Using Social Media Data to Measure and Influence Community Well-Being

Ziad Matni
School of Communication & Information (SC&I)
Rutgers, The State University of New Jersey
4 Huntington St, New Brunswick, NJ 08901
zmatni@rutgers.edu

ABSTRACT
Information about our local environment is everywhere on the Internet and other information and communication technologies (ICT), especially in certain forms of social media. Our engagement and interaction with this information in our daily social lives, as individuals and groups, has grown tremendously in the past decade and continues to do so. This is an early research proposal that aims at looking beyond detection of events in the "real-world" via social media and wants to ask how we can gauge from the social media data how well individuals and groups of community-based individuals are doing. My goal is to do this by finding useful measures of social indicators, in particular community well-being and its possible relationship with the collective stress and/or tranquility of communities, in social media. Moreover, I would like to see if we can influence individuals and groups through their social networks in order to change or modify their behavior on the basis of this data. This research, then, seeks to engage in the fields of social media use, the evaluation of community well-being, and the mechanisms of influence in social networks. This project is important in helping us better understand how social media plays a role in communities and how we might increase the utility of this ubiquitous and very popular online technology.

General Terms
Measurement, Experimentation, Human factors

Keywords
social media, community well-being, social networks

1. INTRODUCTION
Social media has afforded us expanded ways in which to communicate and exchange information with each other. The rich content that it provides comes from us, its users, who share it with one another through interactive websites with integrated social networks. Social awareness streams (SAS), a special type of social media, encompass "real-time" information streams generated by Internet services like Twitter[1], Facebook[2], FourSquare[3], and Flickr[4] [41]. SAS are at the forefront of changing the framework of how our society utilizes information. They are used by many millions of people on a daily basis to help them stay connected by communicating via brief messages, typically in public or semi-public forums. These real-time technologies also yield very large amounts of data from, and about, the people communicating and their communities, both virtual [28] and geographically-localized [46]. But looking beyond events being detected, how can we gauge from the data how well individuals and groups of community-based individuals are doing in their everyday lives?

I want to find *reliable measures of community well-being* by looking at data in user-generated informational technologies, like social media/SAS (or in other instances, in personally-carried monitoring devices like smart phones, health monitors or other wearable computing devices). There are many aspects to what might be termed "community well-being": a measure of civic unrest, the health of its aggregate population, the financial stability or wealth of its aggregate population, the level of education of its population, and so on. I propose to detect some of these aspects in social media/SAS data and hope to show that when these well-being measures are low, we begin to see signs of stress and anxiety on individual and community bases. Likewise, when these measures are high, we see what I term as "tranquility" of individuals and community groups. The tranquility of a community is important for community leaders looking to assuage possible concerns from the citizenry about abnormal health and/or safety events concerns that can arise from widespread anxiety. This anxiety can be due to health epidemics, a rise in police or fire fighter activity, natural disaster occurrences and their aftermaths, or extreme weather occurrences: in short, anything that might disrupt the lives of people's daily lives.

Furthermore, I would like to see if we can *influence individuals and groups to change or modify their behavior* on

[1]www.twitter.com
[2]www.facebook.com
[3]foursquare.com
[4]www.flickr.com

the basis of this data and information we might gleam from their social networks. Once data is aggregated and analyzed, what manner of feedback via the social network might prove to be the most effective and why? The motivation is to better understand how influences through social networks can help ease or moderate the level of anxiety that people and groups of people are feeling vis-à-vis their community. Likewise, this might help us better understand if this feedback mechanism can be used effectively to agitate communities in order to give rise to community activism (e.g. get people more motivated about community-level causes/perceived threats or even global ones, like climate change)?

2. RESEARCH QUESTIONS

My research questions can be put in two groups: ones that address the issues of identifying the stress or tranquility of communities from their online social behaviors, and ones that address how we might influence the stress or tranquility of communities through modification of their online social behaviors. In order to formulate my first group of research question, I will need to identify/define three sets of categories of things that can be found in social media/SAS sources: the set of dimensions of community well-being $\mathbf{d_{CWB}}$, the set of indicators of individual stress, $\mathbf{i_S}$, and the set of indicators of individual tranquility, $\mathbf{i_T}$.

RQ1: *How do the dimensions of community well-being (set $\boldsymbol{d_{CWB}}$), as found in a social media/SAS message, relate to the indicators of stress and/or tranquility (sets $\boldsymbol{i_S}$, and $\boldsymbol{i_T}$) in the same message? Additionally, what might a wide-reaching measure of an aggregated measure or index of online community well-being from social media/SAS sources be?*

RQ2: *How can one model a standard method for measuring online community well-being that parallels a "golden standard" of community well-being (like those used in sociological or quality-of-life studies)?*

For my second group of research questions, I aim to answer:

RQ3: *How can we enact an online influence on a community's well-being based on what we might learn from the first two research questions?*

3. WHAT IS SOCIAL MEDIA?

The Internet has afforded us social interactions like never before: both with the ease to "reach out and touch someone" and, increasingly, with richer media that enhances the computer-mediated communicative experiences. These experiences are not just about sharing content, but importantly also about creating it. Moreover, I think that user-created content is the biggest differentiating factor with this new set of online media, which within this last decade of what has been termed Web 2.0, have loosely been termed "social media". Like many of its Web 2.0 cousins, I posit that social media usually meets the following criteria:

1. It operates in connected data networks, most critically, the Internet,
2. It is accessed from a myriad of computing platforms, including mobile communication devices,
3. It is inherently multimedia in nature: in other words, it not only deals with textual information, but also pictures, videos, and audio,
4. It is designed with an emphasis on its users' participation and interaction with one another,
5. It not only allows, but promotes and bolsters the creation and sharing of user-created content, both individually and collaboratively, and
6. It allows us to do new things (i.e. enables new affordances) in our social interactions, whether between individuals or groups, and in all aspects of life (i.e. when engaging in routine communication, or for entertainment purposes, or in the workplace).
7. It allows us to create content and share it based on part or the whole of one or more articulated social networks.

Social media allows users to seek information with relatively little effort. The interfaces are easy to use for navigation of the information and the information itself comes in inherently rich structures. Because of this, social media sources have an inherent advantage over more traditional collections of documents when performing information retrieval tasks, even if the quality of the content varies [2]. This new media has become an important conduit to preserve work and social connections in everyday life.

Furthermore, the flexibility of these less-bounded, spatially dispersed social networks creates a demand for more collaboration in communication and information seeking. That demand does not seem to be ebbing any time soon; for example, Facebook continues to grow and its users shows no signs of fatigue. Additionally, social media's pervasive presence in people's online activities is a large contributor to their motivations of everyday information seeking behaviors [37].

4. PAST WORK

4.1 Location-Based Social Media Data

Although social media and its use in information seeking, searching, and retrieval is a relatively new phenomenon, the combination of its ubiquity and ease-of-access, its inherent richness of information, and its natural facilitation of social networking make it a powerful (and very low-cost) emerging way to seek, share and analyze information.

The widespread popularity of social networking services on social media offers new opportunities for studying human behavior in urban settings [27, 46]. Past scholarship has examined people's activity on social media to model the dynamic borders of cities [14]. Others have observed the patterns of keywords on social media that emerge quite predictably across cities in the United States and England [42], or across several cultures [22].

In terms of methodologies used to extract patterns that made sense, the literature emphasizes that high-specificity of search terms, high-volume of collected data, and the right clustering algorithms were essential to getting good results [56, 27].

Certainly, using social media and SAS as sources for information has its challenges. This type of data is known to be noisy [41], the reliability of present geo-location tags in the data is known to be questionable, and people post in different quantities at different times, giving way to temporal fluctuations in the volume of messages produced, which can have undesirable effects on data collection [42].

4.2 Community Well-being

The concept of well-being as a social indicator has been studied extensively in sociology to aid in generally defining and measuring quality of life. While well-being may be argued to be a subjective concept, social indicators are established sets of measurement. Both social indicators (seen as "objective" guides) and subjective well-being measures are necessary to evaluate the quality of life of a society [16].

Models for measuring social indicators of well-being have been proposed in various ways, for instance in a recent report by the Aboriginal Affairs and Northern Development Canada (AANDC) [1], which conducts studies on aboriginal Canadian communities, the four main components of a community's well-being were income, education, housing, and labor force. A study by White [54] emphasizes 3 dimensions, which encapsulate the aforementioned Canadian report's components quite well: material, social, and human concerns. White et al. apply these 3 dimensions to studying communities in order to address the practical concerns of standards of living, social relations and public goods, and human capabilities and attitudes to life, respectively. To further show more nuanced detail in these components or dimensions, Miles et al. [39] propose a 'six-by-six' model for measuring community well-being in a Queensland, Australia based study. This model is composed of six dimensions of well-being (covering wealth, public safety, personal health, diversity, governance, and environment or infrastructure), with each dimension additionally made up of six indicator headings.

However, while there exist many studies that focus on dimensions of well-being, not too many of them present unified definitions. Many scholars agree on the importance of "well-being" in assessing the balance and happiness of both individuals and groups. Of the few definitions in the literature, Dodge et al.[17] offer this concise one: the stable well-being of individuals is when they have the psychological, social and physical resources they need to meet a particular psychological, social and/or physical challenge.

4.3 Behavioral Research of Influence in a Social Network

To answer my third research question (RQ3), I am considering one of two approaches: either a social networking experiment using online crowdsourcing, or a lab study where I can observe participants behavior as they interact with one another in a simulated social network.

4.3.1 Social networks

The mesh of our relationships to one another is a big part of what defines us as social beings; in fact, the eminent sociologist Georg Simmel [47] tells us that society itself is nothing more than a web of relations. Social network theory espouses the idea that a social actor's position in a network partially determines the constraints and opportunities that he or she will encounter [6]. This is why identifying and analyzing that position is useful for predicting actor outcomes, such as performance, influence, or indeed most any other social behavior. Actors in networks are always discussed in regards to the links or relationships that exist between them. These relationships are the fundamental component of network theory and distinguish network analysis from other research approaches. The theoretical concepts, the data under study, and the analysis performed is all about the relationships among the units in the studied network. The ways in which actors in a network are able to influence the flow of resources because of what manner of a central role they play amongst other actors describes the social structure of the environment [51]. We are generally not interested in what the actor might do; only that the actor is part of a social structure. Moreover, a social network is often studied as a snapshot (or series of snapshots) of the structure of relations in one particular point in time [51]. As it happens, a common criticism of social network research is that not enough attention is paid to network dynamics [52].

Usually we ascribe "strong ties" to relations between people of a core network: like members of a family, or close friends. Amongst each other, strongly tied people exhibit higher levels of intimacy, more emotional exchanges, more self-disclosure, more frequent interactions, and higher reciprocities [26]. There is a range to all of these characteristics and where they delineate between different tie strengths is open to interpretation by different scholars [26, 34, 21]. For instance, the frequency of interactions between strongly tied social actors does not seem to be as important in kinship ties as it is with other "inner-circle" people that need certain maintenance of the tie strength, such as the case of friends and work colleagues. Ties and their strength attributes can therefore vary over time, growing as people get to know each other better and declining as the reason for some strong associations reaches some conclusions [32]. Strongly tied social actors are usually self-motivated, often because of positively affective feelings for one another ("philos relationship"), to share resources with each other and usually make themselves available to one another [34].

It is easy therefore to intuitively understand "weak ties" as those between mere acquaintances, for instance. Strongly tied pairs provide high velocity paths to information already circulating in their tightly knit network, which means that these actors have access to the same resources. If they wanted new information or fresh resources, they would necessarily have to go outside their strong tie network. The "strength of weak ties", then, is that they provide connections to others outside the strong tie network along with their new resources. These weak ties can help someone generate creative ideas, find a job, or get information on the competition [24, 25, 9]. A corollary to this seems to be that when people search for informational resources, they will need to access networks beyond their strong-tie ones [31].

4.3.2 Social Capital

Each social actor has control over certain resources and when two actors are tied to one another in a social network, they have the potential to exchange these resources, be they material (e.g. funds or goods) or non-material (e.g. information). Social capital facilitates purposeful actions of actors and is not inherent in one thing or one person (the way, say, human capital is), but rather it derives its value from being. The definition of social capital is expressed by its function: at once it is the structure of contacts in a network (this is about how you can reach to actuate resources) and the resources they each hold (about who you can reach) [12]. Scholars of social networks regard the importance of social capital very highly [10].

Putnam describes two dimensions of social capital: bonding and bridging. Bonding social capital allows access to the resources from one's core network: that is, from one's strong ties, so it is usually associated with trustworthiness and social norms. Bridging social capital refers to those resources most likely accessible from outside the core network: in other words, from one's weak ties, so it is typically discussed with outcomes such as volunteering, or acquiring new information [43]. There is, unfortunately, no definitive consensus on how to best measure social capital. Worse still is that even with validated measures and techniques available, scholars are not consistent in their use although there usually is intra-scholar consistency [3]. Past efforts focused on equating social capital with levels of social trust or participation in voluntary associations [43]. Due to their sometimes controversial findings (such as the sweeping declaration of a decline in American social life), these measurement techniques have been critiqued for not having convergent validity within the measures and that these measures could not be generalized beyond their social constructs and contexts [3].

4.3.3 Influence in Social Networks

When we want to ask the question, "How do people's attitudes change in a social network setting?" it turns out that the attributes of people can only explain a small amount of the differences in their attitudes. This is because people's attitudes are mostly shaped through interpersonal processes that take place within people's social networks [19]. There are individuals that are better conduits of change in others' attitudes and behaviors and are referred to as influencers, or "influentials".

Influence is typically dispersed in a social network as described in the diffusion curve: an S-shaped curve whose y-axis is number of people in the network because once adoption starts happening, it tends to cascade rather quickly in an exponential fashion, then it plateaus [44]. Influentials are good at communicating with external change agents, have greater social participation than others, tend to be better-off financially, and are good at switching their innovative sides on and off as the occasion calls for it. Watts and Dodds found that that influenced events are actually driven by a critical mass of easily influenced individuals [53]. What makes them easily influenced is their position in their social networks. It also seems that the joint distributions of influence and susceptibility have a big say about the propagation of influential behaviors. Aral and Walker looked at both people who influence others and people who are susceptible to influence in a large data set of 1.3 million Facebook users. The authors showed that the social network structure underlies all of their results. As it turns out, influential people were less susceptible to influence themselves and that they clustered together in the network, while susceptible people did not [4].

I could posit that the influencers should show high centrality measures. The flow of how influence moves in a network typically comes down to the concept of centrality in network analysis. There are several types of centrality measures, such as degree centrality, closeness, betweenness, or eigenvector centrality [5]. Borgatti [5] cautions researchers that the types of flow processes must be first identified before the type of centrality measure is decided upon because of the different assumptions made by these different measures. Centrality is a property of a node's position in a network. The centrality of a node is, loosely speaking, about the contribution the node makes to the structure of the network [6]. The simplest measure of centrality is degree centrality, which is merely the number of tie of a given type that a node has. It can be further delineated as in-degree and out-degree centrality measures, which classifies incoming versus outgoing links to and from a node, respectively.

Degree centrality is very widely used when using measures of network characteristics (be they social, organizational, or other kinds of networks) to inform the researchers about the actors. Examples include studies that located influential people within a group with regards to their social structure [23]. Degree centrality has been criticized as not adhering to a strict definition because it does not take into account any measures of the whole network beyond the adjacent nodes, but it is nevertheless easy to calculate and popularly used. Eigenvector centrality attempts to answer that criticism and is a more sophisticated version of degree centrality, in that it is calculated as the number of nodes adjacent to (i.e. linked to) a given node, but then each adjacent node is weighed by its own centrality [6]. Examples of using eigenvector centrality measures in research includes studies wanting to uncover influential authors of emerging messages on Twitter [40] and to show academic departmental prestige is an effect of its position within networks of association and social exchange [8].

In addition to characterizing the nodes and edges of a network, one can characterize the whole network as well. Network density is a measure of cohesion of the network and offers a general picture of the network [6]. Network density is a popular measure to use because it can say much about other network characteristics like the presence of structural holes or network diversity; both of which have a negative correlation with density. Social networks with high network diversity are associated with many positive outcomes that include more cultural tolerance [13] and access to diverse information which can be helpful in finding a job [25].

4.3.4 Crowdsourcing for behavioral research

Crowdsourcing, in the context of behavioral research, is generally understood to be a job outsourced to any number of everyday people using their time to do questionnaires, take tests, and engage in experiments for the benefit of the researchers. This large set of people are usually willing to perform short and/or simple tasks for low pay. The largest provider of online crowdsourced pools of participants is Amazon's Mechanical Turk (AMT) [33, 36]. In a study aimed at finding out if network topology has an effect on contributions of cooperation between people in a social network, Suri and Watts ran an experiment using AMT to recreate these different network topologies and had participants interact as conditional cooperators. They found that network topologies did not influence cooperation dynamics, but that positive effects of cooperation of individuals in the network were indeed influential, but only to their direct neighbors. Furthermore, people increased their contributions more in response to high-contributing neighbors (the opposite was found to be true as well) [48].

5. PROPOSED METHODOLOGY

This work is still in progress and is intended as a proposal for a dissertation. As such, I am presenting a proposed methodology to pursue. For RQ1 and RQ2, I am fairly certain about the path I should take with the data collection and initial analysis, as the next 2 subsections discuss in more detail. However, for my third research question concerning an examination of how to best influence people's well-being vis-à-vis their community, I have not made a commitment as to what methodology to use. I have instead opted, for this particular forum, to explore more than one idea to tackle RQ3. For example, I would like to engage in an analysis of social networks. However, recreating the personal social network of every individual who posts relevant messages in social media in this context of community well-being is highly impractical. This is why the approaches I am entertaining are of an experimental nature. One such approach may possibly use crowdsourcing and various simulations of influence through a variety of personal network structures. Another approach may involve observing participants in a lab setting using an online tool where I can control for environmental factors and vary information exchanges online and subjective affects of the participants and observe for influence and change in subjective well-being factors.

5.1 Data Collection

As previously mentioned, I will have to create a set of keywords that describe dimensions of community well-being, $\mathbf{d_{CWB}}$, a set of keyword indicators of individual stress, $\mathbf{i_S}$, and a set of keyword indicators of individual tranquility, $\mathbf{i_T}$. The subjective (individual) and community well-being literature will be my main source for these [16, 15, 55, 39, 54, 17].

I propose collecting social media/SAS data from more than one source to lessen the effect of bias. Interesting candidates include Twitter and Reddit[5] because of their popularity and because of their relatively straight-forward application programming interface (API). Twitter has about 288M monthly users and is a heavily visited website, ranked 8th worldwide and in the United States, according to Alexa.com[6], a commercial provider of Web traffic data. Twitter allows its users to post short textual messages, or "tweets", which are up to 140 characters long. Additionally, Twitter users can annotate their messages in various ways in order to, among other things, impart additional information about themselves in the form of metadata such as names, nicknames, gender, affiliations, what their messages are about or how they might best be classified (implemented with the use of hash-tags, for example, "#Hillary2016"), when their messages were sent (using time stamps), and where they are located (which can take the form of a location name or geographic coordinates of latitude and longitude). Reddit sees about 168M users a month and is also one of the most visited websites, ranking 10th in the US, according to Alexa.com. It is a social voting site that calls itself the "voice of the internet" and is comprised of over 9,500 communities, or "subreddits". Each subreddit is independent and moderated by a team of volunteers. Many of these subreddits are community-specific and are ready repositories of local information.

The APIs that both Twitter and Reddit supply to developers and researchers allows for very large amounts of data to be collected. Collected data via these APIs, especially in the case of Twitter, may be incomplete (i.e. I may not be able to get 100% of all tweets sent out at any given moment), but the volume of data that I can acquire over time will, hopefully, be more than sufficient in terms of scale and diversity for this project.

When collecting the data, I will utilize queries made up of words and groupings of words derived from the set of dimensions of community well-being $\mathbf{d_{CWB}}$. Additionally, I will focus on collecting data from a limited set of geographical locations to target messages coming from specific local communities. Ideally, I would like these different geographic locations to be randomly selected and to reflect a fair cross-section of society, but there is an inherent limitation with a lot of social media/SAS: its use is not fully, nor equally, represented by all groups in society. As such, my expectations might best be set to run this research in more or less urban settings in North America where there are more quotidian users of social media. Additionally, the users of services like Reddit might introduce a skew to the findings, as they are overwhelmingly young (18 to 29 year old) males [18]. If I run my automated collection efforts via a customized crawler program over several days, I expect to collect at least a few tens of thousands of messages, based on prior work I have done in collecting social media data.

5.2 Coding of Social Media Data

Once collection is made, the next step is for these messages to be processed in order to remove duplicates or irrelevant returns. Following this, I will use content analysis methods to classify the messages based on keywords and keyword combinations to primarily ascertain what keywords were used, with what frequencies, and in which messages. These and other general statistical information will enable me to also compare the occurrence of these keywords against the set of keyword indicators of individual stress, $\mathbf{i_S}$, and the set of indicators of individual tranquility, $\mathbf{i_T}$.

I will develop a coding scheme to classify the collected messages. This will be done using content or thematic analysis [38, 7]. I will also recruit at least one other information professional to utilize my coding scheme and classify the messages. I will aim to get an acceptably high measure for reliability using Krippendorff's alpha [35] or Cohen's kappa [11].

With the coded data, I now have new variables representing the various keywords, their frequencies of appearance, and

[5] www.reddit.com
[6] www.alexa.com

other properties, as well as all the message identifiers and their other properties. I would first run Pearson correlations, which could give me an indication of linear association between the variables. I will also need to do some analysis of variance (ANOVA) to examine the differences between variables' means and other variations. Following that effort, the next step will be to try to fit these variables into a prediction model using linear regression (LR) and/or binary logistic regression (BLR) analysis, should there be binary variables in my data. Depending on initial findings, factorial ANOVA, and maybe other nonlinear regression techniques may need to be used [50].

In order to give some validation to the data analysis and any index calculation I might make of online community well-being, I would like to compare my findings with secondary sources. In North America, there are some governmental and non-governmental public databases with ample data on their communities. In the United States, the US Census data is made available through the Internet via a public FTP site[7]. This includes the very rich data of the American Community Survey (ACS), a continuous statistical survey that samples a small percentage of the population of the United States every year (the latest data currently available is from the 2013 survey). The ACS contains demographic, economic, financial, education, and other data on respondents. The data is made available in aggregated form (i.e. US-wide) as well as broken down by various geographic areas (state, county, county subdivision, city, school districts, to name but some of them). Canada's population census data is also available to the public online[8]. Additionally, both the United States and Canada have open databases online[9],[10] that also include data on aspects of community well-being.

5.3 Analysis of Social Network Data

For RQ3, I want to study the possible ways to influence individuals' sense of community well-being by examining their social networks. Information about social networks is best gathered by questionnaires, interviews, diaries, observations and via computer monitoring [20]. The characteristics of someone's social network can be ascertained using certain survey techniques that can give a sense of someone's bonding social capital and bridging social capital. Bonding social capital is best measured using multiple generators and several studies have used name generators for such a purpose [30, 29]. For example, a survey might do best to have two questions that are name generators. Those would ask about "important matters" and "especially significant ties". Bridging social capital is best measured with the position generator [49], a 22-item questionnaire used in prior work I have previously published [3]. The sum of the items on this scale can be used as a measure of bridging social capital, which can be an indicator of the extent of one's weak ties.

6. IMPLICATIONS

This project is important in helping us better understand how social media plays a role in communities. The implications of this research are that, on a practical level, we might be able to observe in real time (and possibly predict) disruptive community events by gauging community well-being through social awareness stream and other social media data. If this proves successful in the long term, I can imagine using similar techniques to focus on other indicators of community concerns, such as political stability, economic health and public transportation flow. This research can also increase the utility of social media platforms for understanding local communities. Moreover, it can add to local community information systems, founded on evidence-based, human-centered research and computation. From an Information Science theoretical perspective, this project might augment the everyday life information seeking (ELIS) model [45] to consider information seeking behavior in regards to local communities, not just individuals.

In conclusion, the ubiquitous nature of information on our local environment is well represented in online social media. This research will add to our existing understanding of this phenomenon as our engagement with this information, as individuals and communities, grows.

[7]http://www.census.gov/data/developers/data-sets.html

[8]http://www12.statcan.gc.ca

[9]http://open.canada.ca

[10]http://www.data.gov

7. REFERENCES

[1] Aboriginal Affairs and Northern Development Canada. The Community Well-Being Index: Well-Being in First Nations Communities, 1981-2011. Technical report, Her Majesty the Queen in Right of Canada, represented by the Minister of Aboriginal Affairs and Northern Development, Ottawa, 2015.

[2] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. Finding high-quality content in social media. In *ACM International Conference on Web Search and Data Minig (WSDM)*, page 183, New York, 2008. ACM Press.

[3] L. Appel, P. Dadlani, M. Dwyer, K. N. Hampton, V. Kitzie, Z. Matni, P. Moore, and R. Teodoro. Testing the validity of social capital measures in the study of information and communication technologies. *Information, Communication & Society*, pages 1–19, Feb. 2014.

[4] S. Aral and D. Walker. Identifying influential and susceptible members of social networks. *Science*, 337(6092):337–41, July 2012.

[5] S. P. Borgatti. Centrality and network flow. *Social Networks*, 27(1):55–71, Jan. 2005.

[6] S. P. Borgatti, M. G. Everett, and J. C. Johnson. *Analyzing Social Networks*. SAGE Publications, Thousand Oaks, CA, 2013.

[7] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[8] V. Burris. The Academic Caste System : Prestige Hierarchies in PhD Exchange Networks. *American Sociological Review*, 69(2):239–264, 2004.

[9] R. S. Burt. The Social Structure of Competition. In *Networks and organizations: Structure, form, and action*, chapter 3, pages 8–49. Harvard Univeristy Press, Cambridge, MA, 1992.

[10] R. S. Burt. The contingent value of social capital. *Administrative science quarterly*, 42(2):339–365, 1997.

[11] J. Cohen. Weighted kappa: nominal scale agreement

with provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4):213–220, 1968.

[12] J. S. Coleman. Social Capital in the Creation of Human Capital. *American Journal of Sociology*, 94(1988):S95–S120, 1988.

[13] R. R. Côté and B. H. Erickson. Untangling the Roots of Tolerance: How Forms of Social Capital Shape Attitudes Toward Ethnic Minorities and Immigrants. *American Behavioral Scientist*, 52(12):1664–1689, July 2009.

[14] J. Cranshaw, R. Schwartz, J. I. Hong, and N. Sadeh. The Livehoods Project : Utilizing Social Media to Understand the Dynamics of a City. In *The International AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 58–65, 2012.

[15] E. Diener. Subjective well-being. The science of happiness and a proposal for a national index. *The American psychologist*, 55(1):34–43, 2000.

[16] E. Diener and E. Suh. Measuring quality of life: economic, social, and subjective indicators. *Social Indicators Research*, 40(1/2):189–216, 1997.

[17] R. Dodge, A. P. Daly, J. Huyton, and L. D. Sanders. The challenge of defining wellbeing. *International Journal of Wellbeing*, 2(3):222–235, 2012.

[18] M. Duggan and A. Smith. 6% of Online Adults are reddit Users. Technical report, 2013.

[19] B. H. Erickson. The Relational Basis of Attitudes. In B. Wellman and S. D. Berkowitz, editors, *Social structures: A network approach*, chapter 5, pages 99–121. Cambridge University Press, 1988.

[20] L. Garton, C. Haythornthwaite, and B. Wellman. Studying online social networks. *Journal of Computer-Mediated Communication*, 3(June):1–32, 1997.

[21] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, pages 211–221, New York, New York, USA, 2009. ACM Press.

[22] S. A. Golder and M. W. Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1882, 2011.

[23] R. V. Gould. Power and social structure in community elites. *Social Forces*, 68(2):531–552, 1989.

[24] M. S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[25] M. S. Granovetter. *Getting a Job: A Study of Contacts and Careers*. University of Chicago Press, 1974.

[26] M. S. Granovetter. The strength of weak ties: A network theory revisited. *Sociological theory*, 1:201–233, 1983.

[27] N. Grinberg, M. Naaman, B. Shaw, and G. Lotan. Extracting Diurnal Patterns of Real World Activity from Social Media. In *The International AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 1–10, 2013.

[28] A. Gruzd, B. Wellman, and Y. Takhteyev. Imagining Twitter as an Imagined Community. *American Behavioral Scientist*, 55(10):1294–1318, 2011.

[29] K. N. Hampton. Comparing Bonding and Bridging Ties for Democratic Engagement. *Information, Communication & Society*, 14(4):510–528, June 2011.

[30] K. N. Hampton, L. F. Sessions, and E. J. Her. Core Networks, Social Isolation, and New Media. *Information, Communication & Society*, 14(1):130–155, Feb. 2011.

[31] C. Haythornthwaite. Social network analysis: An approach and technique for the study of information exchange. *Library & information science research*, 18(4):323–342, 1996.

[32] C. Haythornthwaite. Strong, Weak, and Latent Ties and the Impact of New Media. *The Information Society*, 18(5):385–401, Oct. 2002.

[33] J. Howe. The Rise of Crowdsourcing. *Wired*, 14(06):1–5, 2006.

[34] D. Krackhardt. The strength of strong ties: The importance of Philos in organizations. *Networks and organizations: Structure, form, and action*, 216(239), 1992.

[35] K. Krippendorff. *Content Analysis: An Introduction to Its Methodology*. SAGE Publications, Thousand Oaks, CA, 2nd edition, 2004.

[36] W. Mason and S. Suri. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, 44(1):1–23, Mar. 2012.

[37] Z. Matni and C. Shah. For the Love of Information: Motivations and Affective Dynamics of Surfing the Web for Pleasure. In *Proceedings of the 77th ASIS&T Annual Meeting*, Seattle, 2014. ASIS&T.

[38] M. Miles and A. M. Huberman. *Qualitative Data Analysis: An Expanded Sourcebook*. SAGE Publications, Thousand Oaks, CA, 2nd edition, 1994.

[39] R. L. Miles, L. Greer, D. Kraatz, and S. Kinnear. Measuring Community Wellbeing: A Central Queensland Case Study. *Australasian Journal of Regional Studies*, 14(1):73–93, 2008.

[40] M. Naaman, H. Becker, and L. Gravano. Hip and trendy: Characterizing emerging trends on Twitter. *Journal of the American Society for Information Science and Technology*, 62(5):902–918, 2011.

[41] M. Naaman, J. Boase, and C.-h. Lai. Is it Really About Me ? Message Content in Social Awareness Streams. In *CSCW '10: Proceedings of the 2010 ACM Conference of Computer Supported Cooperative Work*, pages 189–192. ACM, 2010.

[42] M. Naaman, A. X. Zhang, S. Brody, and G. Lotan. On the Study of Diurnal Urban Routines on Twitter. In *The International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2012.

[43] R. D. Putnam. *Bowling Alone*. Simon & Schuster, New York, New York, USA, 2000.

[44] E. M. Rogers. Diffusion Networks. In *Networks in the knowledge economy*, chapter 8, pages 300–364. 2003.

[45] R. Savolainen. Everyday life information seeking: Approaching information seeking in the context of âĂIJway of lifeâĂİ. *Library & Information Science Research*, 17:259–294, 1995.

[46] R. Schwartz, M. Naaman, and Z. Matni. Making Sense of Cities Using Social Media: Requirements for Hyper-Local Data Aggregation Tools. Technical Report Ucl, 2013.

[47] G. Simmel. *On Individuality and Social Forms*.

University of Chicago Press, Chicago, 2011.

[48] S. Suri and D. J. Watts. Cooperation and contagion in web-based, networked public goods experiments. *PLoS ONE*, 6(3), 2011.

[49] M. Van Der Gaag, T. a.B. Snijders, and H. Flap. Position generator measures and their relationship to other social capital measures. *Social capital: An international research program*, 2008.

[50] R. Warner. *Applied Statistics: From Bivariate Through Multivariate Techniques*. SAGE Publications, 2012.

[51] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, MA, Mar. 1994.

[52] D. J. Watts. *Six degrees: The science of a connected world*. WW Norton & Company, New York, 2003.

[53] D. J. Watts and P. S. Dodds. Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*, 34(4):441–458, 2007.

[54] S. C. White. Analysing wellbeing: a framework for development practice. *Development in Practice*, 20(2):158–172, 2010.

[55] J. Wiseman and K. Brasher. Community wellbeing in an unwell world: trends, challenges, and possibilities. *Journal of public health policy*, 29(3):353–66, Sept. 2008.

[56] J. Yang and J. Leskovec. Patterns of Temporal Variation in Online Media Categories and Subject Descriptors. In *ACM International Conference on Web Search and Data Minig (WSDM)*, pages 1–13, 2011.