Mori, K., Matsubayashi, T., & Sawada, Y. (2017). Tweeting celebrity suicides: Users' reaction to prominent suicide deaths on twitter and subsequent increases in actual suicides. *Social Science & Medicine, 189*, 158. Retrieved from http://web2.trentu.ca:2048/login?url=https://search.proquest.com/docview/1957218980?accountid=14391

Villanti, A. C., Johnson, A. L., Ilakkuvan, V., Jacobs, M. A., Graham, A. L., & Rath, J. M. (2017). Social Media Use and Access to Digital Technology in US Young Adults in 2016. *Journal of Medical Internet Research*, *19*(6), e196. http://doi.org/10.2196/jmir.7303.

Wang, L. (2017). Heterogeneous Data and Big Data Analytics. Automatic Control and Information Sciences. 3(1), 8-15.

Weng, J., Lim, E-P., Jiang, J., & He, Q. (2010). TwitterRank: finding topic-sensitive influential twitterers. Proceedings of the third ACM international conference on Web search and data mining, February 04-06. New York, New York, USA.

World Health Organization: Mental health action plan 2013-2020, Tech. rep.. 2013 http://www.who.int/mental_health/publications/ action_plan/en/.

Yin, A., Ray, B., Statnikov, A., & Krebs, P. (2014). Text classification for automatic detection of alcohol use-related tweets: A feasibility study. Information Reuse and Integration (IRI), 2014 IEEE 15th International Conference. 93-97.

Zaydman, M. (2017). Tweeting About Mental Health: Big Data Text Analysis of Twitter for Public Policy. Santa Monica, CA: RAND Corporation.