Fall 12-5-2013

Attributing Meaning to Online Social Network Analysis for Tailored Socio-Behavioral Support Systems

Sahiti Myneni

Attributing Meaning to Online Social Network Analysis for Tailored Socio-Behavioral Support Systems

By

Sahiti Myneni, MSE

APPROVED:

Trevor Cohen, MBChB, PhD, Chair

Sriram Iyengar, PhD

Jiajie Zhang, PhD

Kayo Fujimoto, PhD

Date approved: 11/1/13

Attributing Meaning to Online Social Network Analysis for
Tailored Socio-Behavioral Support Systems

A
Dissertation

Presented to the Faculty of
The University of Texas
Health Science Center at Houston
School of Biomedical Informatics
in Partial Fulfilment of the Requirements for the Degree of

Doctor of Philosophy

By

Sahiti Myneni, MSE

University of Texas Health Science Center at Houston

2013

Dissertation Committee:

Trevor Cohen, MBChB, PhD[1], Advisor
Sriram Iyengar, PhD[1]
Jiajie Zhang, PhD[1]
Kayo Fujimoto, PhD[2]

[1]The School of Biomedical Informatics
[2]The School of Public Health

Acknowledgements

My social network has played a tremendous role in facilitating the completion of my dissertation. I am deeply indebted to my mentors, collaborators, friends, and family who provided me the encouragement critical to completing this research.

Special thanks to my primary mentor Trevor Cohen for his brilliant insight, dedication, and support. I would also like to thank my committee members: Sriram Iyengar, Jiajie Zhang, and Kayo Fujimoto for their expert guidance. I was also very lucky to have interacted with many other faculty, staff and students at The University of Texas School of Biomedical Informatics.

I am also very grateful to the faculty, staff, and fellows involved in the Innovation for Cancer Prevention Research Training Program at The University of Texas School of Public Health. In particular, I would like to thank Patricia Mullen for her guidance and counsel. Special thanks also to Nathan Cobb (MeYouHealth), Ashutosh Sabharwal (Department of Electrical Engineering, Rice University) for their insightful mentoring.

I owe a great debt to my parents, Karunasuhasini and Sreepathy Myneni, who have provided me with so much motivation and nurture. I take this opportunity to thank my family for supporting me in all the endeavors I embark upon. Above all, I am deeply grateful to my husband Dinesh Gottipati, for his undying optimism, endless love, unwavering confidence, and compassion.

Abstract

Ubiquitous online social networks provide us with a unique opportunity to deliver scalable interventions for the support of lifestyle modifications in order to change behaviors that predispose toward cancer and other diseases. At the same time these networks act as rich data sources to inform our understanding of end-user needs. Traditionally, social network analysis is based on communication frequency among members. In this work, I introduce communication content as a complementary frame for studying these networks.

QuitNet, an online social network developed to provide smoking cessation support is considered for analysis. Qualitative coding, automated content analysis, and network analysis were used to construct QuitNet sub-networks based on both frequency and content attributes. This merging of qualitative, quantitative, and automated methods expands the depth and breadth of existing network analysis techniques thereby allowing us to characterize the nature of communication among network members. First, grounded-theory based qualitative analysis provides a granular view of the QuitNet messages. Using automated text analysis, the communication links between network members were divided based on the similarity of the content in the exchanged messages to the identified themes. This automated analysis allowed us to expand the otherwise prohibitively labor-intensive qualitative methods to a large data sample using minimal time and resources. The follow-up one-mode and two-mode network analysis allowed us to investigate the content-specific communication patterns of QuitNet members.

Qualitative analysis of the QuitNet messages identified themes ranging from "Social support", "Progress", and "Traditions" to "Nicotine Replacement Therapy (NRT) entries" and "Craves". Automated annotation of messages was achieved by using a distributional approach incorporating distributional information from an outside corpus into a model of the QuitNet corpus to generate vector representations of messages. A k-nearest neighbour approach was used to infer themes relating to each message. The recall and precision measures indicate that the performance of the automated classification system is 0.77 and 0.71 for high-level themes. The average agreement of the system with two human raters for high-level themes approached the agreement between these human coders for a subset of 100 messages suggesting that the system is a reasonable substitute for a human rater. Subsequent one-mode network analysis provided insights into different theme-based networks at population level revealing content-specific opinion leaders. Two-mode network analysis allowed us to investigate the content affiliation patterns of QuitNet users and understand the content-specific attributes of social influence on smoking abstinence.

These studies provide insights into the nature of communication among members in a smoking cessation related online social network. Ability to identify critical nodes and content-specific network patterns of communication has implications for the development and maintenance of support networks for health behavior change. Analysis of the frequency and content of health-related social network data can inform the development of tailored behavioral interventions that provide persuasive and targeted support for initiating or adhering to a positive behavior change.

Vita

2007……………………………………….....B.E., Electronics and Communication Engineering, Osmania University, India

2009……………….....................................MSE, Electrical Engineering, Arizona State University, AZ

2009-11…………………………………….Research Fellow, School of Biomedical Informatics, University of Texas Health Science Center at Houston, TX

2011 to present...Pre-doctoral Fellow, School of Public Health, PhD Student, School of Biomedical Informatics University of Texas Health Science Center at Houston, TX

Selected Publications

1. Myneni, S., Cobb, N., Cohen T. (2013, accepted). Finding meaning in social media: Content-based social network analysis of QuitNet to identify new opportunities for health promotion. *Medinfo2013.* Copenhagen, Denmark (**IMIA Best Student Paper)**

2. Myneni, S., Iyengar, S., Cobb, N. K., & Cohen, T. (2013). Identifying persuasive qualities of decentralized peer-to-peer online social networks in public health. In *Persuasive Technology* (pp. 155-160). Springer Berlin Heidelberg.

3. Myneni, S., McGinnis, D., Cohen, T., Almoosa, K., Patel, B., Patel, V.L. (2013, accepted). Effective use of clinical decision support in critical care: using risk assessment framework for evaluation of a computerized weaning protocol. *Special Issue of Annals of Information Systems on Healthcare Informatics.*

4. Myneni, S., Patel, V.L. (2010). Organization of biomedical data for collaborative scientific research: a research information management system. *International Journal of Information Management.* 30; 256-64.

Field of Study

Health Informatics

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Existing epidemiological evidence indicates that modifiable risky health behaviors are placing a major socio-economic burden on global human health and wellness (WHO, 2009b, 2011a). According to existing studies, understanding human behavior in real-time settings is essential to improve public health outcomes with respect to these risky behaviors (Centola, 2013; Shiffman, Stone, & Hufford, 2008). Technological advances and ubiquitous internet presence offer potentially valuable datasets in the form of electronic traces of the activities of online social communities, which may help to understand intra- and inter-individual behavioral intricacies. Given the ubiquity of digital communication technologies, most communication events that an individual has with his/her social contacts are being digitally recorded, thus offering scientific community an unprecedented opportunity to understand human behavior at individual and population level. On the other hand, it is important to note that analyzing network data requires methods that can a) enable multiple analytical dimensions from socio-behavioral, technological, and network science perspectives, and b) scale with the social media data deluge (e.g. 175 million tweets every day in 2012 (Gruzd, Wellman, & Takhteyev, 2011; Hagan, 2011). Such multi-disciplinary analytics provides us with capabilities to understand user needs at a comprehensive level. Research studies analyzing network data to date have ignored the content of communication, focusing instead on the frequency of communication events between social actors. However, most behavior change theories propose the use of specialized content to stimulate and support individuals in accordance with the nature and stage of the change (Abraham & Michie, 2008; Glanz, Rimer, & Viswanath, 2008). Interestingly, the findings of contemporary work on social media data do not address the fundamental concerns of these theories.

In this research, I seek to develop an empirically grounded model of the mechanisms underlying behavior change at individual-level and community-level based on the communication between users of QuitNet (www.QuitNet.com), the first online social network designed to support smoking cessation while considering both communication content and frequency. I anticipate that the insights gained from this research effort will enhance our understanding of behavior change, with implications for the design of interventions to help promote and sustain healthy behavior changes that can help individuals manage a variety of health conditions.

The focus of this research is on applying qualitative, computational, automated text analysis techniques, and network-based methods in analyzing the communication patterns underlying microstructure of human behavior in real-time environments. The practical contributions of this research include development of analytic tools for public health professionals and social science researchers to effectively analyze online network content and enabling design of novel persuasive strategies that form the basis for tailored socio-behavioral consumer health interventions. There are three major components to this work: first, studying human communication within user-generated data in QuitNet using qualitative methods; second, applying automated informatics and quantitative network modeling techniques to identify patterns of communication pertinent to behavior change using the social network content posted by the users and third, utilizing the identified patterns to develop support strategies that promote and sustain health behavior changes in individuals while harnessing social influence.

Firstly, I analyze the message content of QuitNet using the qualitative techniques of grounded theory (Glaser & Strauss, 1967; Strauss & Corbin, 2008). Open coding, axial coding, and constant comparisons are employed to arrive at concepts and themes that capture the nature of QuitNet communication in terms of behavioral, inter-personal, and individualistic concepts indicated in the messages. The derived communication themes are interpreted in light of the existing behavior change theories such as Social Cognitive Theory (Bandura, 1986, 2000) and the Trans-theoretical Model of Change (DiClemente & Prochaska, 1998; J. O. Prochaska & Velicer, 1997). This analysis has allowed me to gain a deeper understanding of the behavior change process that an individual undergoes, while embedded within a social group, to cease smoking. This is also the first study that has attempted to identify theoretical consistencies in the world of virtual communication, which gives us a sense of the validity of existing theories in the context of digital health behavioral interventions and support systems.

Secondly, in this research I have employed recent developments in automated text analysis to measure the extent to which key concepts of interest are expressed within text-based messages between users of an online social network, regardless of the specific terms used to express these concepts at the surface-level. Latent Semantic Analysis , a method of distributional semantics (Cohen & Widdows, 2009), has been used in conjunction with machine learning algorithms to derive a measure of relatedness between a given message and the previously identified QuitNet concepts (using qualitative methods) to estimate the distribution of different types of content across QuitNet. A novel distributional approach incorporating distributional information from an outside corpus is developed, and evaluated for the accuracy of its semi-automated annotation of messages in the corpus. The use of automated techniques facilitates scaling up the applicability of qualitative analysis to large-scale social media datasets. Consequently, we have developed a novel approach to structure a network by combining the content of the communication with existing network analysis methods. Consequently, one-mode networks are constructed with users linked if they have exchanged messages related to the themes derived from qualitative analysis. Two mode network models, which have been previously used to understand network patterns underlying adolescent smoking behavior (Fujimoto, Unger, & Valente, 2012) have also been used to identify social ties related to themes of interest. In this way, content-specific patterns have been examined to characterize opinion leaders and network nodes that are likely to exert an influence on behavior. In summary, this component of my research enables the analysis of social network structure of QuitNet interactions as it relates to the communication of smoking cessation-related ideas.

Finally, I have attempted to leverage the content-inclusive network analysis of QuitNet to engineer novel strategies that enable tailoring network and standalone interventions to user needs. For example, the identification of those network members who are key players in providing relapse assistance and motivation can help us make the right connections with users discussing about craves thus improving network assistance to its members. Such research strategies will mediate the origination and refinement of a new generation of translational interventions in public health and behavior science. The proposed strategies can enable us to efficiently channel meaningful and critical behavioral health-related information to members of online social communities such as QuitNet. Possible implementation architectures for the proposed strategies have also been discussed in this research in the context of face-to-face and digital platforms, and at individual and population levels.

In summary, these research studies inform our understanding of the ways in which health behaviors are socially disseminated and influenced. The insights gained from these investigations should generalize to the propagation of health-related behaviors in other conditions such as diabetes management, hypertension, and safe sex practices. Consequently, this research will provide much-needed empirical support for the development of behavior change interventions that transcend the limited ability of methods that target individuals to impact on health behaviors that are embedded in a network of social influence. Aspects of the proposed methods and resulting semantic models will generalize to the study of other health-related behaviors, as well as to the study of non-health-related online communities. The socio-behavioral and cognitive models of behavior change should be sufficiently generic to be tested with other related behaviors in the context of social sciences, behavioral interventions, prevention research, and lifestyle assisting technology development. The use of methods from distributional semantics not only scales up the applicability of qualitative methods, but also facilitates the inclusion of content in network-based models of social influence.

Chapter 2: Literature Review

In 2011, Wal-Mart introduced a punitive approach to make its employees more responsible for their well-being. Under this program, employees who are smokers were asked to pay higher share of the insurance premium (which can amount to around $2000 a year) than their co-workers who meet the company's health goals. The New York Times dubbed this move as "The Smokers' Surcharge" and "more stick, less carrot" approach. While health care organizations such as the American Cancer Society and the American Heart Association were against these acts, the proponents of these programs mention the economic and social burden placed by these unhealthy behaviors on society as a legitimate reason to impose hard rules. The degree of media attention devoted to this incident highlights the lack of consensus in our understanding of how best to alter harmful behaviors – both from a scientific and a public policy perspective. In this chapter, I describe the effect of health behaviors on global human wellness, and provide an overview of behavioral interventions over the past few decades. Later in the chapter, I describe the role of social relationships on these health behaviors and how we can utilize the connected media of the 21st century to better understand human behavior. Finally, I conclude the chapter by summarizing the current state of methods available for analysis of online social network data.

2.1. Health Behaviors and Human Wellness

Preventable non-communicable diseases caused around 63% of global deaths in 2008 (Alwan et al., 2010). Unhealthy and modifiable behaviors (e.g. smoking, physical inactivity, poor diet) are the major risk factors for these conditions. World Health

Organization (WHO) projected that lifestyle-induced illnesses such as hypertension, diabetes, and cardiovascular diseases will be responsible for a significantly increased total number of deaths between 2010 and 2020 (WHO, 2011a). Examples of unhealthy behaviors include smoking, physical inactivity, poor diet, and alcohol consumption. These behaviors contribute toward 835,000 deaths in the United States annually (Mokdad, Marks, Stroup, & Gerberding, 2004) and are associated with an increased risk of premature mortality and chronic diseases such as hypertension, diabetes, stroke and cancer. These behavioral risk factors are present in about 80% of coronary heart disease and cerebrovascular disease globally (Ezzati, Lopez, Rodgers, & Murray, 2004). They lead to four key metabolic or physiological changes: raised blood pressure, weight gain culminating in obesity, hyperglycemia, and hyperlipidemia. In addition, some of these risk factors are highly relevant to the prevention of cancer of the lung and a number of other cancer sites (tobacco smoking), and both breast and colorectal cancer (unhealthy diet, overweight and physical inactivity).

Of these behaviors, smoking is the leading cause of death globally. According to a recent report from the WHO, there are around one billion tobacco users around the world (WHO, 2011b). Approximately one person dies every six seconds due to tobacco-related illness and this accounts for one in 10 adult deaths(WHO, 2009b). Tobacco use is the major risk factor for non-communicable diseases including cancers, cardiovascular diseases, respiratory diseases such as asthma, and diabetes (Health & Services, 2004). It is also a major contributor to social inequalities in mortality worldwide (WHO, 2011b). A recent study published in the Lancet survey of 16 countries that are home to a total of three billion people found that 48.6 percent of men and 11.3 percent of women are tobacco users (Giovino et al., 2012). In the United States alone, smoking causes around 443,000 preventable deaths and places a 200 billion dollar economic burden on the economy each year (Siegel, Naishadham, & Jemal, 2012). Given that these behaviors are modifiable, several public health policies have been implemented to reduce smoking-related deaths. Several efforts have been made to curb smoking by promoting evidence-based cessation programs, levying high taxes, sponsoring mass media anti-tobacco advertisements, and banning public smoking (Fiore, 2000). A brief review of the interventions utilized to promote smoking cessation is provided in the following section.

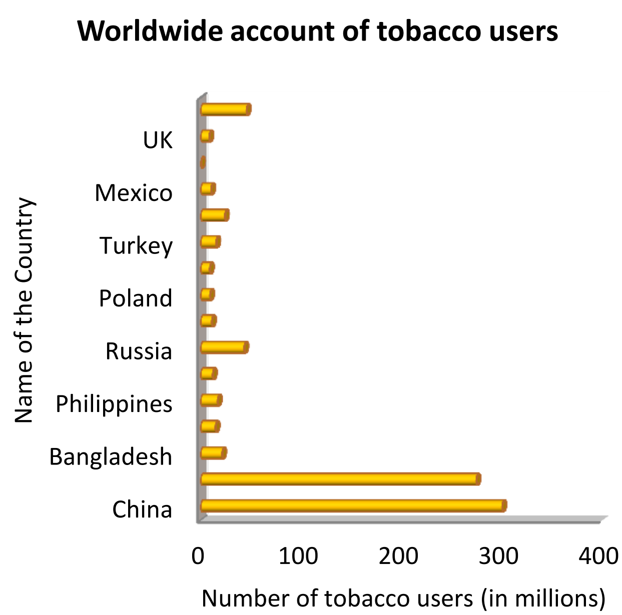


Figure 1. Number of tobacco users in the UK, USA, and 14 GATS countries, GATS=Global Adult Tobacco Survey (WHO, 2009a)

2.2 Overview of Interventions for Smoking Cessation

Simple advice from doctors during routine care, behavioral and psychological interventions such as in-person and group counseling for motivated smokers, structured therapy from clinicians, self-help materials, nicotine replacement therapy, and other pharmaceutical interventions are the primary intervention categories for smoking cessation. A recent Cochrane review examining the effectiveness of each of these interventions found that advice from doctors, structured interventions from nurses, and individual and group counseling are effective interventions (Lancaster, Stead, Silagy, & Sowden, 2000; Stead & Lancaster, 2005). According to the review, while generic self-help materials are no better than brief advice, personalized self-help materials are more effective than standard materials. With the rise of technology, a variety of computer-based interventions have been designed. Walters and colleagues provide a comprehensive review of smoking prevention or intervention programs that use a computer or the internet to calculate or deliver the intervention (Walters, Wright, & Shegog, 2006). The interventions were classified into the following major categories:

1) **First generation**: Participants have no direct contact with computer. They answer a paper-based screening questionnaire, which is fed to computer software to derive feedback. This feedback is then mailed to the participants

2) **Second generation**: Participants interact directly with the system, and get immediate feedback from the program.

3) **Third generation**: Adaptive programs that provide iterative feedback to the user.

The systematic review revealed mixed results in terms of real-world implementation of these interventions (Walters et al., 2006). Research suggests the use of explanatory models of behavior change, that are theoretically- and empirically-grounded, can align the design and evaluation of smoking cessation interventions with important factors underlying the adoption or rejection of a given behavior (Revere & Dunbar, 2001; Rhodes, Fishbein, & Reis, 1997; Skinner & Kreuter, 1997; Walters et al., 2006). Smoking cessation interventions have also been implemented in the context of mobile technologies to take advantage of the ubiquity of those devices in supporting a healthy behavior change (Heron & Smyth, 2010; Whittaker et al., 2009). Mobile platform is seen as a key enabling technology for the delivery of ecologically valid interventions that provide real-time support (Balch et al., 2004; Heron & Smyth, 2010; S. Shiffman et al., 2008; Vuckovic, Polen, & Hollis, 2003). More information on ecological strategies is provided subsequently in the chapter. Similar to web-based interventions, the evaluation of cell phone based smoking cessation programs is also limited (Whittaker et al., 2009). Self-reported smoking status was often used as an outcome measure to evaluate the effectiveness of a smoking cessation intervention. However, studies conducted to assess the reliability of self-reported cessation data indicated over-reporting of quitting (Free et al., 2009; Rodgers et al., 2005). Objective measurement approaches such as measurement of exhaled carbon monoxide and salivary cotinine have been used in several trials but in cases where anonymity is critical, obtaining biomedical data is deemed as a risk of therapy abandonment (Whittaker et al., 2009). In addition to improved accuracy, objective measurement of health behaviors such as smoking, physical activity, and dietary intake can play a vital role in patient empowerment, self-efficacy, and counseling.

2.3. Evidence-Based Strategies in Smoking Cessation Interventions

2.3.1 Personalized and Custom Tailored Support

Context-dependent personalization of messages is often seen as an important technique to improve the levels of persuasion and behavior change. Lustria et al reviewed computer-tailoring in a series of health interventions delivered via the web (Lustria, Cortese, Noar, & Glueckauf, 2009). Computer-tailoring enhances the creation of tailored messages by facilitating the collection and assessment of individual data and then using evidence-based decision rules to create strategic health messages (Fernandez-Luque, Karlsen, & Bonander, 2011; Kreuter, 2000). For example, Cobb et al. conducted a survey study evaluating the delivery of tailored content using the publicly available version of QuitNet, a smoking cessation website (Cobb, Graham, Bock, Papandonatos, & Abrams, 2005). Personalized program content was provided to the participants by providing a “quit calendar” (with coping strategies) based on the individual’s quit date and a list of steps for smoking cessation. Tailored content also focused on gender, age, quit history, and medication status. Results indicated that sustained use of QuitNet, especially use of the support tools (such as the calendar), had positive effects on abstinence rates. As Lustria and colleagues point out, the interventions designed for long-term discretionary use often require sophisticated tailoring techniques - acquisition, analysis, and visualization display of user data is essential to help users achieve their desired healthy lifestyle (Lustria et al., 2009).

2.3.2 Social Communities

The influence of social relationships and social support on health behaviors has been well documented (Heaney & Israel, 2002; Umberson & Montez, 2010). While associations between social relationships and health are complex and not necessarily causal in nature (Lyons, 2011), evidence suggests that the positive health enhancing effects of social relationships can be used for the promotion of healthy behaviors (Christakis & Fowler, 2008; Cobb, Graham, & Abrams, 2010; Heaney & Israel, 2002). Several community- based interventions at the population level have been developed and evaluated in the context of smoking cessation. Such programs usually target certain populations that are characterized by gender, age, culture, and so forth. Examples include the Minnesota (Perry, Kelder, Murray, & Klepp, 1992), Pawtucket (Carleton, Lasater, Assaf, Feldman, & McKinlay, 1995), Stanford Heart Health (Killen et al., 1989), and COMMIT programs (Royce, Hymowitz, Corbett, Hartwell, & Orlandi, 1993), which have the potential to intervene at population levels. Church-based and school-based cessation programs have also been implemented over the past decade (Pallonen et al., 1998; Prokhorov et al., 2008; Voorhees et al., 1996). The mechanisms used to engage the population in these cessation interventions vary from counselling to computer-based animations.

With tremendous growth in both electronic and mobile health, the number of “virtual” peer-to-peer community venues and online social networks is increasing. These platforms allow users to share experiences, ask questions, or provide emotional support and self-help advice to one another. Cobb et al provide a detailed analysis of QuitNet, the first online evidence-based smoking cessation program, which has been in existence for the past 14 years (Cobb et al., 2005). Evaluation of QuitNet indicated that participation in the online community was strongly correlated with abstinence, and that individuals who used any aspect of QuitNet's online community (i.e., forums, chat rooms, internal e-mail) were more than three times more likely to be abstinent using 7-day point prevalence abstinence, and four times as likely to be continuously abstinent for 2 months or longer than individuals who did not participate in the community (Cobb et al., 2005; Graham, Cobb, Raymond, Sill, & Young, 2007). However, participation in the social network remained uncommon. Structural analysis conducted on QuitNet revealed certain crucial concepts related to 'Social Integrators' –who played a key role in the process of assimilation of new members into the network (Cobb et al., 2010).

2.3.3 Stage-based and Ecological Momentary Interventions

Stage-based interventions views behavior change such as quitting smoking as a continuum, rather than discrete transition from smoking to non-smoking (Prochaska et al., 2005). These interventions are tailored to meet the needs of the user based on his/her readiness to quit. Stage-based approaches to behavior change have received widespread approval given the practical utility of the model and individualized plans improve user engagement and adherence (Bunton, Baldwin, Flynn, & Whitelaw, 2000; Dijkstra, De Vries, & Bakker, 1996; Riemsma et al., 2003), The effectiveness of these stage-based strategies for behavior changes has been widely studied in the literature (Adams & White, 2005; West, 2005).

Ecological Momentary Interventions (EMIs) are another kind of intervention speciality provided to people during their everyday lives in natural settings (Shiffman et al., 2008). Therefore, these interventions are ecologically valid and are provided at specifically identified moments in everyday life, allowing EMI to provide real-time support in the real world (e.g., a person participating in a smoking cessation intervention receives a text message on her mobile phone with tips for dealing with craving during a time when he/she typically smokes a cigarette) (Heron & Smyth, 2010; Shiffman et al., 2008). One particularly appealing aspect of EMI is that the content and timing of the intervention can be custom-tailored to patients. By integrating the accessibility and ubiquity of mobile technology, EMI can be developed to support positive health changes in real-time and real-setting depending on user needs (Cohn, Hunter-Reel, Hagman, & Mitchell, 2011; Freedman, Lester, McNamara, Milby, & Schumacher, 2006). EMI strategy has been used for studying human behaviors such as smoking, alcoholism, physical activity (Shiffman et al., 2007; Smyth & Stone, 2003; Wenze & Miller, 2010).

It is important to note that all of the aforementioned strategies can be used alone or be combined with one another. In the next section, I provide a brief description of other strategies used for behavior change and consumer support in disciplines outside smoking cessation including, but not limited to, general health and marketing.

2.3.4 Persuasive Social Gaming Techniques

Use of computers to trigger actions that initiate and/or sustain health-related behavior change has been shown to be effective (Fogg, 2002; Looije, Neerincx, & Cnossen, 2010; Purpura, Schwanda, Williams, Stubler, & Sengers, 2011; Toscos, Faber, An, & Gandhi, 2006). Consolvo et al outlines strategies for design of persuasive technologies that help people willing to change their behavior in support of the desired lifestyle (Consolvo, McDonald, & Landay, 2009). Some examples of support technologies that incorporate these design principles include Breakaway (Jafarinaimi, Forlizzi, Hurst, & Zimmerman, 2005) and Fish 'n' Steps (Lin, Mamykina, Lindtner, Delajoux, & Strub, 2006).

Social gaming is another strategy that can be used in combination with social networks for behavior modification (Baranowski, Buday, Thompson, & Baranowski, 2008). It involves strategies that encompass individuals' actions which measure progress at individual and group levels are being used to trigger and maintain certain healthy behavior changes. Figure 2 shows a persuasive data visualization that maps a player's daily foot step count to the growth and activity of an animated virtual character, a fish in a fish tank. The "Fish'n'Steps" (Lin et al., 2006) application links the size of a fish to the step count. Some fish tanks also include other players' fish, to create an environment of both cooperation and competition. While such social tactics are prominently used for general wellness and physical activity tracking, these methods might lead to better outcomes in smoking cessation interventions as well.

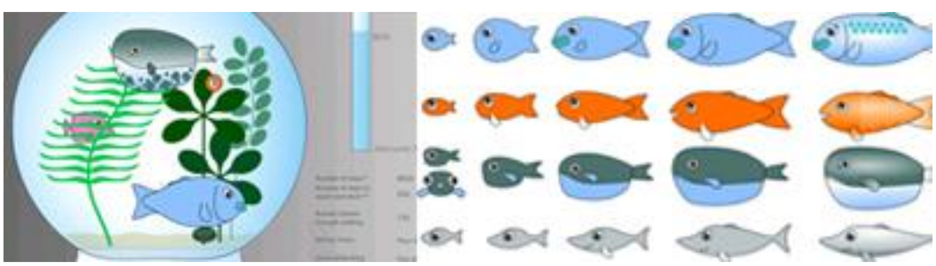


Figure 2. Social gaming strategy for increasing physical activity (adapted from Lin et al.)

2.3.5 Care Provider-Patient Communication

Clinician-patient communication plays a major role in behavior change treatments (Street Jr, Makoul, Arora, & Epstein, 2009). Patients trying to cease smoking, a highly relapse-prone behavior, were often successful when they were well connected to their health advisor (Bauman, Fardy, & Harris, 2003; Campbell, Hotchkiss, Bradshaw, & Porteous, 1998). Clinician-patient communication can contribute to improved health through at least seven ‘pathways’- access to needed care, increased patient knowledge and shared understanding, enhancing therapeutic alliances (among clinicians, patient, and family), enhancing emotional self-management, activating social support and advocacy resources, increasing the quality of medical decisions, and enabling patient self-efficacy and empowerment (Lawson & Flocke, 2009; Street Jr et al., 2009). Rather than looking at smoking cessation intervention as a treatment course, patients may be empowered with the use of a collaborative solution where they work along with their clinicians to achieve a desired behavior change.

2.4 Social relationships and Human Health Behaviours

Despite the proliferation of anti-smoking sentiment and restrictive smoking policies, the prevalence of smoking among adults has remained relatively stable during the past five years at approximately 20% of the adult population (Fiore, 2000; Mariolis et al., 2006). Evidence suggests that the majority of smokers (around 70%) wish to quit, and that 52% do attempt to quit each year (Centers for Disease Control and Prevention CDC, 2011a, 2011b). However, long-term rates of cessation were significantly lower and people willing to quit smoking often relapse (Fiore, 2008). Incorporating and adhering to certain health-related behavior changes such as smoking cessation is a challenging task and often requires a range of support mechanisms to elicit and nurture such a change. Even after making a decision to quit, people are prone to relapse for a variety of reasons, including nicotine addiction, and smoking continues to be a chronic relapsing condition (Center for Disease Control and Prevention CDC, 2013; Lancaster & Stead, 2005; Tim Lancaster et al., 2000; Stead & Lancaster, 2005). Research suggests that an individual's social relationships are the primary factor in the adoption and sustenance of healthy behaviors (Berkman & Glass, 2000; Blanchard, Albrecht, Ruckdeschel, Grant, & Hemmick, 1995; Burg & Seeman, 1994). For example, Christakis and Fowler's analysis of the Framingham dataset shows an association between an individual's social network behavior and the likelihood that a person will quit smoking (Christakis & Fowler, 2008). For example, smoking cessation by a spouse, sibling or friend decreased a subject's chance of smoking by 67%, 25% and 36% respectively as compared with having no social contacts who attempted to quit.

Qualitative analysis of in-person social networks identified several aspects of social support and their effects on behavior change (Chuang & Yang, 2012; Chuang & Yang, 2010; Hwang et al., 2010; Leimeister, Schweizer, Leimeister, & Krcmar, 2008; Yang, Tang, Huang, & Unger, 2011). House defines four kinds of social support- (i) emotional support: expressions of empathy, love, trust, and caring; (ii) instrumental support: tangible aid and service; (iii) informational support: advice, suggestions, and information; (iv) appraisal support: information that is useful for self-evaluation (House, 1981). The positive effects of social support on coping and recovery practices are well documented in the literature (Heaney & Israel, 2002; Thoits, 1995). In contrast to the positive effects of social networks, peer group studies on smoking behavior revealed that an individual's smoking status is associated with having friends who are smokers (Alexander, Piazza, Mekos, & Valente, 2001).This result is an example of homophily, which is defined by the principle that a contact between similar people occurs at a higher rate than among dissimilar people (McPherson, Smith-Lovin, & Cook, 2001). The concept of homophily might have been a result of social influence induced by observational learning as posited by Social Cognitive Theory or simple selection of social contacts based on common attitudes and beliefs. Several observational studies have shown the effects of social constructs such as social influence, selection, norms, and consequences on an individual's involvement in risky behaviors (Valente, 2010). Studies conducted on social relationships have shed light on peer influence, affiliation influence, and positional influence. Peer influence is a form of direct influence that is explicitly based on direct friendship relations. Positional influence defines a form of influence that is exerted on an individual as a result of their occupying a central position in a social network (Freeman, 1979). Affiliation-based influence takes into account indirect sources of influence including inclusion in or organized group activities/events (Fujimoto et al., 2012). Therefore, it is evident that social networks are double-edged swords that can have positive and negative influences on health behaviors.

2.5 Health-related Online Social Communities

In recent years, the penetration of online social media into everyday lives has been astonishing. Around a billion people (1/7th of the world's population) now use a single social networking service, Facebook (Facebook, 2013). Around 70% of American adults have an active account in a social networking service website such as Facebook, Twitter, Linkedin, and Google+ (Trends In Telecommunication Reform, 2012). To curb the growing health care costs and improve the efficiency of health and wellness programs, it may be possible to exploit the advantage of technologies such as online social networks that can deliver the support required for a person to stay abstinent at all times. In addition, these networks have the capability to deliver interventions to large populations. Virtual social networks are online communities that allow millions of users to share experiences, ask questions, or provide emotional support and advice to one another. Today's social media platforms can be broadly classified as 1) open social networks, or 2) intentionally designed social networks (Centola, 2013). Open networks such as Twitter and Facebook support social interactions on any topic, while intentionally designed networks such as Patientslikeme and QuitNet provide platforms for participants seeking targeted interactions pertinent to health-related goals. Efforts have been made to leverage social influence and support to promote and sustain healthy behavior changes. A variety of socio-behavioral interventions have been developed to support healthy lifestyle changes by facilitating attitudinal change, behavioral adherence, and the availability of a support network (Abraham & Michie, 2008; Prochaska & Prochaska, 2011). On account of the availability and accessibility of the World Wide Web via mobile phones, virtual network interventions occur in real-time, and can provide a rich documentation of certain crucial moments in everyday life that can have positive or negative effects on a person's journey toward abstinence (Heron & Smyth, 2010). Therefore, online social networks can form the basis for ecological momentary assessments and interventions. These networks form a core component of Health 2.0, which is defined as “user-generated health care promoting patient empowerment and participation” (Van De Belt, Engelen, Berben, & Schoonhoven, 2010). As the networks mature with scale, their social value increases, and their data can provide valuable insights into fundamental questions of human behavior. Network data based experiments provide improvised versions of traditional behavioral experiments because of their scale, structural control, measurement, replicability, and behavioral fidelity (Centola, 2013). To date, scientists have attempted to analyze online social network data to solve a variety of health-related questions. A methodological review of the studies conducted on virtual networks is presented in the next section.

2.6 Online Social Communities: A Methodological Review

Given the qualitative and quantitative richness, analysis of online social networks demands multidisciplinary approach to deal with its intricacies. Researchers from diverse fields including social science, psychology, behavior science, epidemiology, computer science, and network science have analyzed real world network data yielding valuable insights into social influence, information spread, behavioral diffusion, and structural clues to identify key players and opinion leaders (Cacioppo, Fowler, & Christakis, 2009; Centola, 2010; Christakis & Fowler, 2008; Christakis & Fowler, 2009; Cobb et al., 2010; Morris, 2005; Rosenquist, Murabito, Fowler, & Christakis, 2010). There have been two predominant streams of studies on online social networks. The first category focuses exclusively on content, while the second category concentrates on structural and functional aspects of a network with no consideration of content. Below I provide a brief overview of qualitative, automated, and quantitative methods that have been used to analyze health-related social network data.

2.6.1 Qualitative Studies of Online Social Media

Several studies have adopted qualitative methods to conduct research on health-related social media. Some of these adopt a passive approach where researchers attempt to understand information seeking patterns on websites or interactions on discussion groups (Eysenbach & Till, 2001). Such studies attempt to examine the help mechanisms and content of online self-help groups for alcoholism (Chuang & Yang, 2012), cancer (Blanchard et al., 1995), and other health disorders such as Huntington's disease (Coulson, Buchanan, & Aubeeluck, 2007). The majority of these studies characterize the different types of social support embedded in the communication exchanges within the forum. Findings are consistent across different domains, showing that the majority of the content conveys emotional and informational support to its users. Instrumental support is the least common kind of support found in these online networks. Another type of network-based study involves researchers identifying themselves as such and gathering information in the form of online semi-structured interviews, online focus groups, or internet based surveys to attempt to understand consumers' use of social media. Hwang et al. conducted a network-based survey on the Sparkpeople forum, where members focus on weight loss regimen. The qualitative survey data were analyzed for social support themes using grounded theory techniques. Results indicated that the major social support themes were encouragement and motivation, information and shared experiences (Hwang et al., 2010).

2.6.2 Socio-behavioral Theoretical Frameworks

Outside the context of online networks, several health behavior theories and models have been formulated to explain behavior change in general. These frameworks have served as guides for the development and evaluation of face-to-face and online interventions. These theories can be broadly classified into individual and interpersonal models. Models such as the Health Belief Model (Hochbaum, Rosenstock, & Kegels, 1952; Janz & Becker, 1984), Theory of Planned Behavior (Ajzen, 1991), and the Transtheoretical Model (Prochaska & Velicer, 1997) belong to the former category, while Social Cognitive Theory (Bandura, 1986, 2000) and Social network, support models (Heaney & Israel, 2002) are classified as interpersonal models (explained below in detail). While each of these theoretical frameworks has its own merits and limitations, researchers have indicated concerns about the applicability of these models to the consumers in the digital era (Cobb, Graham, Byron, Niaura, & Abrams, 2011; Riley et al., 2011). Important concepts and definitions in the existing inter- and intra-individual theories of behavior change are presented in Table 1.

The Transtheoretical Model of Change (TTM): TTM tries to explain the behavior change mechanisms by synthesizing several constructs drawn from other theories (DiClemente & Prochaska, 1998). Stages and processes of change are the two main components of TTM. The former block explores the temporality of behavior change, while the later encompasses cognitive and behavioral concepts such as decisional balance, self-efficacy, and rewards program. Precontemplation, contemplation, preparation, action, maintenance, and termination are the six stages of change, where each stage involves a process of progress. This theory has been successfully used in several behavior change settings especially in the physical activity and nutrition domains (Graaf, Gaag, Kafatos, Lennernas, & Kearney, 1997; Marshall & Biddle, 2001).

The Theory of Reasoned Action (TRA): TRA suggests that the behavior of a person is determined by one's behavioral intention (Fishbein, 1979). Intent of a behavior is a function of the person's attitude toward the behavior, their subjective norm associated with the behavior, and their perceived behavioral control. Application areas for this theory include disease prevention behavior, birth control (Albarracin, Johnson, Fishbein, & Muellerleile, 2001; Fisher, Fisher, & Rye, 1995).

Social Cognitive Theory (SCT): The SCT is a theory based on reciprocal determinism between a behavior, the environment, and a person (Bandura, 2000). This theory emphasizes self-efficacy, an important concept related to self-confidence. Self-efficacy is defined as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances" (p. 391). Current literature agrees on a common definition that self-efficacy "refers to what a person believes he or she can do on a particular task" (p. 506). Goal attainment and confidence building through self-monitoring and continuous feedback is often used to improve a person's self-efficacy. Other important constructs in SCT include behavioral capability, observational learning, reinforcement, outcome expectations and expectancies, emotional coping and self-control. The construct of 'observational learning' has been used by network scientists to provide an explanation for social influence and network clustering of people engaging in the same health behavior (Bandura, 2002). According to SCT, observational learning in behavior change occurs when an individual watches another person engage in a given behavior and receive reinforcements. Another component of SCT called reciprocal determinism takes into account the interactions among individuals, their environments, and behavior goals. The environment in SCT refers to a conglomeration of factors that are external to the individual including his/her social network - family, friends, and peers, and physical objects that might affect behaviors. In case of smoking the physical objects can include availability of patches, access to smoking-designated areas in the work place, and so forth.

Health Belief Model (HBM): This is one of the most widely used conceptual frameworks for explaining and changing individual health behavior. HBM evolved from a cognitive theory perspective and is a value-expectancy theory, which attempts to explain and predict individual's attitudes toward objects and actions (Hochbaum et al., 1952). Major components in HBM include perceived susceptibility, perceived severity, perceived benefits, perceived barriers, cues to action, and self-efficacy. An individual’s perceptions of a behavior can be used as predictors of behavior change outcomes under certain conditions that are dependent on demographic (e.g. age, gender) sociopsychological (e.g. personality, social class), and structural variables (e.g. prior knowledge, experience). HBM has been applied to many important healthcare problems focussing on behavioral adherence (e.g. seat belt use) (Fernandes, Hatfield, & Soames Job, 2010).

Table 1. Theoretical constructs from behavior change theories

Adapted from (Revere & Dunbar, 2001)

CONCEPT	DEFINITION
Health Belief Model	
Perceived susceptibility	One's opinions of chances of getting a condition
Perceived severity	One's opinions of how serious a condition and its consequences are
Perceived benefits	One's opinion of the efficacy of the advised action to reduce risk or seriousness of impact
Perceived barriers	One's opinion of the tangible and psychological costs of the action
Cues to action	Strategies to activate readiness
Self-efficacy	Confidence in ability to take action and persist in action
Stages of Change Model	
Pre-contemplation	Unaware of problem, hasn't thought about changes
Contemplation	Thinking about changes
Preparation	Making a plan to change
Action	Implementations of a specific action plan
Maintenance	Continuation of desirable actions, or repeating periodic recommended step(s)
Consciousness Raising	Increasing awareness via information, education, and personal feedback about the healthy behavior
Dramatic Relief	Feeling fear, anxiety, or worry because of the unhealthy behavior, or feeling inspiration and hope when they hear about how people are able to change to healthy behaviors
Self-Reevaluation	Realizing that the healthy behavior is an important part of who they are and want to be
Environmental Reevaluation	Realizing how unhealthy behavior affects others

Social Liberation	realizing that society is more supportive of the healthy behavior
Self-Liberation	believing in one’s ability to change and making commitments and recommitments
Helping Relationships	finding people who are supportive of their change
Counter-Conditioning	substituting healthy ways of acting and thinking for unhealthy ways
Reinforcement Management	increasing the rewards that come from positive behavior and reducing those that come from negative behavior
Stimulus Control	using reminders and cues that encourage healthy behavior as substitutes for those that encourage the unhealthy behavior
Theory of Planned Behavior and Theory of Reason Action	
Behavioral intervention	Perceived likelihood of performing the behavior; prerequisites for action
Attitude	One’s favorable or unfavorable evaluation of the behavior
Behavioral belief	Belief that the behavioral performance is associated with certain attributes or outcomes
Normative belief	Subjective belief regarding approval or disapproval of the behavior
Subjective norm	Influence of perceived social pressure weighted by one’s motivation to comply with perceived expectations
Perceived behavioral control	One’s perception of how easy or difficult it will be to act
Social Cognitive Theory	
Reciprocal determinism	Behavior change results from interaction between individuals and environment
Behavioral capability	Knowledge and skills to influence behavior
Expectations	Beliefs about likely results of action
Self-efficacy	Confidence in ability to take action and persist in action
Observational learning	Beliefs based on observing others

Reinforcement	Responses to a person’s behavior that increase or decrease chances of recurrence
Emotional coping responses	Strategies or tactics that are used by a person to deal with emotional stimuli.

While majority of qualitative work on content analysis of online social media (as discussed earlier in the section) has focused on understanding social support, it is important to note that, in addition to social support, behavior change in online social networks encompasses several other aforementioned socio-behavioral concepts such as reward management, stimulus control, outcome expectations and expectancies, social influence, and observational learning. For example, Bandura suggests using small-scale tasks that people can succeed in to enhance their self-efficacy. On the other hand, the Transtheoretical Model of Change suggests the use of counterconditioning to remove cues to indulge in a risky behavior. Hence, it is evident that different cognitive constructs in existing behavior change theories suggest different techniques. Recent online survey research examined users’ perception of different social influence mechanisms to understand relationship between network participation and smoking cessation self-efficacy (Phua, 2013). Results indicated that participation in health issue-specific social networking sites significantly influenced each social factor, which in turn resulted in greater smoking cessation self-efficacy. However, it is not known the extent to which any of the theoretically-grounded strategies empirically manifest in the communication among network users.

2.6.3. Informatics Methods for Automating Social Media Analysis

Recent advances in automated text analysis allow for large-scale analysis of the content of communication between members. Semantic analysis of social networks using automated methods has been previously applied to the study of research communities in the field of enterprise interoperability. Velardi et al. performed content-based social network analysis with the aid of linguistic analysis, text mining, and clustering techniques, in which the semantic relatedness between terms was measured using a taxonomy-based approach (Velardi, Navigli, Cucchiarelli, & D'Antonio, 2008). Meta-data based approaches have also been used to derive person-word relations (e.g. author-specialization) by extracting social network information using semantic approaches (Matsuo et al., 2007). Classification of conversational and informational questions on social Q&A websites such as Yahoo! Answers has also been attempted using a combination of human coding, statistical analysis, and machine learning (Harper, Moy, & Konstan, 2009). Another application area of automated natural language processing method is the development of a consumer health vocabulary based on threaded discussions in online social network websites (Doing-Harris & Zeng-Treitler, 2011). Doing-Harris et al. have developed a computer-assisted update (CAU) system that consisted of three main parts: a Web crawler and an HTML parser, a candidate term filter that utilizes natural language processing tools including term recognition methods, and a human review interface. The CAU system was applied to the health-related social network website PatientsLikeMe.com to develop and dynamically update the health vocabulary (C. A. Smith & Wicks, 2008). Another avenue for automated methods in analysis of online social media content is assessing similarity between two separate texts to derive and understand content structure. Content similarity has been used as a filtering metric along with link analysis to rank influential users in a web forum (Tang & Yang, 2010). Research employing both network models and estimates of semantic relatedness can be found in the psychological literature. For example, Pathfinder networks employ a scaling technique that builds on relatedness between nodes. If each node represents a concept, the weights of links (or edges) present in the network are defined using human estimates of the relatedness between all pairs of concepts (Schvaneveldt, 1990).

A number of methods have been developed to automate the derivation of similarity metrics between terms based on distributional statistics of unannotated electronic text. Spatial semantic models define terms as vectors in high dimensional space according to the distribution of their occurrence across a large corpus. Different semantic space models use different approaches to derive this multi-dimensional space. In Latent Semantic Analysis (LSA), each document in a text collection is considered as a unique context (Dumais, 2004; Landauer, Foltz, & Laham, 1998). The coordinates of a term-vector in semantic space are determined by the distributional statistics for this term and therefore similar vector representations are created for terms that occur in similar contexts (second-order relationships between terms that do not co-occur directly are also identifiable once Singular Value Decomposition of the original term-by-document matrix has occurred) (Cohen, Schvaneveldt, & Widdows, 2010; Vasuki & Cohen, 2010). Evidence suggests that the semantic relatedness measured derived using LSA agree with human estimates, and can be used to obtain human-like performance in a number of cognitive tasks (Landauer, 2002; Landauer & Dumais, 1997). Hyperspace Analogue to Language (HAL), another distributional semantics approach has been combined with machine learning algorithms to automatically classify consumer health webpages of based on language use patterns (Chen, Warren, & Evans, 2008, 2009; Chen, Warren, & Riddle, 2010). The HAL model uses as its context a smaller neighborhood of words surrounding the target term. The matrix generated is a term-term matrix rather than a term-document matrix, and term frequencies are calculated according to the degree with which they co-occur within a sliding window. Distributional semantics methods such as LSA and HAL have been used for information retrieval, learning assessment in psychology, and emotion analysis (Cohen, 2008; Landauer, 2002; Lund & Burgess, 1996). While the methods for automated analysis of free text are still evolving, these methods have the capability to deal with large amounts of data generated by social media. Combining these methods with traditional network analysis may be of value to the research and practice of behavioral and social sciences.

2.6.4. Network Science Models

Network analysis studies on health-related online social networks have focused primarily on exploring the structural and functional composition of networks. Traditionally, social network analyses represent the same set of entities. Such networks are called one-mode networks. A data matrix is said to be 2-mode if the rows and columns represent different entities (e.g., the rows might correspond to persons while the columns correspond to departments). A network representing a 2-mode matrix is called affiliation matrix (Borgatti & Everett, 1997). A network representing the matrix resulting from the multiplication of an affiliation matrix and its transpose is called co-affiliation matrix. Irrespective of one-mode and two-mode networks, methods for analyzing network characteristics such as centrality measures, core/periphery structures have been employed to study these networks. Detailed definitions of social network metrics and terms are presented in Appendix A.

Several studies have been conducted that apply network analyses to interpret online social media datasets. Cobb et al. studied the online structure of a large online community focused on smoking cessation by characterizing the social network and participants of this community, describing its structure, and identifying network subgroups (Cobb et al., 2010). Centola studied the effects of network structure on diffusion by studying the spread of health behavior through artificially structured online communities (D. Centola, 2010). Large online network dataset have been used to distinguish between influence and homophily effects, thereby enhancing our understanding of the mechanisms that drive contagion in networks and our knowledge of how to propagate or combat it (Aral, Muchnik, & Sundararajan, 2009; Shalizi & Thomas, 2011). Network analytics have also been used to determine the association between social influence and consumer engagement in the context of an online behavior change intervention (Poirier & Cobb, 2012). In addition to network analysis, other mathematical modeling techniques have been applied to Twitter datasets to conduct syndrommic surveillance, measure behavioral risk factors, localize illnesses by geographic region, and analyze symptoms and medication usage (Chew & Eysenbach, 2010; Christley et al., 2005; Paul & Dredze, 2011). While previous studies applying network analysis techniques have been used to examine social influence on health behaviors (Valente, 2006; Valente & Fujimoto, 2010; Valente, 2012), most of those studies are based on face-to-face networks. These concepts need to be understood in the context of online social networks in order to translate them into engaging, sustainable, and effective next-generation behavior change platforms.

2.7 Research Opportunities in Online Health-related Social Network Analysis

Looking at the past and current trends of health-related online social networks, several research avenues can be pursued in order to strategize the use of the networks for improving healthcare. Advancing existing socio-behavioral theories, understanding fundamental mechanisms of behavior change, and formulating and evaluating novel interventional approaches, are important avenues of research opened by these virtual platforms. Are theoretical models of socio-behavioral models of change applicable to both offline and online contexts? Given the ubiquity of online social networks, the network data contain traces of the real-time needs, mental models, and cognitive and socio-behavioral aspects of a person attempting to introduce a new change or sustain an existing behavior modification, thus providing us with an unprecedented opportunity to refine existing theories and models of social networks, social support, and behavior-change that were formulated based on face-to-face communication. How can new methods and metrics be formulated to capture and analyze data patterns derived from online social networks to inform our understanding of behavior change at the individual and network level? New methodological approaches should scale to large online social network datasets. Engineering these approaches is essential if we are to introduce new analytical dimensions into online social network analysis. As mentioned earlier, much of the network analysis studies have ignored communication content. Those studies that considered content have adopted qualitative methods thus limiting their scalability to larger datasets. Content inclusion in network models is another important research direction that needs to be addressed in order to translate existing theoretical strategies into network-based intervention approaches. Most of the behavior change theories suggest strategies that are predominantly custom-tailored at the level of the content that is delivered as part of the intervention. Understanding and leveraging network content may allow us to incorporate theoretically-grounded approaches into interventions at a network level, thus enhancing their impact and efficacy. Other interesting potential research strategies include: a) formulating new methods to identify important nodes within online social networks to tailor or deliver a behavior change intervention by disentangling the factors underlying communication patterns and user engagement, b) using network data to custom-tailor intervention delivery offline and online, c) comparative effectiveness research on interventions to see if findings based on retrospective self-reports with sparse observations in the real world are consistent with those based on dynamic, observed behavioral data collected online, and e) interfacing online social networks with other elements of health care such as physician, insurance, workplace, and offline connections (such as family and friends who are not part of the online network).

Chapter 3: Materials

QuitNet is one of the first online social networks for health behavior change, and has been in continuous existence for the past 16 years. It is widely used with over 100,000 new registrants per year (www.QuitNet.com). QuitNet has members who are current and former smokers seeking to quit or stay abstinent. The members are globally-distributed and come from over 160 countries including Canada, the United Kingdom, Australia and South Africa.

QuitNet website incorporates the United States Public Health Service guidelines for best practice and includes diagnostic tools, social support from peers and experts, and pharmacotherapy (Cobb et al., 2005). It is available to smokers through two main channels: free public internet access and paid contracts. Both versions operate in the same environment and have a single support community (Cobb et al., 2005; Graham et al., 2007), therefore, regardless of the means by which users access QuitNet, they all participate in the same online community. Research materials used for this study were extracted from the publicly available version.

Previous studies on QuitNet indicated that participation in the online community was strongly correlated with abstinence (Graham et al., 2007). The dataset studied in my research was drawn from a previously studied quality improvement database, and is comprised of de-identified messages in the public threaded forums, in which participants post messages and reply directly to each other. A database of 16,492 de-identified public messages between March 1, 2007 and April 30, 2007 was used in our study. All messages were stripped of identifiers but re-coded for ego id (the individual posting) and alter id (the individual whose message is being replied to), self-reported smoking status of sender and receiver (‘0’ for aspiring quitter, ‘1’ for current smoker/non-quitter), date and position within the thread. This dataset included both new users and QuitNet members who registered before the study period.

Figure 3 presents an overview of the QuitNet communication patterns over a period of two months. Around 200 members posted just one message, 576 members posted between 2 and 5 messages, followed by 201 members who posted between 6 and 10 messages. 355 members posted between 11 and 50 messages. Around 70 members posted more than 50 but less than 100 messages. About 60 members posted more than 100 messages during the time period under consideration. It is clear from the dataset that majority of QuitNet users had low levels of activity in terms of direct message posting to the community, thus raising important questions about user engagement and content-specific patterns depending on users’ activity levels.

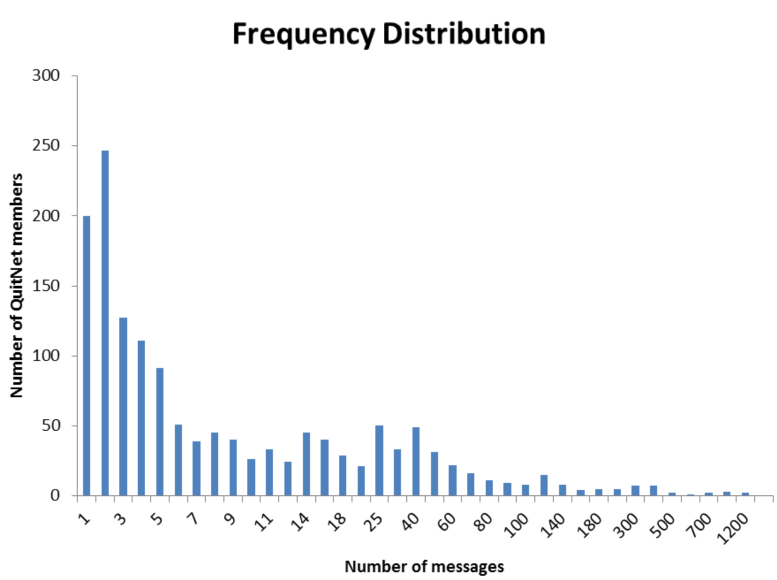


Figure 3. Frequency distribution of QuitNet dataset

QuitNet members were classified into four groups based on their self-reported smoking status. The classification criteria were as follows:

0: Members who were smokers throughout the study period (current smokers)

1: Members who stayed abstinent during the entire study period (ex-smokers)

0to1: Members who switched their status from smokers to ex-smokers (successful quitters)

1to0: Members who altered status from ex-smokers to smokers (relapsers)

Other: Members who changed their smoking status multiple times (frequent relapsers)

The following graph provides a summary of QuitNet members classified into these five groups. Across the five groups, majority of the QuitNet members in our dataset are female aged between 35and 49. Interestingly, all the Group 1to0 members are female between 25 and 49 years.

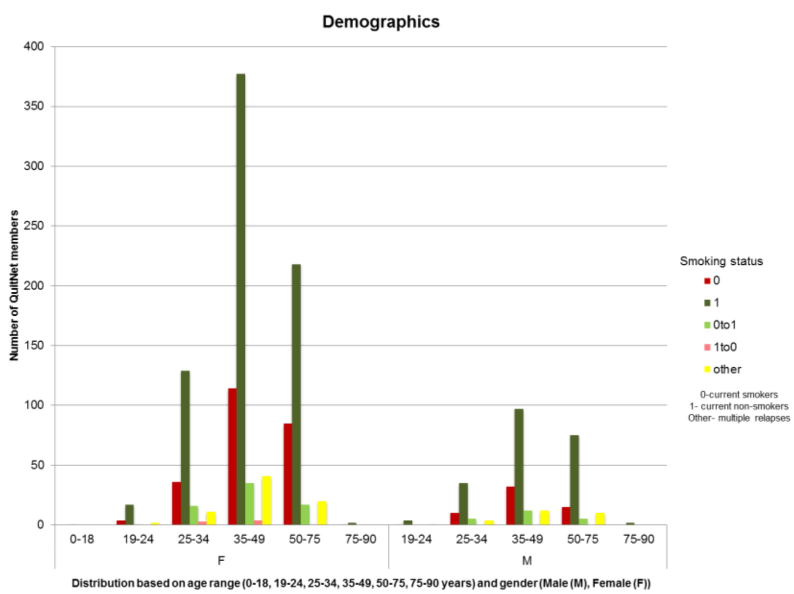


Figure 4. Distribution of QuitNet members based on their smoking status

Chapter 4: Qualitative Analysis of QuitNet Content

Most prior research on social networks has made exclusive use of structure of social ties, where network structure is derived from the frequency of communication among members belonging to a network. However, existing behavior change theories suggest that message content plays a prominent role reflecting the socio-cognitive factors affecting a person's efforts to make a lifestyle change (Heaney & Israel, 2002). An understanding of these factors is required if we are to understand the mechanisms of behavior change in the "Web 2.0" era, where consumers are often the drivers of technological interventions and digitized information sources.

In this chapter, I describe the results derived from a grounded theory-based (Glaser & Strauss, 1967; Strauss & Corbin, 2008) content analysis of QuitNet messages. The findings derived from this analysis using our method are then interpreted in the light of existing behavior change theories, in an attempt to understand the interplay between the behavior changes facilitated by Web 2.0 based interventions and existing health behavior models. This analysis enhances our understanding of the applicability of behavior change theories (such as Socio-cognitive Theory (SCT) (Bandura, 1986) and The Transtheoretical Model of Change (TTM) (Prochaska & Velicer, 1997)) which were formulated in the context of face-to-face communication using laboratory-based social science approaches, in the context of online social relationships.

4.1. Grounded Theory based Content Analysis

The objective of my thematic analysis was to identify key concepts contained in the messages exchanged by QuitNet members, thus capturing essential meaning of communication and factors affecting smoking cessation. This analysis ultimately enables the generation of content-specific networks that relate to themes that emerge from the data itself, as I will describe in subsequent chapters. Such inductive analysis is the principal technique used in the grounded theory method generating themes, where themes emerge from data itself (Strauss & Corbin, 2008). Open coding and Constant comparison are the two main characteristics of the analysis that can be used to ensure the derivation of meaningful representative themes from social network data. Open coding describes data by means of conceptual (rather than descriptive) codes, which are derived directly from the data, and constant comparison enables creation of precise and consistent codes by comparing these codes to observed phenomena and their contexts many times.

Oftentimes, the messages exchanged among network members reflect a local language that is ingrained in the network's unique culture. In a forum where people attempting to lose weight come together, 'morning news' refers to a cue to keep them from consuming junk food. However, when it is interpreted out of context they lose their context-specific meaning. Similarly, in a much more general sense, before the advent of 'Twitter' (an online social networking and microblogging service), the word 'tweeples' was never used. Interestingly, current trends suggest having a high number of 'tweeples' as a metric to measure how well followed a person is. Emergence of local language is a commonly found feature of a community, and the same can be applied to virtual communities as well. Therefore, when analyzing online social network data to understand communication patterns underlying human behavior, understanding community –specific context is mandatory to derive meaningful inferences from the data.

A grounded theory approach was used to analyze QuitNet data to understand the core concepts, the interrelations among concepts and the roles played by these concepts in an individual's smoking cessation activity. The first step in the coding process involved open coding, where a line-by-line analysis was performed on the messages to derive abstract concepts from the data. Each message was reviewed, noting pertinent smoking cessation related concepts in terms of general open codes which were generated dynamically as the data were reviewed.

Examples of open codes included "statistics", "crave", "pregnancy", "boredom", "temper", "patch", and "pledge". This process was repeated until no new concepts were produced from the dataset. Appropriateness of code assignment was ascertained using constant comparison, where instances of codes were compared in an iterative manner to make sure they reflected the same concept. The second step was performed by re-organizing and re-grouping the open codes using axial coding. Axial coding allowed for the identification of unifying, repeated patterns underlying the concepts and their relationships, thereby revealing core themes relevant to smoking cessation. Examples of core themes include "Family and friends", "Obstacles", and "Traditions". Initial coding was performed manually, and later the NVivo software suite for qualitative analysis was used to analyze themes and their patterns of occurrence in the data (Nvivo, 2013). A total of 585 messages were analyzed, revealing 43 distinct concepts. Furthermore, the analysis was carried out for an additional 210 messages to ensure no new concepts emerged. This qualitative coding allowed for an in-depth evaluation of the interactions among people in the QuitNet virtual community and thereby a deeper understanding of the behavior change processes that QuitNet users undergo when attempting to cease smoking.

4.2 Results and Discussions

4.2.1 QuitNet Themes

A total of 43 different concepts were identified, which were then grouped under 12 themes. Examples of the grouping strategy employed to arrive at the thematic level are shown in Figure 5, where the "Obstacles" theme is composed by subsuming multiple concepts: "sleepiness", "weight gain", "temper", "boredom", and "trouble sleeping". These concepts were cited as hurdles that members faced in their attempts to quit smoking. Similarly, with "Traditions", "playing games", "sharing weather details", "attending virtual bonfire events", and "taking part in daily online pledge" were the observed communal practices that are deeply rooted in the QuitNet community. On the other hand, there were a few concepts such as "Craves" where smokers and ex-smokers described the temptation to smoke, which was deemed to be a self-contained theme. Definitions of the themes and example messages for each theme are listed in Table 2.

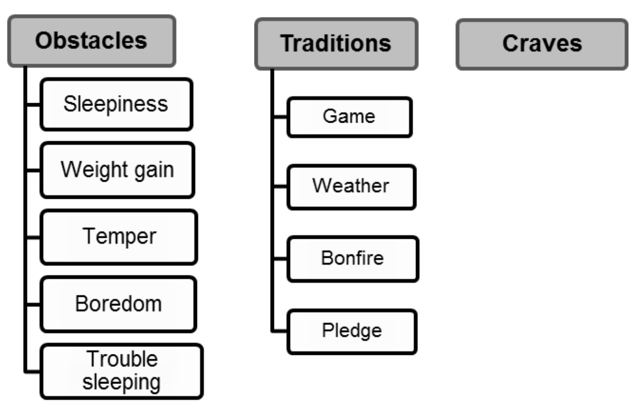


Figure 5. Themes in QuitNet

Table 2. QuitNet themes, definitions, and example messages

Theme	Definition	Example Message
Quit Obstacles	Messages in which members talk about the hurdles they are dealing with or have dealt with to stay abstinent (e.g. sleepiness, weight gain, temper)	*I lost quits in the past because I was so mean and nasty that my family and friends told me to smoke.*
Teachable Moments	Messages where the senders mention about the incentives one gets for not smoking in terms of quality of life (e.g. family bonding), better perception of everyday moments (e.g. savoring food), reduction in health risks (e.g. cancer, blood pressure)	*Food is wonderful.....smell is wonderful...I smoked from 14-46...I never knew what I was missing..*
Motivation	Messages which attempt to provide inspiration to stay abstinent or initiate a smoke-free life	*You can do anything if you would want it bad enough...*
Craves	Messages in which members discuss about dealing with a crave for cigarette	*I want a cigarette very much. I am trying to resist.*
Conflict	Messages that reflect a rift between two group members	*Guess I am naive as I think most everyone in this Q is honest. ... No One likes being called a liar, especially if they are NOT. Go sit*
Relapse (confessions, reasons, retries)	Messages in which members explain why they relapsed and/or share their emotions after they suffered a relapse	*I hate myself, I slipped again. I lighted the nicodemon*
Traditions	Messages which focus exclusively on QuitNet-specific events such as bonfires, pledges, games, and so on	*So nice of you to host this shindig tonight. I've got over 5K unsmoked cigs which I'd be delighted to unload onto a raging bonfire.*
Quit Progress	Messages in which members communicate their progress based on abstinence time and/or number of unsmoked cigarettes	*Gratefully smoke free for 33 days, 17 hours, 1 minute and 6 seconds.*
Family and Friends	Message in which members mention their spouses, children, or friends as motivators	*My hubby should be answering ....poor guy used to get to sleep when I*

		smoked..now he is sleepless but smiling....
Virtual Rewards	Messages in which members mention the virtual gifts (such as bracelet, virtual pet, socks) received on QuitNet marking a milestone	*awesome three days. I like the bracelet.*
Social Support	Messages where the content reflects the elements of praise, advice, empathy, and guidance	*Almost a year already.////Congratulations to you, what a great accomplishment.*
Nicotine Replacement Therapy (NRT) entries	Messages where members explicitly discuss and evaluate various pharmacotherapy options	*I did not use any nrt though I recently went on welburtin after days ct*

A detailed distribution of the themes across messages is shown in Figure 6. “Traditions”, “Social support” and “Progress” were the most frequently found themes, followed by discussions related to “Teachable moments/Benefits”, “Relapse”, and “Craves”. “Conflict”-related messages were the least frequently found, only behind “Virtual Rewards”, “NRT entries”, “Family and friends”, and “Obstacles”.

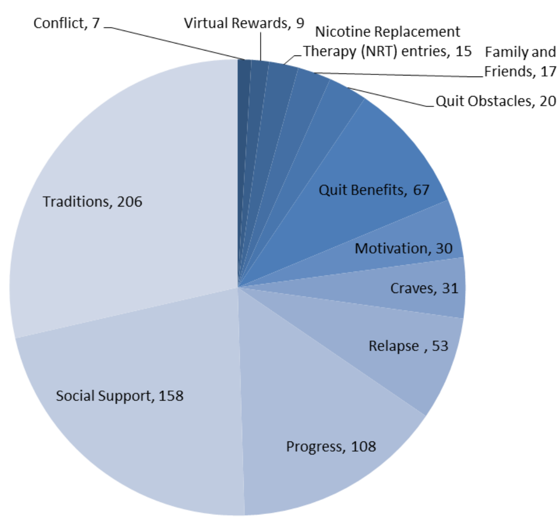


Figure 6. Grounded theory based qualitative analysis of 790 messages

QuitNet members exchange messages pertaining to “Traditions” that are specific to QuitNet. Examples of traditions are as follows- a) “Bonfire”: a virtual event hosted regularly where members bring their unsmoked cigarettes and throw them into a fire, b) “Pledge”: a member virtually extends his or her hand to another member indicating their commitment towards staying abstinent. This represents the support the member offers to the next person in line to help him or her stay smoke-free, and as such is one example of the content of messages belonging to the “Social support” category. These messages provide guidance, express empathy, convey admiration, and promote bonding. Expressions of empathy, love, trust, and caring which form the basis of emotional support were also communicated using phrases such as ‘hugs’, ‘flowers’, and ‘kisses’. Members use measurable metrics such as the number of unsmoked days and cigarettes, the amount of money not spent on cigarettes, and the number of days of life saved by staying smoke-free to measure their “Progress”. These metrics are calculated by an automated script available on the QuitNet website. Members refer to these calculated metrics when providing positive feedback to others, and utilize them for self-monitoring.

Analysis of the QuitNet data provided crucial insights into the “Relapse” experiences of smokers and ex-smokers. Work-related stress, family tragedies, inability to ward off craving, and a false notion of ‘just one puff’ (denotes weak moments where members smoke a cigarette thinking that it would not affect their ability to stay abstinent from then on) that were cited as common reasons for relapse. Relapse is a common problem encountered by smokers who are trying to quit and ex-smokers who successfully quit (Center for Disease Control and Prevention CDC, 2013). In addition to messages indicating risk factors for relapse, messages where members declare their relapse and communicate their emotions (e.g. “tears rolling down”, “cheating the loved ones”, and “feeling like a loser”) after relapse were also included in this theme. Also, messages describing the ‘aha moments’ where members recollect the reasons behind their decision to quit smoking occur in the dataset. Health-related issues such as the onset of smoking-related disease and pregnancy are cited as common drivers for these “Teachable moments/Benefits”, while quality of life concerns such as problems related to exercise, family time, physical appearance, and social awkwardness are also listed as reasons for quitting.

The majority of the challenges in QuitNet members’ journeys towards smoke-free lives were defined by cravings for cigarettes. This theme, called “Craves”, includes messages with content where successful quitters explain to fellow members how they dealt with craving. Some messages even contained information about members’ experiences and efforts as they dealt with craving in real-time. Messages relevant to the “Motivation” theme displayed an effort made by QuitNet members to encourage fellow members by making inspiring, engaging, and thought-provoking comments on the role played by personality traits such as attitude and willpower in a successful quit attempt. Messages also have content through which members mentioned the “Obstacles” they were facing, or have faced, at some point of their abstinence phase. Weight gain, temper, problems with sleep, and boredom were among these hurdles. “Family (e.g. spouse, children) and friends” are mentioned in some of the messages as support network or motivators or obstacles. For instance, members mentioned about not being able to stay abstinent because of watching their spouses smoke. “Nicotine Replacement Therapy (NRT) entries” such as information sharing about the usage of patches and gums and going “cold turkey” (cold turkey means quitting without any pharmaceutical assistance) are also mentioned in QuitNet messages. The members requested information about withdrawal effects and side-effects associated with the use of NRT. Also, successful quitters advised newer members to make use of a patch to fight cravings and avoid relapse. Another emergent behavior exhibited by QuitNet members involved the role of “Virtual Rewards”. Some of these rewards included bracelets, virtual pets, socks, and access to an “elder lodge” where successful quitters meet virtually. Rewards were given when members met milestones such as 3-day, 15-day quit, and 100- day quits, one year anniversaries, and so forth.

4.2.2 Thematic Inter-relationships: Comparison with Existing Behavior Change Theories

The themes identified in QuitNet communication relate to the socio-behavioral and cognitive constructs of the existing behavior change theories. Tables 3 and 4 shows how QuitNet themes can facilitate a driver of behavior change that relates to one of the theoretical constructs. A comparison matrix for inductively-derived themes (seen in columns) and theoretically-derived constructs (seen in rows) is provided based on the comparative analysis of their definitions and the concepts they represent.

Table 3. Theme-Theory matrix (Part 1 of 2, QuitNet themes continued)

QuitNet Themes / Theoretical Constructs	Conflict	Virtual Rewards	NRT entries	Family and Friends	Quit Obstacles	Quit Benefits
Susceptibility						■
Severity						■
Benefits						■
Expectations						
Expectancies						■
Barriers					■	
Cue to action			■			
Self-efficacy		■				
Intention						■
Belief						
Norm	■					
Control						
Decisional balance					■	■
Consciousness raising	■		■			
Dramatic relief						
Self-reevaluation						■
Environmental re-evaluation	■			■		■
Self-liberation						
Helping relationships				■		
Counterconditioning		■	■			
Reinforcements		■				
Stimulus control		■				
Social liberation	■			■		
Environment		■		■		
Behavioral capability						
Self-control		■				
Observational learning		■		■		
Emotional coping response				■		

Table 4. Theme-Theory matrix (Part 2 of 2, QuitNet Themes continued)

QuitNet Themes / Theoretical Constructs	Motivation	Craves	Relapse	Quit Progress	Social Support	Traditions
Susceptibility	■		■			
Severity	■					
Benefits						
Expectations			■			
Expectancies						
Barriers		■	■		■	
Cue to action					■	
Self-efficacy	■		■	■		■
Intention						
Belief	■					
Norm						■
Control	■					■
Decisional balance		■				
Consciousness raising					■	
Dramatic relief		■	■			
Self-reevaluation						■
Environmental re-evaluation						
Self-liberation						■
Helping relationships			■		■	■
Counterconditioning						■
Reinforcements						
Stimulus control				■		■
Social liberation	■					
Environment						■
Behavioral capability			■			■
Self-control				■		■
Observational learning		■	■	■		■
Emotional coping response		■	■		■	■

For example consider the concept “stimulus control” which involves using reminders and cues that encourage healthy behavior as substitutes for those that encourage the unhealthy behavior. For example, for individuals who are accustomed to smoking early in the morning, there exists a QuitNet-specific tradition where members post messages describing early morning weather and reaffirm their commitment to stay abstinent. A shaded cell (see Tables 3 and 4) indicates that a given theme relates to a particular construct. Themes such as “Virtual rewards”, “Traditions”, and “Progress” have components that attempt to improve self-efficacy of an individual. Self-efficacy, which is defined by Bandura’s Social Cognitive Theory as one of the vital ingredients for behavior change (Bandura, 2000; Dijkstra et al., 1996), deals with the belief in one's capabilities to achieve a goal. An individual’s self-efficacy can affect his or her motivation to achieve a goal, such as adhering to a healthy behavior. Persons with high self-efficacy are more likely to persist longer in efforts to achieve the desired goal (McAuley, Lox, & Duncan, 1993; Strecher, DeVellis, Becker, & Rosenstock, 1986). In case of smoking cessation, ability to ward off cravings and stay abstinent can be improved by enhancing a person’s self-efficacy (Dijkstra et al., 1996), which can be achieved by setting and achieving short-term goal settings. Organizing such goals in group environment also induces observational learning. For instance, “Virtual Rewards” such as bracelets and virtual pets accomplish the task of short-term goal setting. Watching other members receive these rewards often motivated QuitNet members to stay abstinent as evidenced by the following quote- *“So proud of you, I won’t light my cigarette, want that lovely bracelet on my hand”*. Bonfires (a component of the

“Traditions” theme) are related to observational learning, where a QuitNet member is motivated by the praise another member received at the event on account of the number of unsmoked cigarettes he or she brought. Similarly, themes such as “Craves” and “Relapse” address the aspect of dramatic relief described by the Transtheoretical Model of Change (DiClemente & Prochaska, 1998). “Teachable moments/Benefits” and “Obstacles” relate to the decisional balance component of behavior change. Environmental reevaluation is also provided by these two themes. The “Social support” theme includes several important constructs such as consciousness raising, cue to action, emotional coping, and helping relationships. For example, when a member was attempting to overcome a crave, other QuitNet members often post messaged that attempt to help the peer member realize how far they have come, the reasons for their quitting, and provide them with a supporting shoulder.

As described above, several constructs from the existing intra- and inter-individual behavior change theories are put together and compared with the themes derived from QuitNet content. The graphs in Figure 7 present the prevalence of the constructs in QuitNet themes. Self-efficacy and observational learning are the most relevant theoretical constructs, followed by observational learning and helping relationships (Figure 7, bottom). On the other hand, “Traditions”-related messages were highly aligned with theoretical constructs, followed by “Relapse”, “Virtual rewards”, and “Teachable Moments/ Benefits” (Figure 7, top). Results indicate that community-based activities such as “Traditions” organized in virtual communities such as QuitNet might play an important role in operationalizing theoretical constructs in the virtual settings. In addition, member-generated strategies such as bonfires, pledges facilitated the highest number of theoretical constructs from a variety of theories. Therefore, it is important to note that no single theory from behavioral science provides a basis for all of the themes emerging from QuitNet messages.

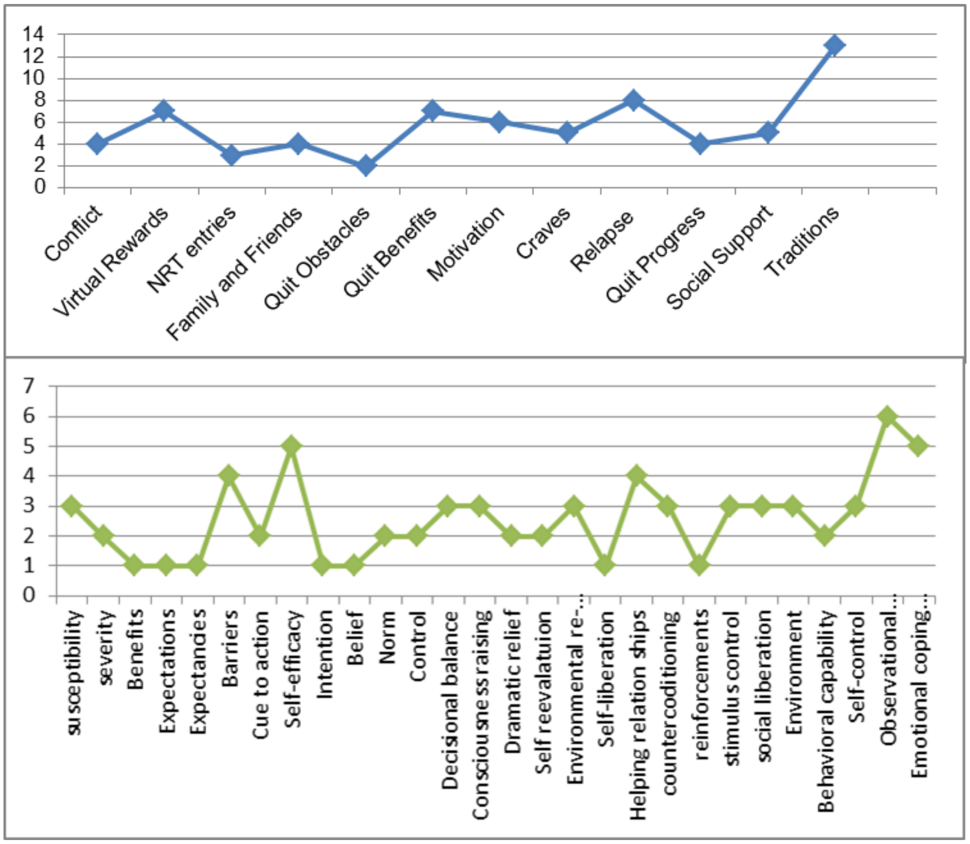


Figure 7. Thematic and theoretical prevalence in QuitNet content

In the case of QuitNet, activities such as pledges and bonfires emerged from within the community and each of those events marks a specific aspect of the smoking cessation process. With the evolution of communication channels from being traditional face-to-face conversations to virtual social networks powered by web-based mobile health systems, the validity of existing behavior change theories in the digital era has been questioned (Riley et al., 2011). This qualitative analysis establishes the validity of behavior change theories in the context of 21st century technologies. In addition, the inductive evaluation of social network content revealed new socio-cognitive constructs, which have not been considered by existing behavior change theories. For instance, the theme “conflict” which deals with the conflicts that arise between QuitNet members and is not found in any existing socio-behavioral theories because the foundations for these theories is group cooperation and not group competition. This finding highlights the need to incorporate mechanisms that build trust among members who communicate with one another using virtual channels. As another example, none of the messages mentioned the role of a physician in their efforts to cease smoking, suggesting the behavior change effort in this community is primarily self-propelled. Like any other virtual community, most content embeds aspects of social support. In addition to support, several other socio-behavioral elements related to behavior change theories were found in QuitNet messages. In addition to emphasizing on progress and positive aspects of smoking cessation, focus on community-building and social togetherness (e.g. bonfire) have help members adhere to their quit attempts.

4.3 Limitations

This qualitative analysis provided useful insights into prominent themes in QuitNet communication. However, manual coding is highly labor-intensive and time consuming. Consequently, the analysis is limited to a small sample size, potentially limiting the generalizability of these results. One limitation of this analysis is the size of the random sample of messages chosen. It is possible that given the low fraction of messages thematically coded, the distribution of the themes might not have been accurately represented. To attempt to address this, 210 messages were coded to reach thematic saturation. However, it may be possible that the remainder of the data set contains additional themes that were not captured. The rapid growth of Health 2.0 technologies will further complicate this issue, as it will generate a data deluge of millions of messages transmitted over the web and mobile health media. Therefore, for large datasets, one needs to complement the qualitative method with an automated technique that can optimize resource utilization.

4.4 Conclusions

This chapter describes a qualitative analysis of online social network communication using a grounded theory approach. The key contributions of this study are as follows- 1) the study describes the first grounded theory based qualitative analysis of the communication in an online social network developed to promote behavior change, and 2) the study attempts to understand the applicability of existing behavior change theories to virtual communication channels. Capturing the essence of the meaning underlying the messages exchanged during different situations and contexts in this manner provides important information to guide further investigations. Qualitative analysis of communication between members of an online social network can provide valuable insights into the mechanisms underlying human behavior change. With the onset of mobile smart phones and ubiquitous internet connectivity, online social network data reflects the intricacies of human health behavior as experienced by real people in real time. Therefore, analysis of these data can also provide us with the much needed theoretical and empirical foundations for design of effective intervention strategies. This study offers insights into the various kinds of behavioral constructs prevalent in the messages exchanged among QuitNet users. In addition, it underlines the need for the use of inductive approaches for the analysis of online social network data to capture community-specific culture. As such, these findings suggest the need for an aggregation of multiple theoretical constructs from more than one inter- and intra-individual theory. Given the context-rich nature of the messages, they yield empirical understanding of human behavior change. This understanding has important implications for both theory and practice. Theoretically, inductive analysis of virtual communities provides us with a basic understanding of human behavior in the digital era. Pragmatically, it sets the stage for real-time digital health interventions promoting healthy lifestyle modifications.

Chapter 5: Scaling Up Social Media Analysis using Automated Methods

Recent advances in automated text analysis allow for large-scale analysis of the content of communication between members of an online social network. Methods of distributional semantics such as Latent Semantic Analysis (LSA) (Landauer et al., 1998) provide us with the capability to derive relatedness measures between terms from unannotated text. With LSA and other geometrically motivated models, this is accomplished by representing the terms in a high dimensional vector space, referred to as a “semantic space”. The coordinates of a term vector in semantic space are determined by the distributional statistics for this term, such that similar vector representations are created for terms that occur in similar contexts (Cohen & Widdows, 2009). Evidence suggests that the semantic relatedness measures derived using distributional semantics techniques agree with human estimates, and can be used to obtain human-like performance in a number of cognitive tasks (Landauer & Dumais, 1997; Landauer, Laham, Rehder, & Schreiner, 1997). Studies have used LSA to automate the coding of communication content among group members to assess team cognition (Kiekel, Cooke, Foltz, & Shope, 2001), suggesting the applicability of the method for communication analysis at scale. In this chapter, I describe the techniques to scale up qualitative analysis of online social networks to large datasets using a reflective variant of LSA. I describe a series of automated analyses conducted on the QuitNet dataset. I also describe the development and evaluation of the automated classification system that has been used to accomplish large-scale qualitative analysis of QuitNet communication.

5.1. Topic Modeling of QuitNet Communication

Topic models are probabilistic models of the distributions of terms across topics and topics across documents (Steyvers & Griffiths, 2007). Training algorithms can help us use these models to develop new ways to search, browse and summarize large archives of texts. Latent Dirichlet allocation (LDA) (Blei, Ng, & Jordan, 2003) is a powerful probabilistic topic modeling method to extract topics from textual data. The intuition behind LDA is that documents exhibit multiple topics. This is similar to probabilistic latent semantic analysis (pLSA) (Hofmann, 1999). A probabilistic topics model was built and then explored. The system was used to detect a total of 50 topics from 16, 492 documents, consisting of the total number of individual messages in the QuitNet corpus. Table 5 shows the different terms that were assigned to the discovered topics with high probability. As can be seen in the table, the results are unsatisfactory and uninterpretable. This may be because of lack of enough semantic content in our QuitNet dataset on account of the messages being short (Zhao et al., 2011). Therefore, in the next experiment I conducted a preliminary study combining two disparate corpora to incorporate background knowledge and employed keyword based modeling to estimate semantic relatedness scores between messages and different themes. The themes were derived based on the initial analysis of 100 randomly chosen messages from QuitNet dataset (Myneni, Cobb, & Cohen, 2012).

Table 5. Unsupervised QuitNet topic detection

Topic 3	Topic 45
0.0018583393:mike	0.001728964:read
0.0018377833:said	0.0017100619:hour
0.0016942666:does	0.0016830735:mike
0.0016744763:spring	0.0016731379:said
0.0016681686:read	0.0016728827:does
0.0016680913:give	0.0016725274:nan
0.0016679233:remember	0.0016533922:ktq
0.0016621412:hour	0.0016343923:join
0.0016557919:why	0.0016296231:give
0.0016365406:home	0.0016229254:start
0.0016330811:weeks	0.0016112429:quitting
0.0016312406:nan	0.0016078708:ever
0.0015939242:dump	0.0015974957:two
0.0015845497:may	0.0015926965:may
0.0015626587:same	0.0015877625:patch
0.0015583031:flames	0.0015609124:find
0.0015453438:cloudy	0.001551453:sis
0.0015424535:patch	0.001550621:home
0.0015424473:two	0.001546045:why
0.0015399951:ever	0.0015333359:dump
	0.0015297277:spring

5.2 Keyword-based modelling

5.2.1 Methods

As illustrated in the previous section, the distributional information in our QuitNet corpus was insufficient for the automated derivation of meaningful measures of semantic relatedness between terms. This was not uniquely a problem for probabilistic models – a superficial examination of the nearest neighbors of selected terms was sufficient to establish that LSA models based on the QuitNet corpus alone were not estimating meaningful similarities between terms. Therefore, distributional information from the Touchstone Applied Science Associated (TASA) corpus was utilized (Landauer et al., 1998). The TASA corpus is a collection of 37,657 articles designed to approximate the average reading of an American college freshman. LSA was used to derive vector representations of terms in the TASA corpus. This corpus has been widely used in distributional semantics research, and when applied to this corpus LSA has been shown to approximate human performance on a number of cognitive tasks (Landauer & Dumais, 1997; Landauer et al., 1997). LSA was performed on the corpora using the Semantic Vectors package (Widdows & Cohen, 2010; Widdows & Ferraro, 2008), an open source package for distributional semantics. The Semantic Vectors package was created at the University of Pittsburgh. Researchers from various universities and corporations contribute towards its development and maintenance. It is hosted by the Google Code website, and is also available for free download as an open source resource. The software is written in Java.

The log-entropy weighting metric was used, and terms occurring on the stopword list distributed with the General Text Parser (GTP) software package (Giles, Wo, & Berry, 2003) were ignored. This stopword list consists of frequently occurring terms that carry little semantic content (See Appendix B). Subsequently representations of the messages in the QuitNet corpus were generated by adding the vectors for the terms they contain, and normalizing the resulting message vectors (called TASA-based QuitNet message vectors). Representations for terms in the QuitNet corpus were then generated by adding the message vectors for each message they occurred in, and normalizing the resulting vector. Subsequently, a second set of message vectors was generated, which is termed QuitNet message vectors. A high-level overview of the use of distributional semantics approach is as provided in Figure 8. This approach was utilized in order to ensure that terms present in the QuitNet corpus, but not in the TASA corpus, could obtain meaningful vector representations on account of their having similar distributions to terms in this corpus that did occur in the TASA corpus. This approach is similar in nature to the reflective approach that has been utilized previously to infer associations between terms that do not co-occur directly (Cohen et al., 2010). This technique provides a convenient means to combine distributional information from two disparate corpora, in this case QuitNet and TASA.

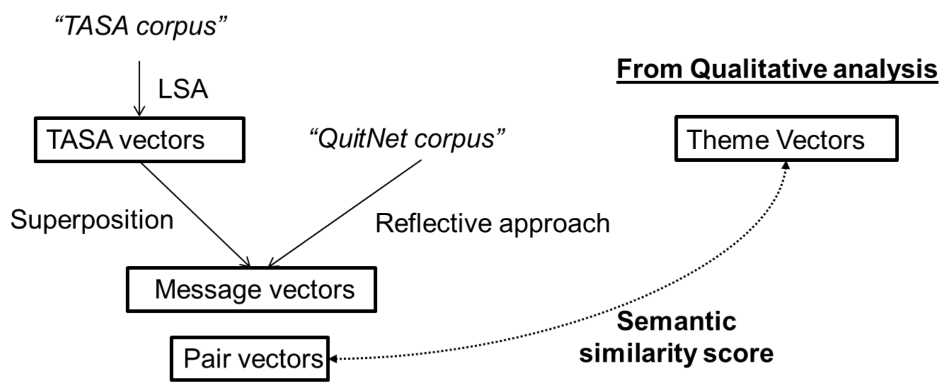


Figure 8. Distributional semantics for automated QuitNet analysis

A more detailed pictorial depiction of the vector generation is presented in Figure 9.

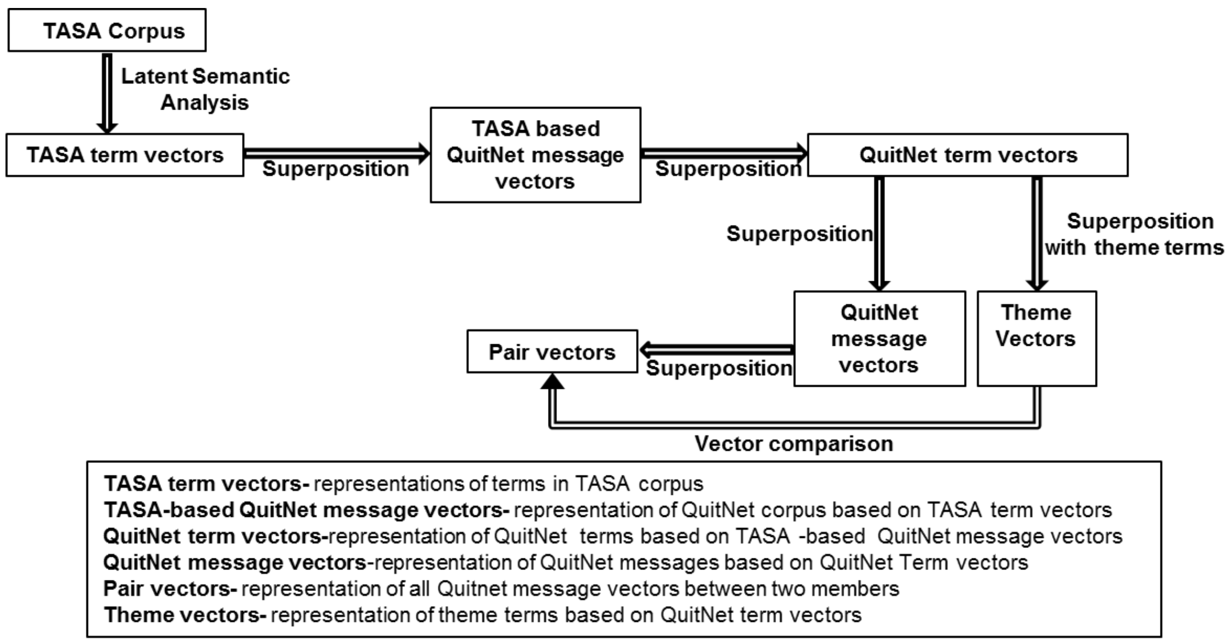


Figure 9. Vector generation sequence for QuitNet semantic analysis

5.2.2 Findings and Discussions

Inspection of the nearest neighbors of key terms from the QuitNet corpus revealed that the measurements of semantic relatedness derived using this approach were intuitive and readily interpretable. Table 6 shows the nearest neighbor terms that were derived to highlight the nature of indirect inference achieved by the system. Table 7 demonstrates the concept of indirect inference, which results from the use of reflective techniques. Indirect inference refers to a measurable semantic relation between two terms that do not co-occur directly together in the corpus used to generate the model concerned. This capability of LSA is likely address the issues with social network text data, which are predominantly unstructured, where content related to the same social concept might be expressed in several ways (e.g. “I understand you”, “I hear you”, “hold my hand”).

Table 6. Nearest neighbors of terms “craving” and “depression” Underlined terms illustrate indirect inference. Association strength is measured with the cosine metric

Craving	Depression
0.901:cigarette 0.890:nicotine 0.862:crave 0.856:craves 0.854:smoker 0.849:habit 0.847:chantix 0.841:cig	0.727:depressed 0.696:horrific 0.688:adjusts 0.669:clinical 0.669:requiring 0.656:emotions 0.645:stress 0.630:emotional

Table 7. Messages returned by the system for “Reinforcement” and “Support” themes

Themes	**Examples of Messages returned by the system**	
	Exact Match	**Indirect Inference**
Reinforcement (Key-terms: ‘congratulation’, ‘congratulations’)	That one goes in my library!!! Super quit and super ramble! Congratulation on you half year quit	Well said, XXXXX and powerful. Maybe painting a picture will reach some people who don't want to hear platitudes.
Support (Key-terms: ‘sad’, ‘emotions’)	No big cravings, but emotions are right there ready to pounce in any direction. Yesterday was let's be angry day. Today is much more mellow	Sorry, I know there is nothing that any of us can say or do to make the pain any better, but hopefully knowing that we are all here supporting you and standing at your side, will help just a little bit. You are not alone, we are always there.

The terms related to each of the themes were identified based on the qualitative analysis of a randomly chosen sample of 100 QuitNet messages. These terms were used to derive a measure of the semantic relatedness between message exchanged by each pair of communicating QuitNet members and a given theme. This was accomplished by constructing a "pair vector" for each communicating pair of users, by adding the message vector for every message they have exchanged, and normalizing the resulting vector. In addition, a "theme vector" for each identified theme was constructed by adding the vectors for terms representing this theme (see Table 8).

Table 8. Terms used for semantic similarity calculation

Theme	**Terms used for calculation of semantic similarity metric**	**Threshold**
Personal experience	experience, opinion	0.624
Adherence	pledge, bonfire, milestone	0.731
Advice	hang, help, advice	0.812
Reinforcement	congratulation, congratulations	0.786
Support	sad, emotion	0.820
Returns	cancer, exercise, weight, stats	0.685

The semantic relatedness between a pair of users and a given theme is the cosine of the angle between their pair vector and the vector representing this theme. In other words, for a random pair of QuitNet members {Q1, Q2}, there would be "T" number of semantic similarity scores where T is the number of themes prevalent in a given set of data. If the summary representation of the messages exchanged by {Q1, Q2} relates more to theme T1 than to theme T2, then semantic similarity score S_{T1}\{Q1,Q2\} would be greater than S_{T2}\{Q1,Q2\}. For example, if a random pair of members {user ID 31, user ID 274} has a semantic similarity score of 0.796 for a theme related to reinforcement and a score of 0.523 for a theme related to personal experiences, this suggests that their communication related more to the former theme. Comparison between the generated vectors (pair vectors and theme vectors) was accomplished using the Semantic Vectors package (Widdows & Cohen, 2010; Widdows & Ferraro, 2008). The similarity scores for each pair of members who exchanged messages in QuitNet relevant to a given theme were retrieved. For instance, the terms “congratulation, congratulations” were used to obtain similarity score for reinforcement theme. Unlike the frequency of communication among QuitNet members, theme-based semantic similarity scores were normally distributed and hence thresholds were applied. The average similarity score for “reinforcement” theme was 0.674 with standard deviation=0.112. The filtering threshold for each theme was different and set at average plus one standard deviation point. As seen in Table 8, the threshold for reinforcement was set at 0.786.

5.2.3 Evaluation of the system

The recall and precision of the system retrieved messages were evaluated for a given theme. A random user, S1, was chosen. Subsequently, all corresponding members that S1 exhanged a message with irrespective of the theme were retrieved. The messages were coded to see how many of them belonged to that particular theme. There were a total of 40 messages exchanged between S1 and 22 R1-R22 other members in QuitNet. Of these, exchanges between S1 and R1, R8, R14, R17, R19 included messages that belonged to the reinforcement theme. In comparison, the system retrieved exchanges between S1 and R1, R3, R8, R14, R17, R18. Thereby, the recall (number of connections retrieved/number of connections discussing theme) of the system was estimated to be

0.80, while the precision (number of connections retrieved/number of retrieved connections discussing theme) was estimated at 0.74. The same process was repeated for the “Support” and “Personal experience” themes. The recall and precision were calculated to be 0.78 and 0.86 for support theme, while for the personal experience theme, the recall and precison measures were 0.67 and 0.84. Overall, a total of 82 messages from 27 unique users were evaluated. These 82 messages were coded to see how many of them belonged to that particular theme (above threshold). On an average, the recall of the system was calculated to be 0.75 and the precision was 0.81.

5.2.4 Limitations

Precision is one important metric that needs to be considered during this analysis since low precision would lead to the formation of erroneous links in the network, which would mean that the derived theme-based network is not truly representative of the messages exchanged. The average precision of the system across the themes is 0.81 and the precision can be further increased by raising the thresholds of individual themes at the cost of the recall. This method might be limited by the terms used to retrieve messages. These terms are derived based on the qualitative coding of QuitNet data. The accuracy of the system may be further improved by more sophisticated choice of search terms, and the combination of the vector representations of these terms using vector space equivalents of negation and disjunction (Widdows & Peters, 2003). In addition, the evaluation of the automated methods in this paper considers messages exchanged by a single user in each theme only. This limitation was addressed in our next experiment, by developing an evaluation framework that builds on a large sample of messages drawn from multiple users.

In the next section of the chapter, I describe the use of a machine learning approach to partially automate the coding of the entire QuitNet dataset using themes derived from the qualitative coding described in Chapter 4. A series of experiments evaluating the accuracy and reliability of the method are described in detail.

5.3. Nearest Neighbors approach

As in the previous experiment employing keyword based modeling, representations of the messages in the QuitNet corpus – namely "QuitNet message vectors" were derived In order to use these generated vectors to support automated coding of QuitNet messages, in this experiment a *k*-nearest neighbors approach was employed. The *k*-nearest neighbor algorithm (*k*-NN) is a machine learning method for classifying objects using a set of training examples. *k*-NN is the most commonly used machine learning algorithm where an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its *k* nearest neighbors (*k* is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of that single nearest neighbor.

For each message, the system provided a ranked list of codes based on pre-assigned manual codes to the nearest neighbors. The score for a particular code was obtained by adding the cosine measures of the nearest neighbors corresponding to that code. The cosine measure represents the relatedness of the message to its nearest neighbor. All of the coded messages other than the message in question were considered (leave one out cross-validation). Figure 10 illustrates the scoring procedure for each code. As shown in the figure, the five nearest neighbors to message ID10515456 were retrieved. For each of the codes (in this case high-level themes), a score was calculated by adding up the cosine values of the nearest neighbors to which the code was attached. For instance, the score for “motivation” was obtained by adding the cosine scores of the nearest neighbors (10449020 and 10581825). These scores were used in the next stages to fine-tune the system for accuracy and reliability as explained in the next section of this chapter.

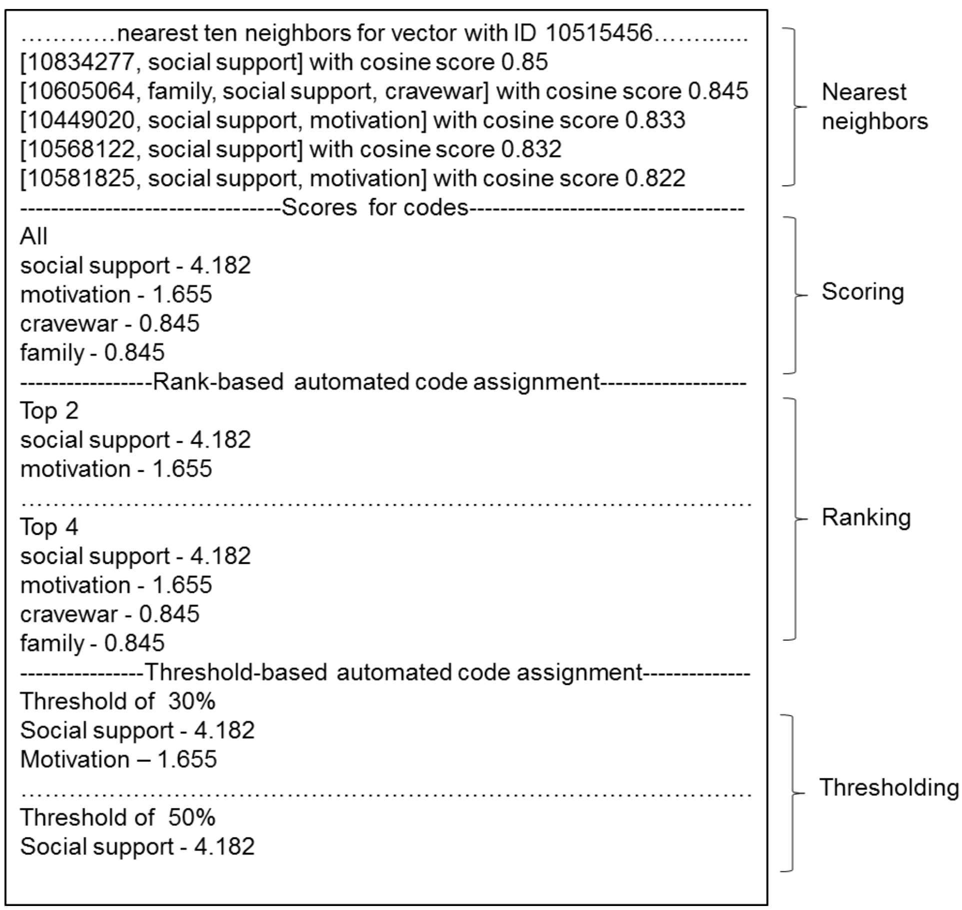


Figure 10. Overview of the scoring and optimization procedures used for automated classification system

All: list of all the codes that the system assigned to the message in question; Top 2: ranking cut-off that retains the codes with the top 2 highest scores; Top 4: ranking cut-off that retains the codes with the top 4 highest scores; threshold of 30%: cutoff based on association strength that retains the codes with scores greater than 30 percent of the highest score; threshold of 50%= cutoff based on association strength that retains the codes with scores greater than 50 percent of the highest score.

Using the LSA technique described in the methods section, vectors representing all of the messages in our QuitNet dataset were generated. Two experiments were conducted to evaluate the extent to which these vector representations could be used to accomplish the automated analysis as explained below.

5.3.1 Experiment 1: Evaluation of the System Accuracy

Methods

The dataset of the 790 manually coded messages were again coded by the automated classification system. Leave-one-out approach was taken for producing test and training sets. The system returns a scored list of codes for each message (see Figure 10). However many messages are coded with a small number of codes. So some cutoff point is required for meaningful evaluation. In the preliminary experiment we used ranking as a cutoff, but subsequently based the cutoff on association strength. In the ranking-based code assignment, we had the system rank the codes at multiple levels (e.g. top 2, top 4) as seen in Figure 10. Threshold-based cutoffs were also tested at various levels of the association strength. For instance, at 30% threshold only the codes with a score greater than 30% of the highest score were retained. As shown in Figure 10, when a 30% threshold was applied, codes with score greater than 0.3*highest score (0.3*4.182=1.255, social support, motivation) were retained. The recall and precision of the system in assigning codes (both for low-level concepts and high-level themes) was calculated at various levels of threshold and ranking. The best recall and precision values across all these cases was calculated to be 0.53 and 0.58 respectively. Based on error analysis, messages belonging to two themes “miscellaneous” and “game” were then excluded from the dataset because the message content is not amenable to content-based analysis. The code “game” was assigned to those posts where QuitNet members play a word association game with each other to engage themselves in an activity to curb the cravings. However, single word postings such as this cannot be dealt with by the system effectively as individual message content does not provide sufficient semantic context to interpret the purpose of these words. Similarly, messages that were coded as “miscellaneous” were also excluded because the content does not relate to any of the smoking cessation related themes. The experiment was then repeated with the dataset excluding the messages that belong to the “miscellaneous”, and “game” categories, leaving 533 messages.

Results

The graphs in Figure 11 provides recall, precision, and F-measure values across all parameter settings, with rankings based on k-nearest neighbors (where k=1,3,5,10,and 20) for low-level concepts and high-level themes derived from grounded theory analysis. These were analyzed separately in order to evaluate the performance of the system at different levels of granularity. In particular, we are interested in high-level themes for our purpose because most high-level themes relate to the social, cognitive, and behavioral constructs proposed by behavior change theories as described in previous sections of the paper. Having the system code high-level themes enables us to recognize patterns in the QuitNet communication and analyze them in the light of the existing theories and models of behavior change, social support, and influence. The average recall and precision were calculated to be at 0.77 and 0.71 respectively for high-level themes when considering 5-nearest neighbors at a threshold of 50%. The F-measure was found to be 0.74 in this case. The theme-specific recall and precision measures were also calculated. The most frequently-occurring themes had higher recall and precision measures than the less commonly found themes. The system performed at its best for themes such as Social support, Traditions and Progress.

Conclusion

The recall, precision, F-measures indicate that the performance of the automated classification system is reasonable for high-level themes, at a threshold cutoff of 50%.

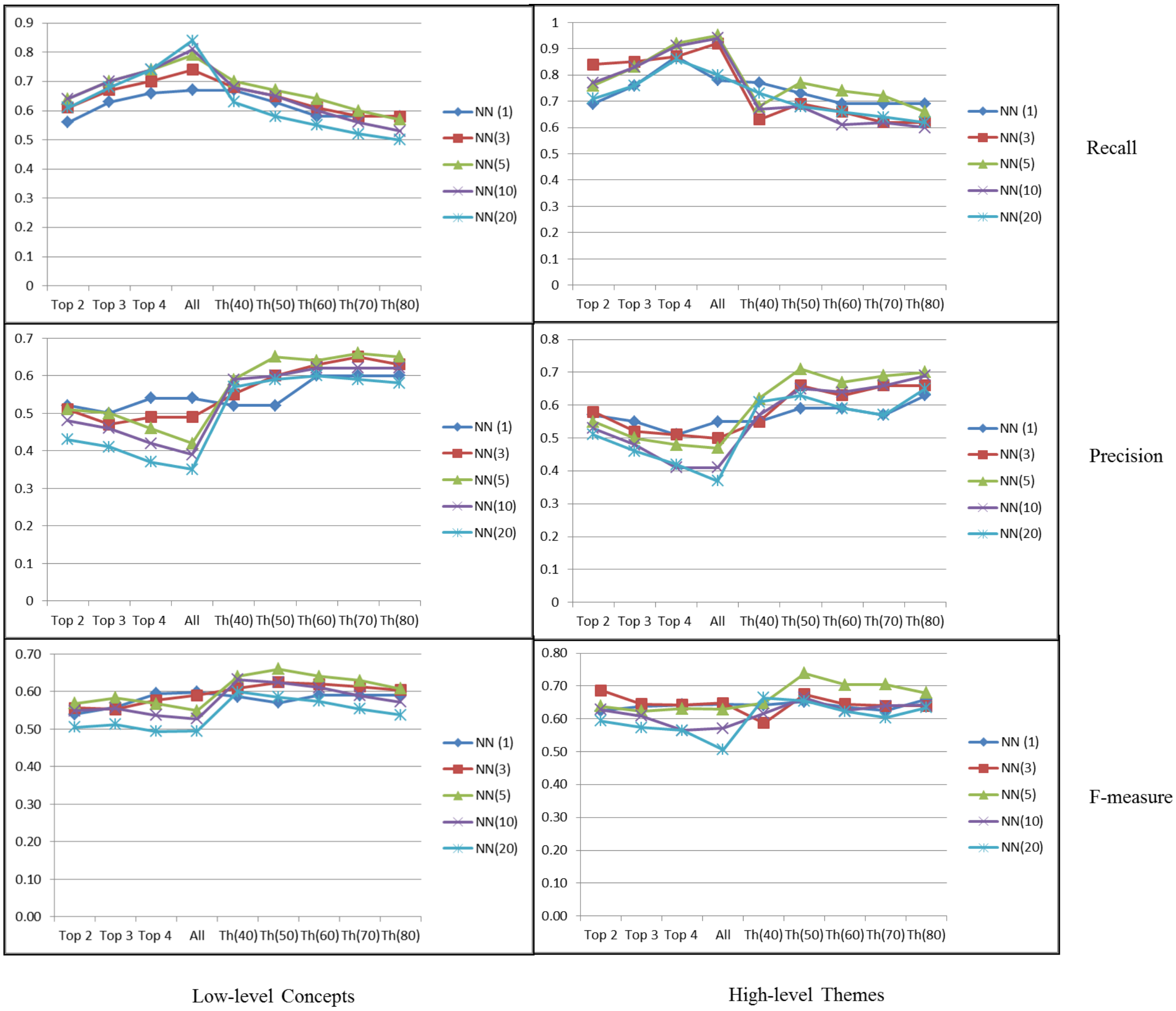


Figure 11. Automated classification accuracy metrics

The three graphs in first column correspond to recall, precision, and F-measures respectively for low-level codes. The three graphs in the second column correspond to recall, precision, and F-measures respectively for high-level themes. Top X (X=2, 3, 4) indicates the ranking-based cutoff, Th(Y), Y=40%-80% indicates association strength based threshold, NN (Z) (Z=1, 3, 5, 10, 20) indicates the number of nearest neighbors retrieved by the system to formulate scores for code assignment

5.3.2 Experiment 2: Evaluation of the System Reliability

Methods

A separate dataset of 100 messages were coded by two researchers using the high-level themes that emerged from the grounded theory analysis. The same data set was fed to the automated classification system that was optimized for better accuracy in the previous experiment. The system assigned high-level themes to these 100 messages which were then used to calculate human-system reliability using Cohen's Kappa measure.

Results

The following graph in Figure 12 provides an overview of the reliability measures. Two sets of measures were determined in this experiment, one with the messages inclusive of "game" and "miscellaneous", and the other with the dataset exclusive of these themes. Inclusive of "game" and "miscellaneous", the inter-human reliability was calculated be 0.82, and the average human-system reliability was 0.71. Exclusive of both these themes, the inter-human reliability was calculated be 0.83, and the human-system reliability was averaged at 0.77. Surprisingly, results indicated that the system agreed better with second coder, who coded just these 100 messages, than with the coder of the initial 790 messages upon which the system was trained.

Conclusion

The reliability measures obtained in this experiment indicate that the average agreement of the system with human raters for high-level themes approached the agreement between human coders.

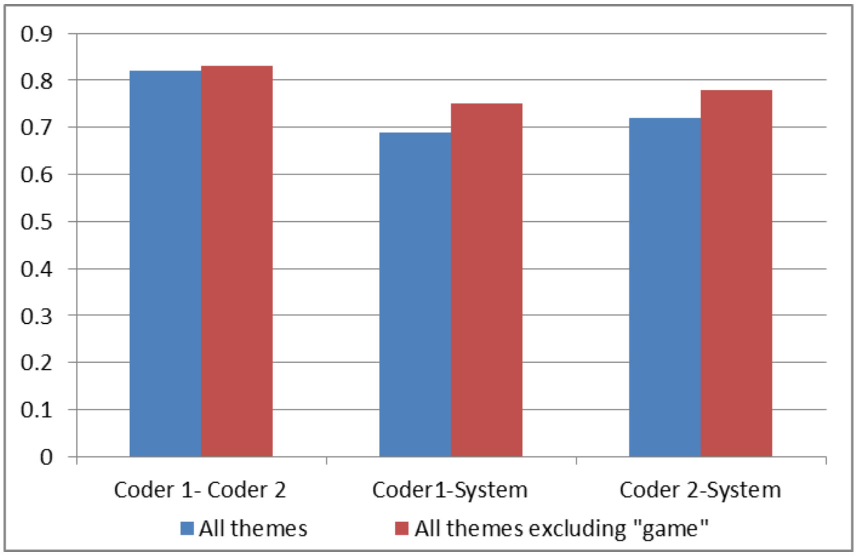
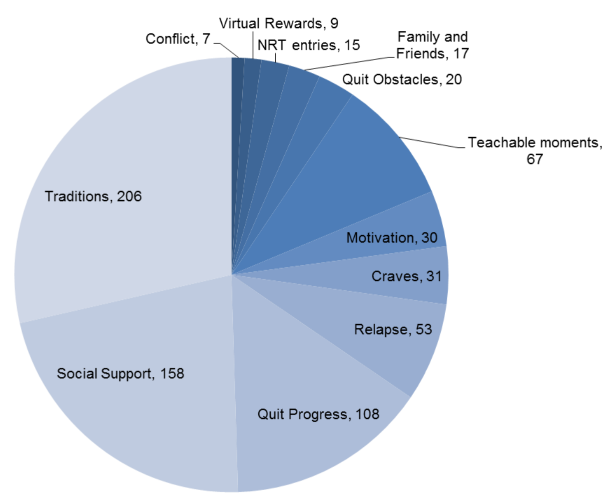


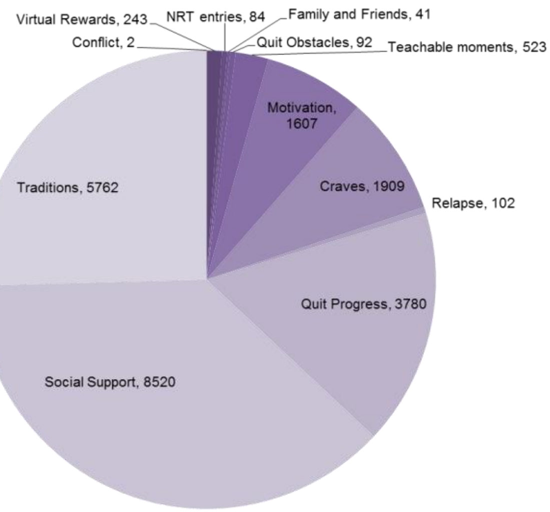
Figure 12. Reliability measures for the automated classification system

5.4. Automated Qualitative Analysis of QuitNet communication

Our entire database of QuitNet messages, consisting of 16,492 messages were processed by the automated classification system described in previous sections of the paper. Figure 13 illustrates the distribution of the QuitNet communication themes based on manual and automated analyses. As seen in the figure, the distribution of the themes from the manual analysis (Figure 13(a)) is different from the resulting distribution from the automated approach (Figure 13(b)). “Traditions” were the most commonly occurring theme in the manual analysis, while the system-based analysis indicated that the social support theme was the most frequently found. Also, “Teachable moments” and “Relapse” were the leading themes in the manual analysis when compared to “Motivation” and “Craves”, however, automated analysis of the entire QuitNet dataset suggested that the messages related to “Craves” and “Motivation” were most commonly found when compared to “Teachable moments” and “Relapse confessions”. The distribution of the remaining themes such as “Family and friends”, “Obstacles”, “Nicotine Replacement Therap (NRT) entries”, “Virtual rewards”, and “Conflict” was comparable in both the analyses. This study of the distribution of codes is presented as a proof-of-concept for our method of high-throughput qualitative analysis. However, this approach also presents the possibility of more sophisticated analyses, which we will explore in subsequent studies. For instance, QuitNet dataset can now be investigated for discernible relationships between smoking status and communication themes. This has been accomplished using a set of network analytics as described in Chapter 6. Although qualitative coding was conducted until thematic saturation, it is possible that the analysis might not have captured all QuitNet themes. Researchers can also perform reverse search on the messages to which no theme was assigned by the system to see if the content reveals new concepts that were previously not captured by the random sample also collected for the purpose of qualitative research.



(a) Distribution based on qualitative analysis of 790 messages



(b) Distribution based on large-scale qualitative analysis of 16,492 messages using automated classification system

Figure 13. Distribution of QuitNet communication themes

In addition to deriving the distribution of the communication themes in the QuitNet messages, the large-scale qualitative analysis facilitated by the automated methods allows us to understand the theoretical constructs that are embedded in each of the QuitNet communication events. Using the theme-theory matrix described in Chapter 4, an estimate of each of the theoretical constructs belonging to inter- and intra-individual behavior change models was calculated as depicted in the following graph in Figure 14. The X-axis represents the count of each construct found in QuitNet messages and the Y-axis identifies the construct. As can be seen from the graph, the QuitNet messages facilitate the transmission of behavior change constructs such as emotional coping responses, helping relationships, observational learning, stimulus-control, and self-efficacy.

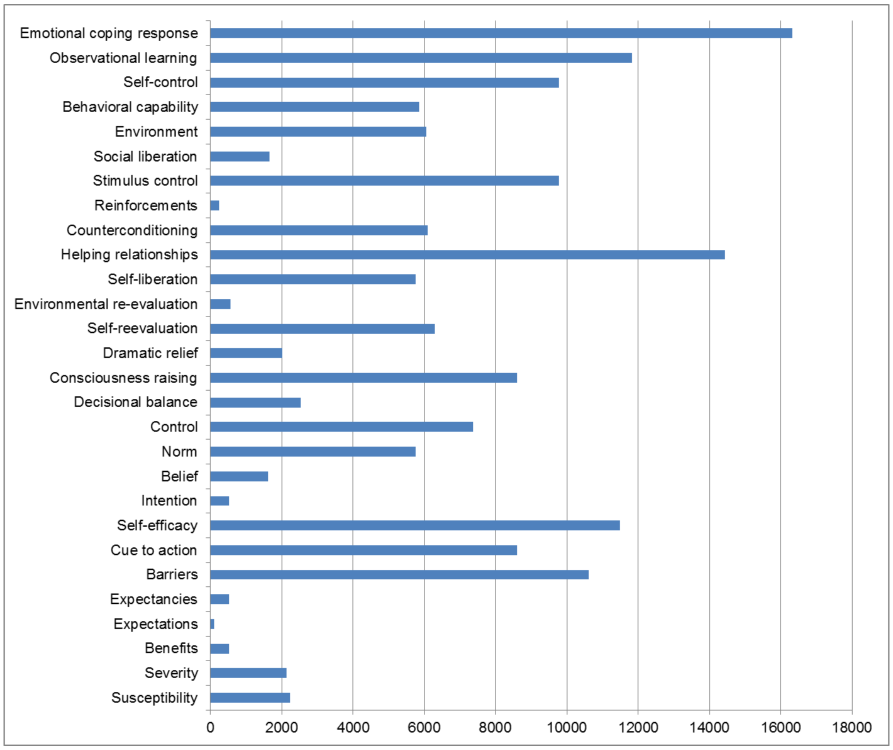


Figure 14. Instances of theoretical constructs in QuitNet dataset

5.5. Conclusions

Given the voluminous nature of online social network data, it is important to develop automated methods that can address the large quantities of free text data that are available online. Applying existing machine learning methods to the QuitNet corpus alone did not result in interpretable results. Therefore, I employed a novel approach by combining two disparate corpora thereby providing sufficient background context for subsequent employment of machine learning algorithms. While the performance of the automated method is reasonable, the accuracy of the system may be further increased by adopting more sophisticated machine learning algorithms. Importantly, the automated method provides a feasible and intuitive mechanism that facilitates code assignment to all the messages in the dataset, thus enabling resource-preserving qualitative analysis of large datasets. The reflective approach adopted in the automated classification system scales in linear proportion with the size of the dataset and therefore the automated system can be applied to analyze millions of messages available on an online social network such as QuitNet. The application of this approach to the analysis of social networks related to health and wellness is novel. The use of distributional approaches that are capable of indirect inference appears to be of value to this method, given that human beings can express similar concepts in different ways. Also, large-scale qualitative analysis of the entire QuitNet dataset serves as proof-of-concept for application of the automated method to large datasets. One implication of large-scale qualitative analysis of online social networks is that it facilitates the inclusion of semantic content into network models of online communities. The distributional semantics method described in this paper provides a convenient means to derive intuitive interpretations of QuitNet corpus and renders network content in a mode that makes it amenable for network analysis. As I will discuss in the chapters that follow, content-based network analysis of QuitNet, which is made feasible by large-scale qualitative analysis using automated methods, has been shown to yield content-specific tailoring strategies that can be used for health promotion and behavior change. Online social media provide health researchers and informatics professionals with rich health-related data. Extracting meaningful information and understanding the underlying communication patterns can add to our understanding of human behavior. While qualitative methods can provide us with a highly granular snapshot of communication patterns, manual coding of online social media data is highly labor-intensive. As I have shown, the automated methods extend the applicability of the applicability of qualitative methods well beyond the analysis of a small sample. These methods can be applied to the data generated by emerging Health 2.0 technologies and such application results in analytics that are both scalable and granular. Methods that facilitate the extension of qualitative analysis to large-scale digital health data can extend the research and application frontiers of social media, thereby further enhancing their positive impact on health-related behaviors.

Chapter 6: Content-inclusive Derivation of Social Network Models

In this chapter, I describe how we can modify existing network analysis methods to incorporate content attributes into the analytic framework of online social networks. A multitude of studies have been conducted on online social networks geared towards the understanding of behavioral diffusion and effectiveness of network-based interventions (Centola, 2010; Cobb et al., 2010; Cobb et al., 2005; Poirier & Cobb, 2012). Statistical tools, computational algorithms, network metrics (e.g. cohesion, clustering coefficient, network density, path length, centrality measures), and mathematical models have been utilized for analysis of communication patterns in a given network (Valente, 2010). However, Cobb and colleagues have argued for deeper analysis of online social networks to promote theoretically-grounded web-based social interventions (Cobb et al., 2011).

To date, most network analysis studies on health-related online social networks have focused primarily on exploring the structural and functional composition of networks without considering communication content. It is evident that different cognitive constructs that can be transmitted through content play an important role in promoting behavior change in existing behavioral science theories (Abraham & Michie, 2008; A Bandura, 2000; Prochaska et al., 2005). However, it is not known if communication in online social networks is consistent with the constructs suggested by these theories. Recent studies focused on the content of health-related online social networks and analyzed the different kinds of support available in these networks (Chuang & Yang, 2010; Hwang et al., 2010; Phua, 2013). It was found that informational support and emotional support (see Chapter 2, Section 2.4) were the two most common kinds of support. However, in addition to social support, behavior change in online social networks relates to several other socio-behavioral concepts such as reward management, stimulus control, outcome expectations and expectancies, social influence, and observational learning (Heaney & Israel, 2002). Recent survey research conducted on members of health-related online social networks reported the facilitation of these behavioral constructs by their networks (Phua, 2013). Recent studies on web-based behavior change interventions incorporating network ties demonstrated that consumer engagement is driven by social influence exerted on an individual (Poirier & Cobb, 2012) Previous studies have shed light on the concepts of peer influence, affiliation-based influence, and positional influence (Alexander et al., 2001; Freeman, 1979; Fujimoto et al., 2012). Definitions of these concepts are provided in Chapter 2. In order to understand the way in which social influence transpires in online media new methods of analysis are required.

To summarize, there have been two predominant streams of studies on online social networks. The first category focuses exclusively on content, while the second category concentrates on structural and functional aspects of a network without consideration of content. However, efforts have been made to evaluate the quality (in terms of semantic features such as syntactic and semantic complexity, punctuation, and grammaticality) of content in social media (Agichtein, Castillo, Donato, Gionis, & Mishne, 2008), facilitate social tagging, where users to annotate, categorize and share their web content using short textual labels (Fu, Kannampallil, & Kang, 2009; Gibson, 2005), and sentiment analysis (Kale, Kolari, Java, Finin, & Joshi, 2007). Few efforts have been made to bridge content-rich and content-free analyses to characterize social networks. It is my hypothesis that a hybrid methodology that takes into account both structure and content will provide insights into content-specific patterns of social networks, providing the basis for a new generation of tailored socio-behavioral interventions that are empirically-grounded. In the following sections, I describe two studies in which network analyses for conducted by incorporating information related to content into network structure.

6.1 Study 1: Frequency-based Network Model of QuitNet

6.1.1. Methods

I used the QuitNet dataset comprising of 16,492 messages to derive network models based on communication frequency of network users. Gephi, an open source network analysis and visualization (Bastian, Heymann, & Jacomy, 2009) was used to model QuitNet data. The objective of the analysis was to identify active QuitNet users and derive the keywords exchanged by those users to understand content flow.

6.1.2. Results and Discussion

Figure 15 provides a depiction of the QuitNet network based on communication frequency. A total of six different clusters were found with considerable crosstalk. In order to identify high active users and their network clusters, I further refined the model by excluding nodes with less than 5 degrees and edges with weight less than 3 and the subsequent results can be seen in Figures 16(a) and (b). Further, clustering analysis was conducted using the Markov Cluster Algorithm (MCL) implementation within Gephi.

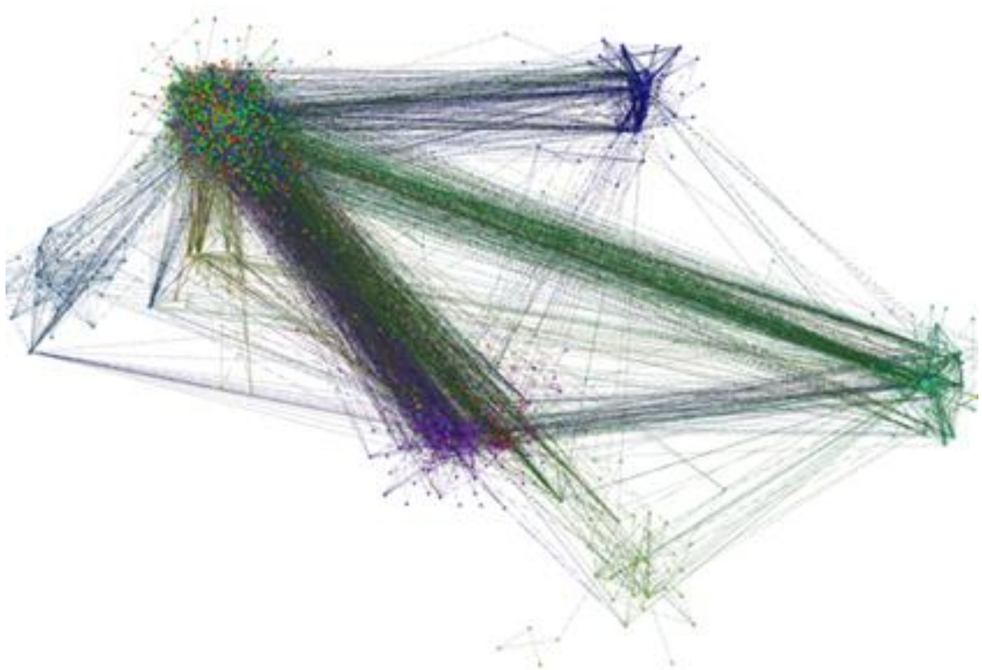
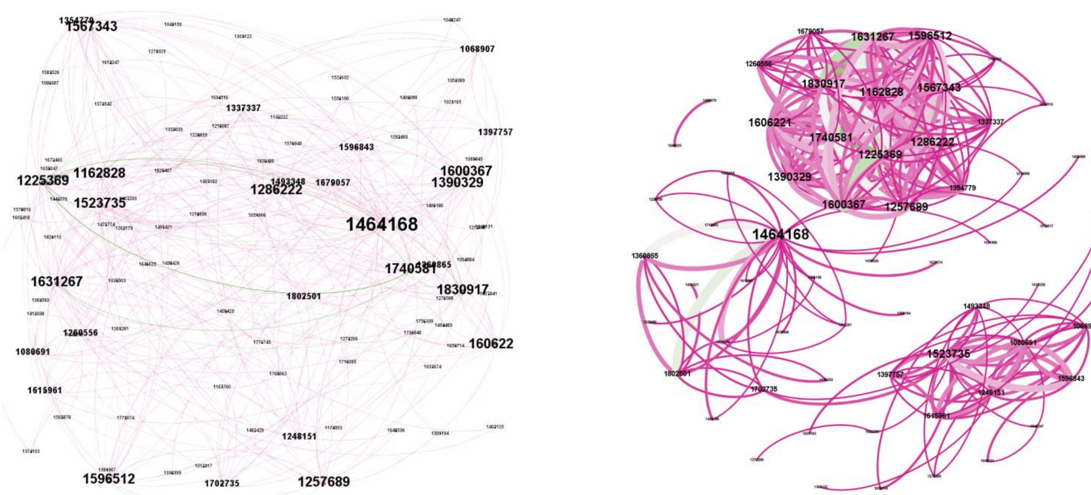


Figure 15. Network overview of QuitNet considering communication frequency



(a) (b)

Figure 16. QuitNet network models refined using communication frequency

Figure 16(a) represents QuitNet with nodes (degree>5) and edges (weight>3). Size of the node font scales with the degree of the node, where the larger the font size the more active the member is in the network. Based on the clustering analysis, three distinct clusters were found and Figure 16(b) provides a depiction of these clusters. Also, thickness of the edges in the figure represents the frequency of communication between two members in QuitNet. Based on Gephi output, the terms related to the messages posted by the most active members (based on node degree) were retrieved using the Semantic Vectors package. The vector space was queried for the representative terms of the messages posted by a given QuitNet user. The user identification information, in this case user id was used to query the term vector space. Table 9 provides an overview of the terms in the messages posted by the top five QuitNet members ranked based on their activity in QuitNet determined based on their communication frequency

Table 9. Concepts retrieved based on frequency analysis of QuitNet

Member	Semantically related concepts	
14XX0011	0.24315883	just
	0.23547024	stayed
	0.22778404	too
	0.21635884	jumpy
	0.21241876	gone
	0.20875898	ever
15XX2233	0.23674795	too
	0.23042281	just
	0.22731751	footsteps
	0.22154514	jumpy
	0.21562326	inflicted
1600XX12	0.39882264	him
	0.39585572	went
	0.391921	looked
	0.38669822	came
	0.3820126	stayed
	0.3807798	little

As seen in Table 9, the terms were not providing meaningful information to infer the communication patterns of active network users. Therefore instead of the top-down approach adopted in this study, I attempt to derive network models based on an empirical bottom-up approach and examine if any insightful inferences can be drawn from such analysis. With the majority of studies on social networks being hypothesis-driven (Laumann, Marsden, & Prensky, 1989), this approach offers a data-driven strategy for network analysis, which in turn can facilitate the development of empirically-grounded interventions. Network models described in Section 6.2 were derived by taking into account the similarity of messages communicated by a user pair (as described in Chapter 5) to a particular QuitNet theme (themes derived by qualitative coding as described in Chapter 4).

6.2. Study 2: Content-based Network Models of QuitNet

6.2.1 Methods

A network model of the QuitNet data was created by representing users as nodes, and the communication between users as edges. The edges were weighted using the similarity score for each of the messages exchanged with respect to the themes derived using qualitative coding. The statistical threshold for a theme was determined by considering the relatedness between it and each of the 6,928 pair vectors. As the distribution of these relationship scores was normal, a threshold value of the mean plus one standard deviation was used for each theme. All edges with a score above the threshold were retained for the network model of each theme, which was analyzed further using Gephi. Once the theme-based networks were derived, we then analyzed the differences in networks across themes using a social network metric - modularity.

Modularity is defined as “the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random, and therefore can be considered as a measure of the cohesiveness of communities within the network” (Newman, 2006). The modularity can be either positive or negative, with positive values indicating the possible presence of community structure. Thus, community structure in a network can be understood by looking for the divisions of a network that have positive, and preferably large, values of the modularity. In addition to modularity, other important network analysis metrics involving centrality measures (in degree, out degree), and network level measures such as clustering and average path length were considered for analysis.

6.2.2 Results and Discussion

Results obtained for reinforcement, support, and personal experience themes were as follows. The average degree of these three theme based networks was 2.911, 2.34, and 3.162 with the average weighted degree being 2.355, 1.965, and 2.088 respectively. The modularity of the networks was 0.64, 0.766, and 0.607. This indicates there was sizable small world phenomenon within the theme-based network. The average path length was 4.475, 5.072, and 4.51 respectively. Figure 17 illustrates the network structure of the theme-based networks. It can be noted that the topology and structure of these three networks is quite different with respect to density and the high-degree nodes were different across the spectrum. In contrast, when the QuitNet data as a whole (no thresholds were applied as the frequency distribution is not normal) was analyzed with the edges weighed based on communication frequency, the modularity was lower (~0.4) than that for the theme-based networks suggesting that content-based networks were more densely connected enabling effective information channelling through the network. Six large communities were found within the network with majority of the nodes concentrated in four modules with considerable crosstalk between them. These sub-communities are of limited utility for the development of intervention strategies. In contrast, sub-community clusters identified in theme-based networks offer content-specific insights (explained in Chapter 7) that are interpretable and actionable. Also, these results indicate that different theme-based networks have different opinion leaders. In the "reinforcement" network, all leaders belonged to the same community, while in other networks the leaders were dispersed across different communities. Consequently, the methods outlined in this section have implications for the design of targeted interventions, which are described in detail in Chapter 7.

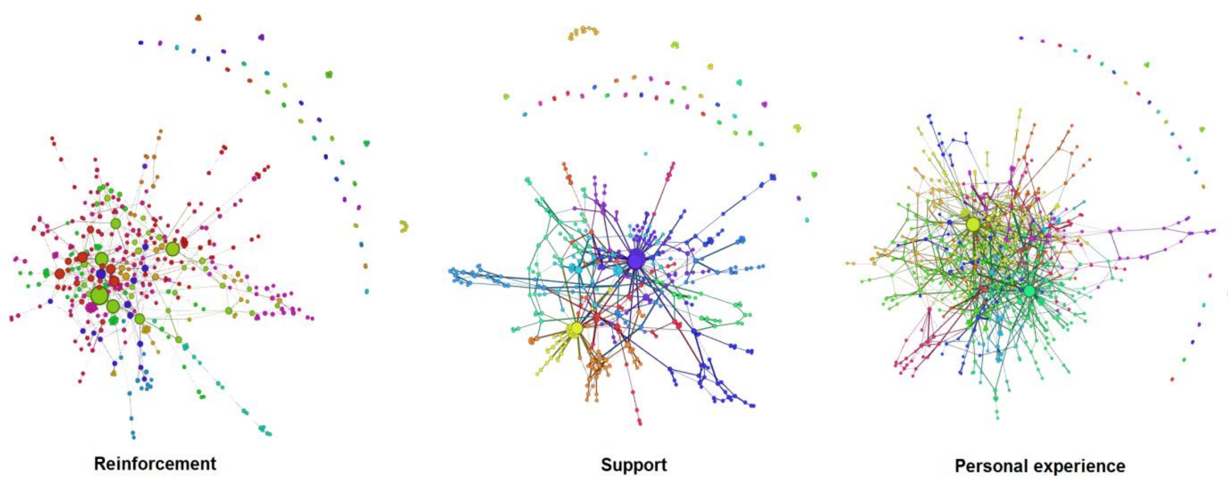


Figure 17. Visualization of QuitNet themes

While network analysis described in this section provides novel content-related network insights, the analysis can be further extended to include metrics beyond the ones described in this study. In the next sections of this chapter, I will use established methods of social network analysis to incorporate communication content into network analysis models and facilitate deeper understanding of the role of content on behavior change.

6.3 Study 3: Two-mode Network models of QuitNet

The basic premise behind these methods is that direct friendship based relations are only one component of the sources triggering social influence and that co-membership in organizations, events, and clubs can increase the likelihood of social influence. Peer influence based on affiliation conveys through co-participation or co-affiliation with other users with same behavioral attribute such as smoking, substance abuse, HIV status. Affiliation exposure models were adopted to interpret social attributes in QuitNet. Previous studies have shown the applicability of these methods to real-world networks (Fujimoto et al., 2012; Fujimoto & Valente, 2012; Fujimoto, Williams, & Ross, 2013) enabling the understanding of the additional sources of influence such as co-participation in same activities such as sports or co-affiliation based on same venues . In the next section, I frame communication of specific themes and concepts as affiliation events and QuitNet members as actors. Therefore, themes adopt the role of a sports club or event in the original use of these methods. Subsequently, the two-mode data derived will be as illustrated in Table 10. The information in this table is based upon the results obtained from semantic analysis using the *k*-nearest neighbors algorithm, as described in Chapter 5. The members were assigned to themes based on the automated categorization of their messages. Users were assigned to a theme if they have at least exchanged one message belonging to the respective theme. Thus, QuitNet members and their communication themes form the two modes for social network analysis using the affiliate exposure models. Affiliation exposure measures the degree to which each actor is exposed to a given attribute based on the extent of co-participation among the actors and can provide an understanding of how co-participation is associated with behavior (Fujimoto et al., 2012). Given that co-participation in our study is dependent on content of communication, this allows for the characterization of the role of content-specific social influence attributes in smoking cessation.

Table 10. Illustration of two-mode data (Row mode represents QuitNet users and Column mode refers to the QuitNet communication themes)

QuitNet Theme / QuitNet Member	Social Support	Obstacles	Rewards
ID00000XX	1	1	0
ID1111XX1	1	1	0
ID112233X	0	1	1
IDXXX2221	1	0	1

6.3.1 Methodological Overview of Two-mode Network Analysis

In the two-mode network data, structural variables are measured on two sets of distinct nodes (Borgatti & Everett, 1997; Borgatti, Mehra, Brass, & Labianca, 2009). Two-mode networks are referred to as affiliation networks (A_{ij}) that represent an actor's membership with an event. By multiplying an affiliation matrix A with its transpose (A_{ij}'), the resulting matrix C_{ij} (=$A_{ij}A_{ij}$') is a symmetric where off-diagonal entries count the number of events jointly affiliated by all pairs of actors (QuitNet users in the present example), and diagonal entries count the total number of events with which each QuitNet user is affiliated.

Fujimoto et al. employed affiliation exposure model that is a two-mode version of the one-mode network exposure model. Network Exposure Model (NEM) (Burt, 1987; Thomas W Valente, 1995) was used to examine network influence on smoking behavior (K. Fujimoto, Chou, & Valente, 2011). Affiliation exposure uses the co-participation matrix (C_{ij}) by multiplying C_{ij} by each user's attribute y_j (e.g. Y represents a vector comprising of a user's behavioral attribute, in our case abstinence behavior), the resulting affiliation exposure vector of F is obtained by multiplying the C matrix with the Y vector. Affiliation exposure (F) measures the percentage of events in which a QuitNet user co-participates with others embracing a certain behavior or attribute. In our application, affiliation exposure measures the percentage of themes that the QuitNet member co-communicates with other users who are attempting to quit. Note that the diagonal values (D) of $C_{ij;\ i=j}$ are ignored for this computation because they do not capture the network component of the data, but will be included as a control variable for later regression analysis. In order to examine the effect of affiliation exposure in the context of understanding multiple factors affecting behavior change, our affiliation exposure terms can be included as one of the covariates in a regression model in the following equation. The affiliation exposure term (F) is the main independent variable of the along with the number of themes each user participated (D), and other factors (X) such as gender and age. The alpha and beta values capture the error and variable coefficients respectively.

$$y = \alpha + \sum_{j=1}^{k} \beta_j X_j + \beta_{k+1} \underline{F} + \beta_{k+2} D$$

6.3.2 Materials and Methods

A total of 1423 QuitNet users and their communication themes were tabulated into a matrix, where QuitNet members were indexed in rows and themes indexed in columns. A behavioural attribute called Reported Abstinence Quotient (RAQ) was constructed to measure the smoking behaviour. RAQ was determined by averaging out the abstinence status self-reported by QuitNet members every time they log in to the system over a period of 60 days, thereby giving us an estimate of a member's self-reported abstinence status during this period. RAQ can range from 0 to 1, where '0' indicates that a member has not reported a period of abstinence during the three month period in which the QuitNet dataset was extracted, '1' indicates that a member stayed abstinent throughout. The higher the RAQ, the higher the proportion of self-reported status updates reported abstinence from smoking during the study period. In addition to the estimation of RAQ, self-reported abstinence status was used to classify QuitNet members into five groups of individuals as follows (see Chapter 4 for more details).

Group A: Members who were smokers during the study period (current smokers)

Group B: Members who stayed abstinent during the entire study period (ex-smokers)

Group C: Members who switched their status from smokers to ex-smokers (successful quitters)

Group D: Members who altered status from ex-smokers to smokers (relapsed smokers)

Group E: Members who changed their smoking status multiple times (members with multiple relapses)

Unlike majority of studies on social networks where the ties between actors were deductive in nature (Laumann et al., 1989), I take an empirical bottom-up approach in this study to explain the communication patterns within QuitNet. The themes that form the second mode of this network dataset are the ones derived from the qualitative analysis of QuitNet messages described in Chapter 4. Examples of themes include “Social support”, “Traditions”, “Craves” and “Relapse”. Demographic measures of age and gender were used in this study as controlled variables. QuitNet data were analyzed using 2-mode network analysis by creating affiliation network composed of two modes. The first mode is the set of QuitNet members (n=1423). The second mode is the set of communication themes (V=12) with which the QuitNet members were affiliated. Multiple methods were employed to examine the affiliation networks among QuitNet members and communication themes.

First, a visual representation of the affiliation networks between QuitNet members and communication themes was constructed. These networks were used to identify the major content types connecting the QuitNet members by estimating the degree-centrality metrics. Second, a *coaffiliation network* among QuitNet member groups (groups derived based on self-reported abstinence status) was constructed, where each pair of QuitNet members sharing at least one common theme is connected. Similarly, a *co-occurrence network* among communication themes was constructed, where each connected pair of themes shares at least one common QuitNet member. Further, clusters of QuitNet members were identified in a coaffiliation network using a spring-embedding algorithm. The goal was to visually represent the structural patterns of affiliation networks in terms of a member's abstinence status or type of communication content (i.e., "Social support" or "Craves"). Graphs were constructed using NetDraw plugin of Ucinet 6 (Borgatti, 2002; Borgatti, Everett, & Freeman, 2002). Finally, the levels of exposures to smoking behaviour within networks formed by QuitNet members affiliating with the same communication themes were measured. In this study affiliation exposure measured the extent to which QuitNet members were exposed to other QuitNet members who have attempted to quit smoking through shared affiliation with QuitNet communication themes. Calculation of affiliation exposure involves two components: 1) the number of themes with which a pair of QuitNet members shared affiliation, and 2) the number of themes with which each QuitNet member affiliated. Regression analysis was employed to test for an association between affiliation exposure and Reported Abstinence Quotient, a measure of an individual's aggregate smoking behaviour.

6.3.3 Results and Discussion

The demographic information of the QuitNet dataset was described in Chapter 4 in detail. More than three-fourths of QuitNet members in the dataset are female and the average age of QuitNet users in the set is 44.3 years. Approximately 70% of the users have a self-reported RAQ of '1' (indicating that they did not report smoking at all) and 22% have a '0' RAQ (indicating that they always reported smoking).

Visualization of Affiliation Network

Figure 18 represents the affiliation networks formed by 1423 users and 12 themes representing QuitNet communication. A few types of content ("Social support", "Progress", "Tradition") made up the most exchanged themes by QuitNet users irrespective of their smoking status. Considering the degree-centrality, Social support and craves were the most popular themes for Group A, progress for Group C, Social support for Groups D and E, while Traditions was the leading theme for Group B.

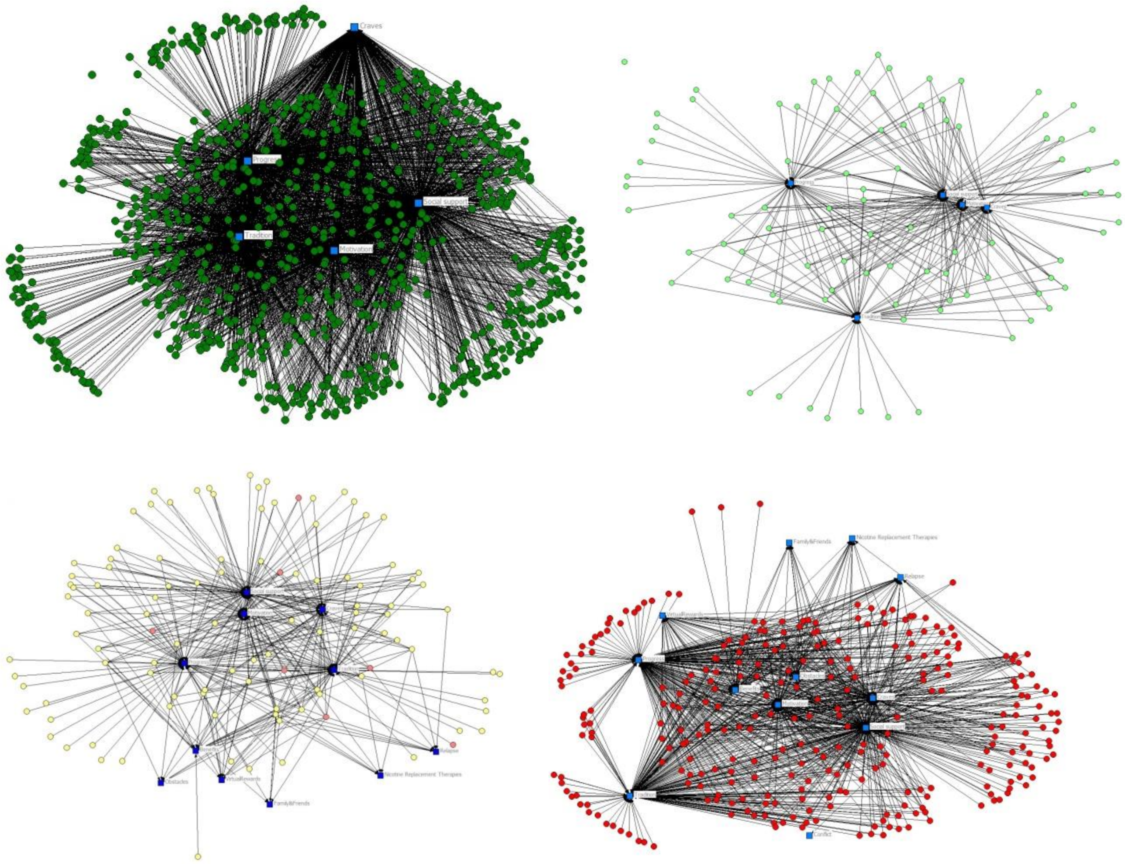


Figure 18. Affiliation network between QuitNet users

Green circles represent QuitNet users who have maintained successful quitting with abstinence status ‘1’, Lime green circles represent QuitNet users who have changed their status from smokers to ex-smokers, Light red circles represent QuitNet users who have relapsed (changed their status from ex-smokers to smokers), yellow circles represent users with multiple relapses (multiple changes in self-reported abstinence status), blue squares represent the QuitNet communication themes

The network depicted in Figure 19 has an edge connecting each pair of themes if they share at least one common user. Blue color indicates that the particular theme facilitates positive user perceptions, while pink circle indicates that the given theme facilitates either positive or negative user perceptions. The perceptions were identified based on qualitative analysis described in Chapter 4. Up-triangle indicates that a theme utilizes group-driven inter-personal mechanisms of behaviour change (see Chapter 4), down triangle means themes facilitates intra-personal behaviour change constructs (see Chapter 4) and diamond indicates individual-driven inter-personal mechanisms of behaviour change in a theme. As seen in Figure 19, thematic content where users exhibited positive perceptions tended to share more users than compared to other content types. Also, majority of the themes equipped with inter-individual behaviour constructs were heavily populated except for virtual rewards and family and friends.

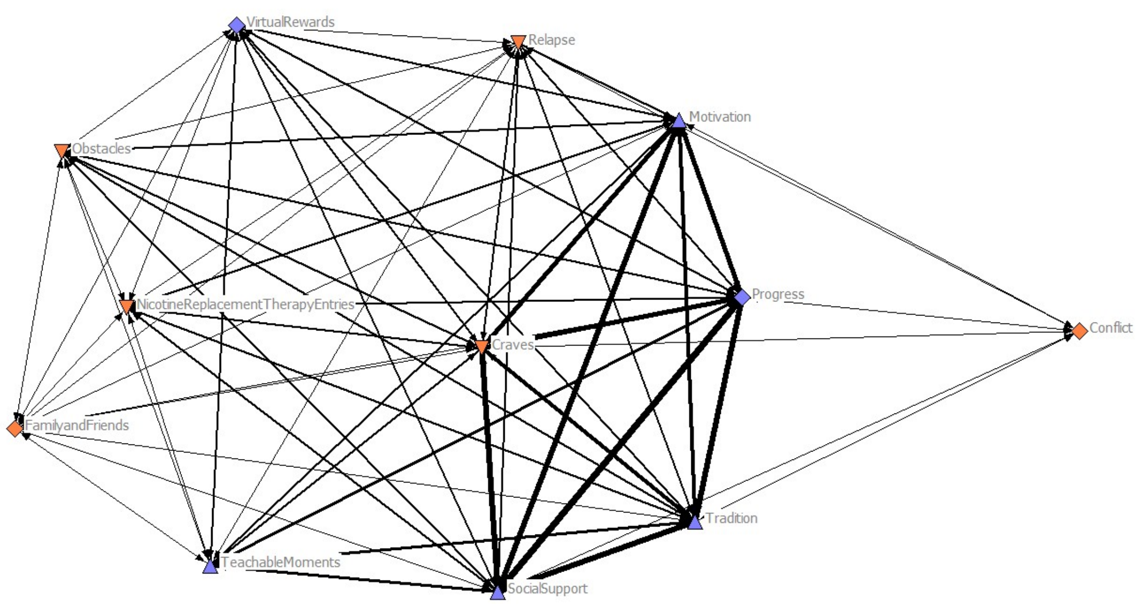


Figure 19. Co-occurrence network of QuitNet members among communication themes

Figure 20 illustrates a coaffiliation network among QuitNet users, where a user pair is connected if they exchange content with at least one theme in common. The following graph represents a network of 1156 users who have exchanged more than five messages, and not less than 25 messages. A frequency-based filtering mechanism was employed in order to derive meaningful clustering models to understand the communication patterns with respect to a given user's activity levels in the network.

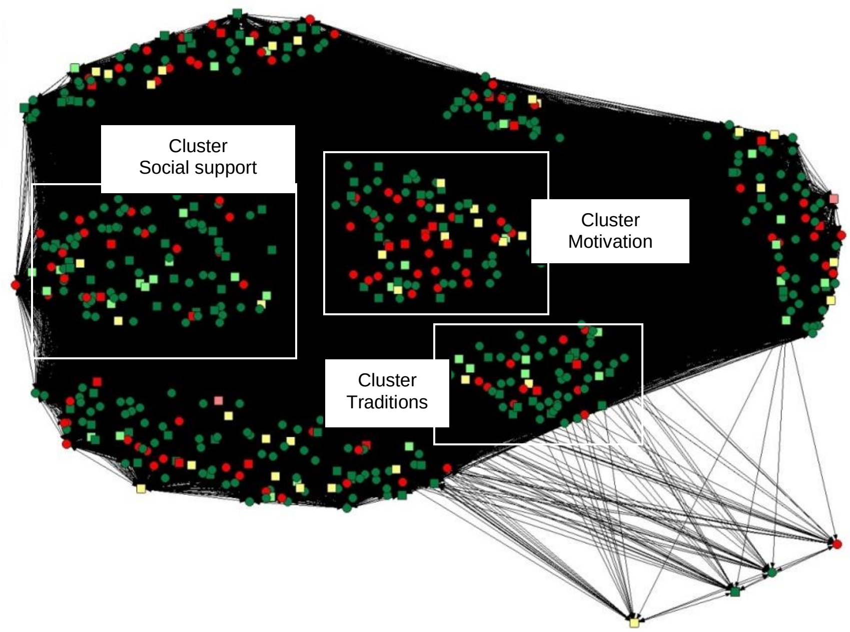


Figure 20. Coaffiliation network with 716 users who have exchanged more than 5 but less than 25 messages

Red: Members who were active smokers, Green: Members who were abstinent, Light green: Members who changed from active smokers to ex-smokers, Light red: Member who changed from ex-smokers to active smokers, Yellow: Members who changed their abstinence status multiple times; Square represents a male member, Circle represents a female member

Examining the clusters revealed similar communication content exchange by users belonging to a common cluster. Three clusters identified in Figure 20 have users who exchanged messages belonging to themes including social support, traditions, and motivation. No particular pattern related to age and genders of the QuitNet members were discernible from the derived network models. As the frequency filtering window slid to more than 25 messages, the network comprised of QuitNet users with higher activity, and no apparent clusters were found (see Figure 21). This might be because of the saturation of thematic exposure. This might have led to a mathematical artefact, increasing density of C matrix. Because when members exchanged more than 25 messages it is possible that they might be exposed to majority of the themes since message frequency in this case is greater than thematic count. Thus, leading to no distinction of clustering which might have resulted in a global QuitNet community with no localized content-based clustering.

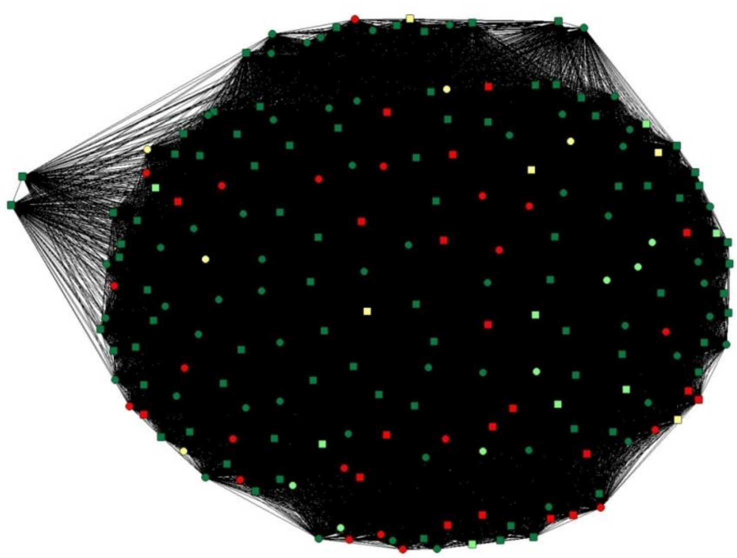


Figure 21. Coaffiliation network with 382 users who have exchanged more than 25 messages

Red: Members who were active smokers, Green: Members who were abstinent, Light green: Members who changed from active smokers to ex-smokers, Light red: Member who changed from ex-smokers to active smokers, Yellow: Members who changed their abstinence status multiple times; Square represents a male member, Circle represents a female member

Two-mode Version of Network Autocorrelation Model

I have employed the separate affiliation exposure model (Fujimoto et al., 2012; K. Fujimoto, 2012) to compute affiliation exposures measuring the extent to which QuitNet members are exposed to content-based coaffiliates with self-reported abstinence from smoking. The themes were divided into four groups based on the nature of the behavior change constructs they are related to as identified in Chapter 4: Group-centric inter-personal (“Social support”, “Tradition”, “Motivation”), intra-personal themes with focus on personal experience stories (“Craves”, “Relapse”, “Nicotine Replacement Therapy (NRT) entries”), intra-personal themes with socio-behavioral beliefs (“Benefits”, “Obstacles”, “Conflict”), and Individual-centric inter-personal (“Progress”, “Virtual Rewards”, “Family and Friends”) . In my statistical analysis, I used network autocorrelation model to account for potential autocorrelation (Doreian, 1980a, 1980b, 1981; Dow, 2007, 2008; Dow, Burton, White, & Reitz, 1984; Mizruchi & Neuman, 2008; T. Smith, 2001). I used two-mode version of network autocorrelation model (Fujimoto et al., 2011). The following table shows the computations using network autocorrelation model. The ‘sna’ package in ‘R ‘, an open source statistical analysis software, was employed for this purpose (Butts, 2008, 2010). A dummy variable was used to control for full original value of full exposure that constitute for 43% of all values of affiliation exposure terms.

Table 11. Affiliation exposure computations for Reported Abstinence Coefficient based on network autocorrelation model

Group-centric Inter-personal	
Affiliation exposure	0.041 (0.013) *
Age	0.008(0.0007)*
No. themes affiliated	0.008 (0.018)
Individual-centric Inter-personal	
Affiliation exposure	0.085(0.035) †
Age	0.009(0.0008)*
No. themes affiliated	0.03(0.038)
Intra-personal (personal experience stories)	
Affiliation exposure	0.04 (0.025) *†
Age	0.003(0.001)*
No. themes affiliated	-0.03(0.03)
Intra-personal (socio-behavioral beliefs)	
Affiliation exposure	-0.02 (0.022)
Age	0.008(0.002)*
No themes affiliated	0.18(0.049)*

**P*<0.01 †*P*<0.05, *†*P*<0.1

The effect of affiliation exposure to coaffiliates was statistically significant for inter-individual themes using a two tailed test. The autocorrelation parameter estimate showed that as QuitNet users are more exposed to other users who stay abstinent through sharing the same group-centric inter-personal themes in communication, they are more likely to stay abstinent themselves. (P<0.01). Similarly, the estimates indicated that as QuitNet users are more exposed to other users who stay abstinent by engaging in communication related to individual-centric inter-personal themes, they are more likely to stay abstinent themselves at a P<0.05. Among intra-personal themes, the parameter estimate indicated that members exposed to other abstinent QuitNet members who were exchanging messages that were facilitating the sharing of personal experiences were more likely to stay abstinent themselves, when compared to getting exposed to members sharing their beliefs about smoking behavior. The autocorrelation model also indicated that older QuitNet members were more likely to stay abstinent. The number of themes a QuitNet member exchanged messages in the intra-personal belief category is significant, indicating that the more themes a member exchanged messages in this category, the more likely a member stays abstinent from smoking behavior. It is important to note that significance of the number of themes is not a network construct. The number of themes was not significant for the rest of the thematic groups. However, their affiliation exposures were significant.

6.4 Conclusions

In this chapter, I describe the attempts to derive content-based network models by adapting existing network modelling approaches to incorporate information derived from communication content. One-mode and two-mode network analysis approaches are used for this purpose. Unlike existing studies in network science, which consider deductive relationships, this study takes an empirically-grounded approach to understanding communication-driven network patterns in behavior change.

The one-mode networks derived using Gephi allow us to derive sub-networks within QuitNet based on the communication content exchanged between two members. The network structures reveal differences across themes, the most striking of which is the difference in the high-degree nodes across the themes. These indicate those users with the most connections with whom they discuss content related to a particular theme.

Consequently, these high-degree nodes represent the opinion leaders of the network with respect to those particular themes. Opinion leaders play a pertinent role in social mobilizations and social networks, they act as gatekeepers for interventions, help change social norms, and accelerate behavior change (Valente, 2003, 2006). It is shown to be very important to identify opinion leaders in order to facilitate information diffusion in a network in social network and social mobilization (Obregón et al., 2009; Rogers, 1995) Previous work on QuitNet also emphasized on the importance of "social integrators" for network formation (Cobb et al., 2010). The one-mode models derived in this chapter, thus enables us to identify opinion leaders in a large social network. In the "reinforcement" network, all leaders belonged to the same community, while in other networks the leaders are dispersed across different communities. Such insight into distribution of key nodes can inform the design of targeted social interventions related to dissemination of specific-theme related information to opinion leaders. In addition, large-scale personalized interventions can be delivered to a given module of members in accordance with this community's interest in a theme as will be discussed in Chapter 7.

The study employing two-mode methods examined the structural features of content-based QuitNet members' affiliation with different kinds of communication content. My results show that members in these content-based affiliation networks are inter-connected by virtue of all kinds of content, with certain kinds of content being more popular in certain QuitNet member groups. The themes facilitating group-centric inter-personal behavioral constructs and a positive perception among QuitNet members were populous when compared to other themes. In addition, distinctive patterns of clustering characterized by common communication content were found among QuitNet members irrespective of age and gender. With respect to factors affecting content-based network influence, results indicate that those members affiliating with members exchanging content related to group-centric inter-personal themes tend to stay abstinent from smoking behavior. Members who directly communicate about health belief concepts are also more likely to stay abstinent. Findings from this study suggest new directions for the development of network interventions in public health that focus on content-based targeted behavior change strategies. In addition, this research shifts the direction of intervention tailoring from being solely deductive and theory-based (Champion, Foster, & Menon, 1997; Strecher et al., 1994; Thoreau, 2002) to being inductive and end-user needs' based at individual and population level.

Chapter 7: Empirically-grounded Intervention Strategies Harnessing Social Relationships for Positive Behavior Change

In Chapters 4-6, I describe a set of methods that are qualitative, quantitative, and automated in nature to analyze the communication content of online social network data. In this chapter, I describe a set of empirically-grounded strategies for individual lifestyle support systems that can be operationalized over the web using mobile health technology. The development of these intervention approaches was facilitated by content-inclusive analysis of QuitNet social network data. The proposed strategies range from simple message design to sensor integration and network structure re-engineering. These application areas provide a portfolio of new evidence-based ventures that provide new directions for research on interventions that health researchers and technology developers can undertake to change human health behavior.

7.1 Application Area 1: Targeted Health Promotion Interventions

Inclusion of content attributes into network models of QuitNet resulted in valuable insights into communication patterns that may enable us to better tailor network-related interventions. Application areas of network interventions include a) identification of opinion leaders, clusters, and group-specific opinion leaders, b) “rewiring” networks to improve or reduce network cohesion, and c) network-attribute interventions (Valente, 2012). Content attributes can be used to derive network-attribute interventions, where members exchanging messages related to a particular theme will be segmented as a group to harness the positive effects of their social influence. In this work, I utilize the findings derived from hybrid methodological application to online social network data to formulate new intervention strategies that can be implemented in real-time as part of network interventions. Examples of these approaches include identifying content-specific key players and creating mentor-mentee relationships based on the needs of the mentee and interests of the mentor.

The one-mode network structures obtained from formal network analysis using Gephi reveal differences across themes, the most striking of which is the difference in the high-degree nodes across the themes, which indicate those users with the most connections with whom they discuss content related to a particular theme. Consequently, these high-degree nodes represent the opinion leaders of the network with respect to those particular themes. Identifying key players within groups was shown to be one important step for effective in tobacco control (Puska et al., 1986). Opinion leaders play a pertinent role in social mobilizations and social networks, they act as gatekeepers for interventions, help change social norms, and accelerate behavior change (Obregón et al., 2009; T. W. Valente & Pumpuang, 2007).

The opinion leaders identified in my work were within a group of members exchanging information related to a specific topic of interest such as “Support”, “Reinforcement”, and “Advice”. For example, the identification of those network members who are key players in providing “Relapse assistance” and “Motivation” can help us make the right connections with users discussing about “Craves”, thus improving the network’s assistance to its members. This new knowledge about content-specific opinion leaders can be transformed into a content-sensitive targeted intervention by incorporating new support features into a social network for providing guidance information to network users with respect to content variety and content-specific opinion leaders. For example, if a network member exchanged messages related to “Progress”, then that member can be directed toward similar content types and the opinion leaders for that particular content. In addition, the “Progress”-related opinion leaders can be alerted about the new member to facilitate a connection between this member and an opinion leader. Similarly, if a member posts messages that indicate “Conflict”, trust-related issues with another member, then directing them toward messages indicating “Social support” and “Motivation” may be of assistance. In terms of factors affecting content-based network influence, results indicate that exposure to abstinent members exchanging content related to group-centric inter-personal themes (e.g. “Social support”, “Tradition”) tend to stay abstinent from smoking behavior. Therefore, online interventions can incorporate an explicit display of member profiles contributing to such content to enhance affiliations to these people. In the context of offline interventions at population level, public support messages incorporating content features highlighting the need to seek social support and be part of a group-based smoking cessation endeavour can help the general public (confronting similar issues) become involved in a support community to sustain abstinence.

7.2 Application Area 2: Design of Individualized and Network-based Persuasive Technologies for Promoting Health-related Behaviour Change

The field of persuasive technology suggests the use of computers as tools that provide people with new or enhanced abilities that enable them in initiating and/or sustaining a behavior change by reducing barriers, changing mental models, and/or increasing individual’s self-efficacy (Fogg, 2002). Online social networks form the basis for decentralized peer-to-peer support by acting as virtual communication platforms that allow users to share experiences, ask questions, and provide emotional support, and advice to one another. In other words, the triggers to elicit and/or sustain a behavior change can come from multiple nodes (peers). In contrast, web-based interventions that are fully automated and technology-driven are examples of centralized support where triggers can come from only one node (the system). There can be a third kind of support infrastructure which is a hybrid of decentralized and centralized structure as found in moderated social networks, where the interactions can be peer-to-peer but there is a high-level hierarchy that guides the communication. Messages in decentralized networks do not necessarily arise from a single authority figure as is often the case with centralized systems. Thus the persuasive content arises in an organic form with many messages arising in reaction to others. Therefore, every message between decentralized network members can act as a vehicle for persuasion. In this section, I present a new framework that enables us to understand the persuasive qualities of messages exchanged in QuitNet like online social network. By learning the existing persuasive characteristics, efforts can be made to support enhanced persuasion in these digital communication media at the level of individuals and groups. For the purpose of mapping the content analysis conducted in Chapter 4 to persuasive technology design concepts, I adopted the Persuasive Systems Design (PSD) model provides a set of persuasive strategies in multiple categories of system-dependent attributes (Oinas-Kukkonen & Harjumaa, 2009).

The mapping process forms a bridge between theoretically-validated themes and persuasive qualities. In order to account for the nature of the network content, the proposed framework deals with three categories of the PSD model- primary task support, dialogue support, and social support. Table 12 shows the framework to map theoretically-grounded themes to persuasive qualities. Consider the theme "Virtual rewards" which was derived from messages in which members discuss virtual gifts such as bracelet to celebrate their achievements. While its connection to the "rewards" strategy in the PSD model is apparent, this theme has additional implicit persuasive attributes such as:
1) the elicitation of observational learning in a member when he/she watches other members receive this reward, 2) getting virtual rewards brings recognition to the network members, 3) peer recognition in turn provides a simulation cue to other members and help sustain their decision to quit, 4) it allows individuals to focus on small scale tasks, ultimately supporting them as they advance towards their target behavior, 5) the reward gets tailored and personalized based on the member's timeline. Similarly, other themes were mapped to the persuasive qualities in the PSD model. The QuitNet themes and persuasive qualities were then re-analyzed using the framework to incorporate self-reported smoking status information. Using the framework, I employed an additive model and estimated the number of persuasive qualities each group is exposed to during QuitNet communication events. More details about this analysis can be found in (Myneni, Iyengar, Cobb, & Cohen, 2013). The proposed framework is our first step towards translating empirically-driven analysis of Quit communication to understanding the

persuasive qualities of contemporary communication platforms. Future studies can use this framework to assess and inform the design of virtual support systems.

The results can be translated into intervention development solutions such as network activities with implicit persuasive behavioral constructs as illustrated below. Given that the results from this study indicated that community-specific customs such as traditions (bonfire, pledge) supported many of the persuasive qualities, I propose a network activity that triggers the consumption of multiple content types (such as messages encompassing group-centric inter-personal and individual-centric inter-personal behavior change constructs), and facilitates network connections to members communicating such content. Affiliation to members communicating such content exposes QuitNet members to abstinent behavior, as described in Chapter 7.

Table 12. A framework to identify persuasive qualities in a decentralized social network

	Virtual rewards	Quit obstacles	Quit benefits	Motivation	NRT entries	Conflict	Craves	Quit progress	Social support	Traditions	Relapse	Family & Friends
Dialogue support available												
Praise									x			x
Rewards	x		x									
Reminders					x			x		x		
Suggestion				x					x			
Similarity								x				
Liking												
Social role				x					x	x	x	x
Social Support												
Social learning	x	x					x	x		x	x	
Social comparison						x		x		x		x
Normative influence			x			x	x				x	x
Social facilitation										x		x
Cooperation										x		
Competition						x						
Recognition	x									x		
Primary Task Support												
Reduction	x							x		x		
Tunneling				x					x			
Tailoring	x	x	x	x	x		x	x	x		x	x
Personalization	x	x	x	x	x		x	x	x		x	x
Self-monitoring		x						x				
Simulation	x		x							x	x	
Reharsal												

I present a social strategy to improve member engagement in the QuitNet and facilitate meaningful connections to peers and meaning affiliations to content (see Figure 22). This can be accomplished by the construction of virtual nurseries that depend upon users' healthy behavior for their wellbeing, in this case abstinence from smoking behavior. A member can grow a flower garden, where the size of the flower links to the progress of this member as measured by their Reported Abstinence Quotient (RAQ). Members achieve the ability to grow different types of flower when they communicate certain types of content (benefits, obstacles), and making connections with members already communicating about content such as Social support or Tradition lets them collect various colors of flower. The size of the flower will grow in proportion with RAQ (positive reinforcement). However, the intervention does not allow for negative rewards by shrinking the flower if RAQ goes down with time. Studies indicate that the value of these games is in their ability to improve member engagement by providing only positive reinforcement, but not negative rewards, especially in the case that members are not meeting their own expectation to foster long-term behavior change (Lin et al., 2006). Each individual's flower garden will then be part of a village or a networked forest, which flourishes in accordance with individuals' collective efforts. The representation of 'flower' used in this strategy can be tailored to gender and age of the network demographics, which in the case of QuitNet community is predominantly female (35-49 years).

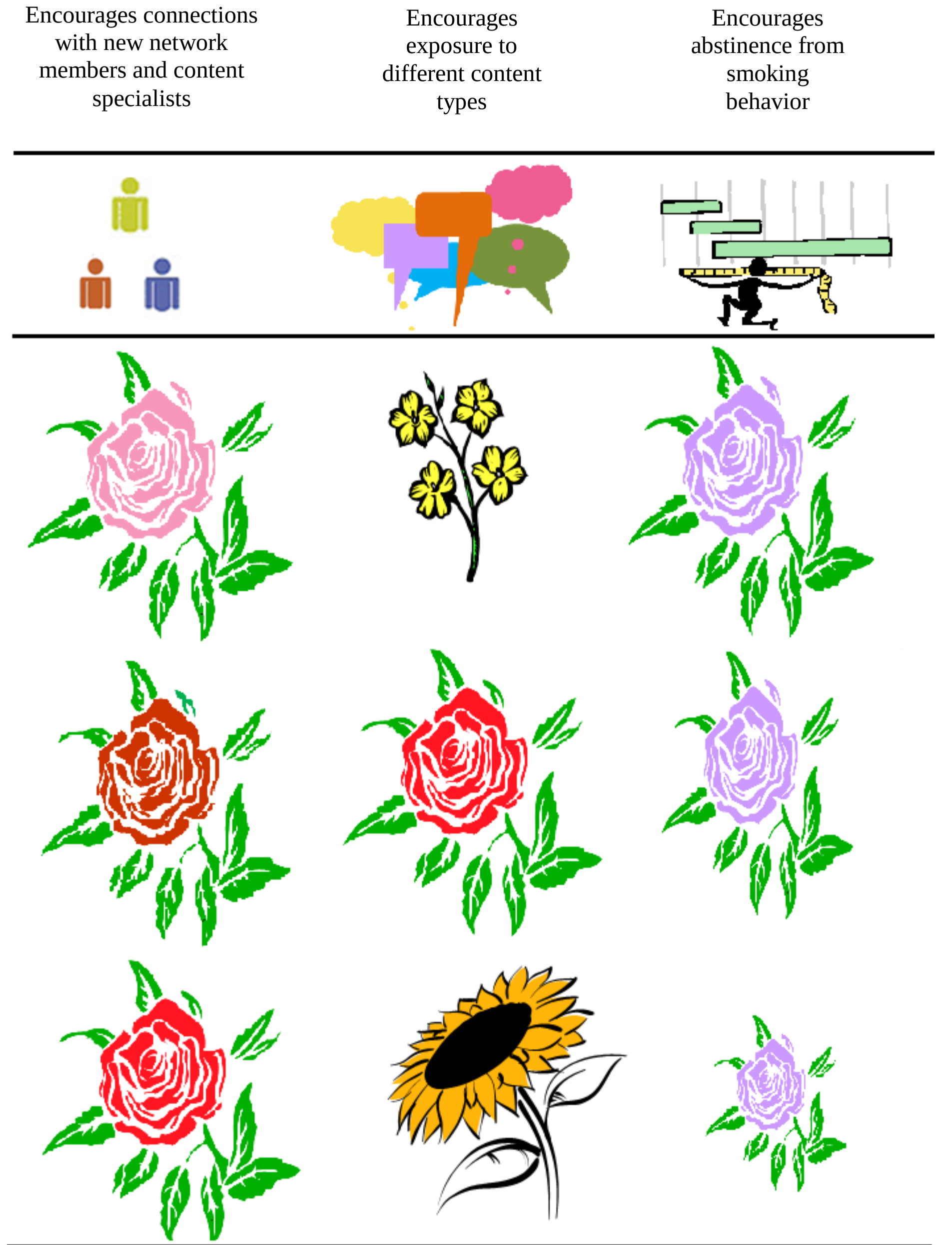


Figure 22. Persuasive strategy facilitating affiliation exposure and content exposure in the context of an online social network

7.3 Application Area 3: Convergence of Health Data Analytics, Engineering Technology, and Socio-behavioural Dynamics to Human-centered mHealth Product

One potential limitation of the forest flowers application (described in Section 7.2) is that it depends upon self-reported smoking status, to motivate for the use of sensors instead. The proposed solution contains multiple components and can be executed in phases. The major components include a hardware sensor and software application with an embedded social network theme. A brief functional mapping of internal components is as follows.

Table 13. Function mapping of proposed mHealth solution

Sensor	Will be an external hardware accessory to the	Mobile phone
Sensor	Will be interfaced with external world using	Application
Mobile phone	Brings portability and ecological validity to the	sensor

7.3.1 Sensor Accessory to a Mobile Phone

One major problem in most of the smoking cessation studies, including randomized control trials is self-reported smoking status (Stone, Shiffman, Atienza, & Nebeling, 2007; Whittaker et al., 2009). Like measurable metrics in other behaviors such as physical activity (number of steps taken, calories spent), diet (caloric intake), carbon monoxide (CO) levels are often used as a means to objectively measure smoking behavior. Oftentimes to attempt to better capture real-time data, Ecological Momentary Assessments (EMA) are used to verify smoking status of subjects., which employ repeated collection of real-time data on users’ behavior and experience in their natural environments However, they are again self-reported and are prone to biases (Whittaker et al., 2009). In addition to evidence from literature, qualitative analysis of QuitNet data reveals that members become wary of other members' smoking status reports due to trust issues (described in Chapter 4). As part of the project, input from four established behavioral scientists and intervention specialists was solicited to understand the for a robust objective measurement device that can be used as an accessory to the mobile phone facilitating EMA of an individual's smoking status. Excerpts from these domain experts are embedded in the following table.

Table 14. Expert input for mHealth smoking cessation support system

Expert	**Feedback**
Behavior Scientist 1: MD Anderson	It is quite important to get regular behavioral data for a smoking cessation program to work. There are a few CO monitors, but they are bulky, costly, and subjects have to come in to get their smoking status verified, which is not very helpful. A pedometer like device will be helpful, and if it can talk with a cell phone it is even better. You can even make them blow, an automated measurement might not even be necessary
Pulmonologist: Baylor College of Medicine	It will be good if we can verify if a smoker is really not smoking, a mobile based device will suit the need. But, what will you do about zero CO measures, CO will quickly plateau
Behavior Scientist: UT School of Public Health	1000's of clinical trials happen every day in Houston and we really don't know which one works because we can't say if the smoker is abstinent. Without measuring abstinence status, it might be difficult to implement any of these interventions on a large scale, it like hitting target in the dark and thinking that we have hit the bulls-eye

Based on initial brainstorming, a list of functional requirements of a mobile sensing technology is created- (1) Portable measurement of smoking status, (2) Real-time data collection as people go about their lives, (3) Accounting for calibration errors, (4) Eliminating the aspect of social stigma, (5) Information sharing with clinicians, and (6) First version should aim at “willing to change” population.

7.3.2 Software Application (Inspired by Lessons from Social Network Analysis)

While the application will be used to present the sensory data to the user promoting behavior change, the functionality can be further optimized by incorporating theory-based information visualization and network science models. For example, QuitNet messages analyzed using content analysis techniques and social network analysis models can be used to devise certain collaborative strategies that can be implemented in the application to harness the socio-behavioral dynamics of social networks as described below.

1. Providing automatic triggers to opinion leaders when a member is becoming dissociated from the network, or when a new member joins the network.
2. A social gaming strategy to create connections among members across sub-networks using automated messages to initiate a conversation.
3. Providing a tracking system to opinion leaders to keep track of their connections, and sending them theme-specific information to help their followers.

7.3.3 Mapping Sensor and Software

In addition to providing objective measures of smoking, the reach of the sensor can be extended by incorporating it as a driver of the software application. Sensor data can be used as a validation and motivational measure. Combining it with the software will allow the user to share their statistics with other users using social gaming techniques such as leader boards, which can play a key role in adherence to the smoking cessation regimen. Sharing these data with their network of close connections comprising of clinicians, family, and friends can serve to improve self-efficacy and reinforcement. However, one major problem with the sensor data is that with 3-4 days of abstinence, the CO levels can be zero in exhaled air. Because of this the user might lose interest and abandon the sensor. Therefore, I propose the construction of a normalized metric (SM) that is based on sensor measurement of CO levels (C), temporal trend of measures (T), and data sharing within a social network (P). A calculation in the lines of SM= C+T+P can be used to give a smoking cessation score to every participant. This way, the score will never hit a plateau and can be used to give digital rewards to the user. The variables of this metric and their association still need to be refined. A detailed association among the individual components of the proposed solution is depicted in Figure 23. The work is currently under progress, sensor selection and calibration has been completed.

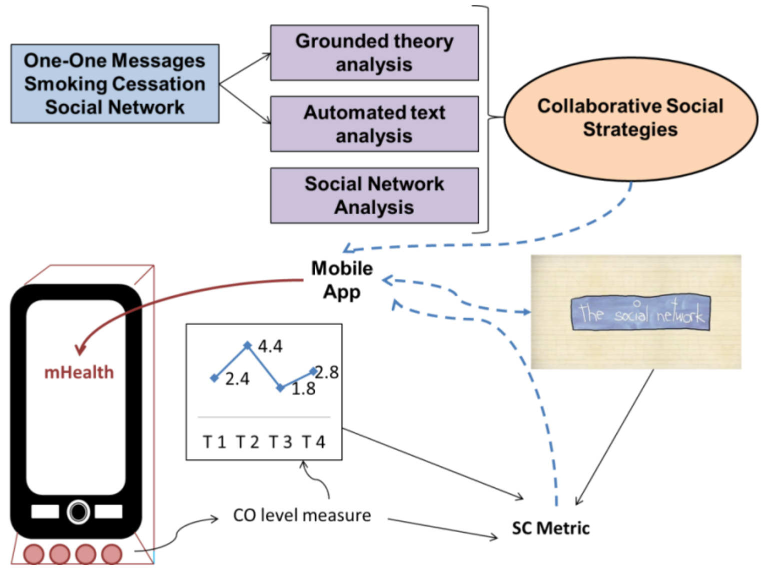


Figure 23. Overview of mHealth personal support technology for smoking cessation

7.4 Conclusions

In this chapter, I presented novel strategies to help to tailor and customize network and standalone interventions. Such research strategies will mediate the origination and refinement of a new generation of translational interventions in public health and behavior science. The development of scalable interventions to help people engage in healthy lifestyles has been a high-priority task for health researchers and professionals. The proposed strategies can enable us to efficiently channel meaningful and critical health-related behavioral information to members of online social network communities such as QuitNet. Possible implementation architectures for the proposed strategies have also been discussed in this research in the context of face-to-face and digital platforms, and at individual and population levels.

The ability to identify critical nodes, content-specific network patterns and influence mechanisms can have implications for how we build and maintain support networks for health behavior changes. Some of the strategies discussed in this chapter describe how mHealth advances can be used to enhance network interventions. In addition, the discussion of persuasive technologies demonstrates how research approaches can be mapped to, and integrated with other areas of health technology development that help us improve patient engagement and empowerment in sustaining a positive health behavior change.

Online social networks provide rich data sources that can be used to understand behavior change. These lessons can be subsequently used to design new support mechanisms and improve existing public health promotion campaigns and community-based efforts by emphasizing on certain content types. Also, online social networks can provide us with a platform to understand user needs. Consequently the design specifications of user-centered technology design can be derived from social network data. Interfacing the mobile sensor and social network can advance smoking cessation interventions to the next level by improving their scalability, personalization, and capacity for dissemination. Self-reported smoking status was found to be a problem in majority of interventions, while social networking, personalized tailoring, and persuasive gaming appear to be suitable methods that can be effectively operationalized for smoking cessation support using mobile health solutions. The multi-component collaborative solution proposed in this chapter addresses the problems with self-reporting and leverage the new insights (e.g. content-specific sub communities, opinion leaders, and network influence) in the context of a mobile social network.

Chapter 8: Summary, Limitations, Conclusions

8.1. Summary

Health-related behaviors such as smoking contribute to the majority of deaths in the United States and around the world. The development of scalable interventions to help people engage in healthy lifestyles is a high-priority task for health researchers and professionals. Online social networks allow us to examine the role of social relationships in influencing health behavior at high granularity. The members of these networks, such as QuitNet, represent a multinational community of users, which brings an international perspective to the analysis. Analysis of voluminous social network data requires methods that can provide a granular view of data. Figure 24 provides detailed information about my research strategies and the reasons underlying their usage. First, communication themes in QuitNet data were derived using grounded theory techniques (Strauss & Corbin, 2008). The derived themes were compared to key constructs related to existing behavior change theories. Combining automated text analysis methods from distributional semantics (Cohen & Widdows, 2009) with the *k*-nearest neighbors machine-learning algorithm, the qualitative coding was extended to the entire QuitNet dataset. At this point, the members were connected based on their relatedness to specific message content. These semantically connected networks were then analyzed using modified affiliation exposure based network analysis methods (Fujimoto et al., 2012). Two-mode data also called duality data (Borgatti & Everett, 1997) consisting of information regarding members communicating messages belong to same theme were used for further analysis to understand the communication patterns among QuitNet members based on the content they communicated, and a behavioral attribute: self-reported abstinence status. These patterns were then used to propose novel intervention strategies that are targeted. The underlying communication patterns among QuitNet members can also inform individual lifestyle support systems that can be operationalized in real-world settings using mobile health applications.

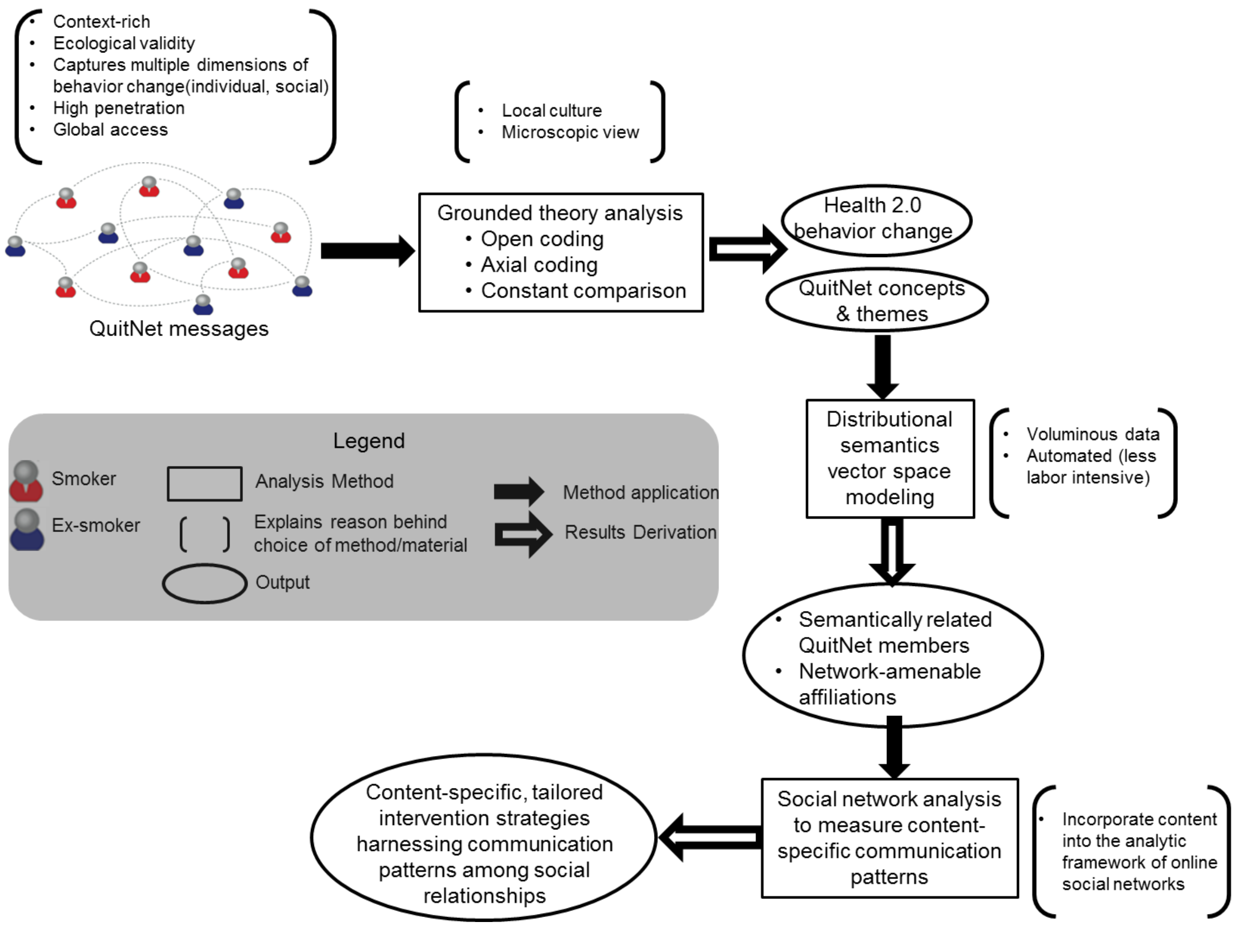


Figure 24. Research strategy overview

8.2. Limitations and Future Work

The QuitNet dataset considered in our analysis was recorded in 2007, and is limited in size. For future studies, attempts will be made to obtain further datasets drawn from recent data. Although the messages were qualitatively coded with the aim to achieve thematic saturation, it is possible that additional themes were not captured. This qualitative analysis provided useful insights into QuitNet communication themes. However, manual coding is highly labor-intensive and time consuming. Consequently, analysis is limited to small sample sizes, potentially limiting the generalizability of these results to large datasets. The growth of Health 2.0 technologies will further complicate this issue, with a data deluge of millions of messages transmitted over the web and mobile health media. Therefore, for large datasets, one needs to complement the qualitative method with an automated technique that can optimize resource utilization. Such an extension brings validity and generalizability to the manual coding in addition to facilitating the analysis of an entire dataset.

The precision of our methods of automated text analysis is of particular importance, as low precision may lead to formation of erroneous links in the network, such that the derived theme-based net-work is not truly representative of the messages exchanged. In case of the keyword-based modeling approach, the accuracy of the system may be further improved by more sophisticated choice of search terms, and the combination of the vector representations of these terms using vector space equivalents of negation and disjunction (Widdows & Peters, 2003). In addition, the evaluation of the keyword-based automated analysis employs messages exchanged by a single user in each theme. This limitation has been addressed in our subsequent work (experiments employing nearest neighbors algorithm) by developing an evaluation framework that builds on a large sample of messages drawn from multiple users. A limitation of the current system is the use of a relatively simple machine-learning algorithm (the k-nearest neighbors approach). This may have limited the system's accuracy and reliability. Future work will evaluate whether the performance of the system can be further improved by employing more sophisticated machine learning classifiers such as Support Vector Machines (S. Dumais, Platt, Heckerman, & Sahami, 1998; Hearst, Dumais, Osman, Platt, & Scholkopf, 1998).

The formal social network analysis employing one-mode network models conducted in this study was limited to two metrics. Future work will include network analysis metrics that take into account the interplay between individual-level and network-level measures (e.g. centrality and centralization). The two mode network analysis conducted in this research has a few limitations-

a) Although empirically-derived from QuitNet dataset, the themes might not have captured all the communication themes on account of the limitation with the qualitative analysis described above. Network influence through content-based affiliation may not be limited to these themes,
b) Because of the cross-sectional nature of the analysis this study might have limited the understanding of causality of content affiliation to abstinence behavior as well as potentially dynamic patterns of content affiliation in a network.

c) Our analysis considers only the affiliations through content based on communication occurring in virtual QuitNet platform. Some of these QuitNet members might have direct face-to-face communication (Leander & McKim, 2003; Wellman, Boase, & Chen, 2002) and my analysis does not take into account communication occurring in person. Future research should consider the cross-sectional nature of communication occurring among a group of individuals occurring both in person and online.

8.3. Conclusions

This work provides a new method that combines qualitative coding, automated text analysis, and network analysis to provide further insight into the mechanisms underlying behavior change in social networks. Results obtained using this method can inform the design of personalized and targeted interventions that persuade people to initiate or adhere to a positive behavior change, thus setting the stage for a new generation of translational interventions in public health and behavioral science.

Firstly, qualitative analysis of communication between members of an online social network can provide valuable insights into the mechanisms underlying human behavior change. With the onset of mobile smart phones and ubiquitous internet connectivity, online social network data reflects the intricacies of human health behavior as experienced by real people in real time. Therefore, analysis of these data can also provide us with the much needed theoretical and empirical foundations for design of high efficacy intervention strategies. This study offers insights into the various kinds of behavioral constructs prevalent in the messages exchanged among QuitNet users. In addition, it underlines the need for the use of inductive approaches for the analysis of online social network data to capture community-specific culture. In the case of QuitNet, activities such as pledges, bonfires emerged from within the community and each of those events mark a specific aspect of the smoking cessation process. With the evolution of the communication channels from being traditional face-to-face conversations to communication within virtual social networks powered by web-based mobile health systems, the validity of existing behavior change theories in the digital era has been questioned (Riley et al., 2011). This qualitative analysis supports the validity of behavior change theories in the context of 21st century technologies.

Secondly, the automated analyses evaluate the performance of methods from distributional semantics for the analysis of QuitNet content. Given the voluminous nature of online social network data, it is important to develop automated methods that can address the large quantities of free text data that are available online. While the performance of the automated method is reasonable, the accuracy of the system may be further increased by adopting more sophisticated machine learning algorithms than k-nearest neigbhor. Importantly, the automated method provides a feasible mechanism that facilitates code assignment to all the messages in the dataset, thus enabling resource-friendly qualitative analysis of large datasets. The reflective approach adopted in the automated classification system scales in linear proportion to the size of the dataset. Therefore the automated system can be applied to analyze millions of messages available on an online social network such as QuitNet. Large-scale qualitative analysis of the entire QuitNet dataset serves as proof-of-concept for the application of the automated method to large datasets. One implication of large-scale qualitative analysis of online social networks is that it facilitates the inclusion of semantic content into network models of online communities.

Thirdly, the one-mode and two-mode network models of QuitNet allow us to take an empirical approach to understanding communication-driven patterns of behavior change. The one-mode sub-networks within QuitNet derived from the communication content exchanged between pairs of members reveal differences across content types with respect to network clustering and opinion leaders. Such insights into the distribution of key nodes will help us design targeted social interventions related to the dissemination of specific-theme related information to relevant opinion leaders. In addition, large-scale personalized interventions can be delivered to a given module of members in accordance with this community's interest. In addition, the use of two-mode methods and affiliation exposure computations lets us examine the structural features of content-based QuitNet members' affiliation with different kinds of communication content, to investigate content-based network influence. Findings from this study suggest new directions in developing network interventions for public health by focusing on content-based targeted behavior change strategies.

In summary, online social networks have been gaining in popularity and present health researchers with a unique opportunity to understand human behavior change and deliver scalable and sustainable interventions. However, as demonstrated by QuitNet, these venues can also provide a forum for a community of dedicated users to assist one another in the pursuit of better health, an activity that ultimately has societal benefit beyond the users of QuitNet itself. The development of better tools to analyze social network content of this nature allows us a greater understanding of the ways in which such social networks mediate behavior change, thereby providing us with the opportunity for empirically-grounded interventions to further assist these communities with the attainment of their laudable goals. Content-based network analysis of QuitNet, made feasible by large-scale qualitative analysis using automated methods, has been shown to yield content-specific tailoring strategies that can be used for health promotion and behavior change.

Chapter 9: Innovation and Contributions

9.1. Innovation

The studies documented in this dissertation incorporate communication content into the network analysis of online social networks. To my knowledge, there has been no previously published research on social network analysis of health-related networks merging content-based approaches with network analysis methods to understand social influence. The use of automated methods to merge the qualitative and quantitative threads of network science forms an important and innovative aspect of this work. Overall, this work has several innovative aspects, as outlined below:

- The development of a “structure plus content” network analysis method to investigate inter- and intra-individual behavioral intricacies.
- The merging of qualitative, automated, and quantitative methods to model health-related online social media data.
- The extension of the scalability of qualitative methods to large-scale datasets, thereby enabling high throughput qualitative analysis.
- A demonstration of the design of empirically-grounded personalized and targeted solutions that trigger or improve adherence to a behavior change by harnessing the power of social relationships.

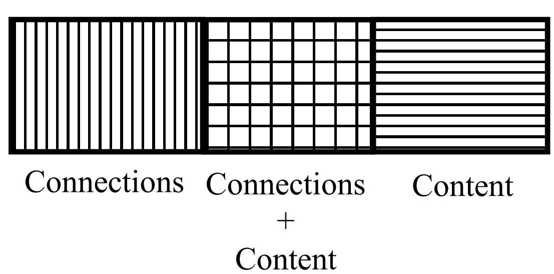


Figure 25. Summarizing innovative aspects of research

9.2. Contributions

9.2.1. Theoretical Contributions

With the evolution of the communication channels from being traditional face-face conversations to virtual social networks, the validity of the existing behavior change theories in the digital era has been questioned. This study establishes the validity of behavior change theories in the context of 21st century technologies. In addition, the inductive analysis of social network content has advanced our understanding of end user needs that fall outside the scope of existing behavior change theories, as well as that of previous network analyses that have occurred in the context of online communication channels. In this work, I also characterize the culture of QuitNet, identifying unique traditions that have implications for behavior change.

9.2.2. Informatics Contributions

This dissertation introduces a new method to analyze the communication content of online social networks. The method takes a bottom-up approach facilitating the development of a content-inclusive empirical approach for the network science and informatics domain. A novel distributional approach incorporating distributional information from an outside corpus is developed, and evaluated for the accuracy of its semi-automated annotation of messages in the corpus. The use of automated techniques facilitates scaling up the applicability of qualitative analysis to large-scale social media datasets. The use of automated methods from distributional semantics to analyze the messages exchanged in an online social network mitigates the bottlenecks posed by qualitative coding.

9.2.3. Public Health Contributions

This study offers new tools and techniques for the analysis of social network data in the digital age. These new approaches enable public health professionals to understand the social factors underlying behavior change. These insights suggest new tailoring strategies for individual- and network-based socio-behavioral interventions, laying the foundation for translational engineering of consumer support systems that promote and sustain healthy behavior change. The studies documented in this dissertation can inform- a) the public health community of ways in which the general public might be exposed to health promotion messages in digital media, b) interventionists' and health researchers' understanding of the social mechanisms underlying the communication patterns of individuals attempting to make a behavior change, and c) technology developers' understanding of the content-specific attributes that can potentially enhance patient engagement and empowerment.

Health-related behaviors such as smoking contribute to majority of deaths around the world. Development of scalable interventions to help people engage in healthy lifestyles has been a high-priority task for health researchers. In addition to use of automated methods, use of qualitative methods to understand behavior in digital media allows us to innovate evidence-based, empirically-grounded technological solutions in the context of virtual communication platforms. These solutions may hold the key to a healthier future.

References

Abraham, C., & Michie, S. (2008). A taxonomy of behavior change techniques used in interventions. *Health psychology, 27*(3), 379.

Adams, J, & White, M. (2005). Why don't stage-based activity promotion interventions work? *Health education research, 20*(2), 237-243.

Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008). *Finding high-quality content in social media.* Paper presented at the Proceedings of the international conference on Web search and web data mining.

Ajzen, Icek. (1991). The theory of planned behavior. *Organizational behavior and human decision processes, 50*(2), 179-211.

Albarracin, Dolores, Johnson, Blair T, Fishbein, Martin, & Muellerleile, Paige A. (2001). Theories of reasoned action and planned behavior as models of condom use: a meta-analysis. *Psychological bulletin, 127*(1), 142.

Alexander, C., Piazza, M., Mekos, D., & Valente, T. (2001). Peers, schools, and adolescent cigarette smoking. *Journal of adolescent health, 29*(1), 22-30.

Alwan, Ala, MacLean, David R, Riley, Leanne M, d'Espaignet, Edouard Tursan, Mathers, Colin Douglas, Stevens, Gretchen Anna, & Bettcher, Douglas. (2010). Monitoring and surveillance of chronic non-communicable diseases: progress and capacity in high-burden countries. *The Lancet, 376*(9755), 1861-1868.

Aral, Sinan, Muchnik, Lev, & Sundararajan, Arun. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences, 106*(51), 21544-21549.

Balch, George I, Tworek, Cindy, Barker, Dianne C, Sasso, Barbara, Mermelstein, Robin J, & Giovino, Gary A. (2004). Opportunities for youth smoking cessation: findings from a national focus group study. *Nicotine & Tobacco Research, 6*(1), 9-17.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ, US: Prentice-Hall, Inc.

Bandura, A. (2000). Social cognitive theory. *Encyclopedia of psychology, 7*, 329-332.

Bandura, Albert. (2002). Social cognitive theory of mass communication. *Media effects: Advances in theory and research, 2*, 121-153.

Baranowski, Tom, Buday, Richard, Thompson, Debbe I, & Baranowski, Janice. (2008). Playing for real: video games and stories for health-related behavior change. *American journal of preventive medicine, 34*(1), 74-82. e10.

Bastian, M, Heymann, S, & Jacomy, M (2009). *Gephi: an open source software for exploring and manipulating networks.* Paper presented at the International AAAI Conference on Weblogs and Social Media, San Jose, California.

Bauman, AE, Fardy, HJ, & Harris, PG. (2003). Getting it right: why bother with patient-centred care? *The Medical Journal of Australia, 179*(5), 253-256.

Berkman, L.F, & Glass, T. (2000). Social Integration, Social Networks, Social Support, and Health. In L. F. Berkman & I. Kawachi (Eds.), *Social Epidemiology*. New York: Oxford University Press.

Blanchard, Christina G., Albrecht, Terrance L., Ruckdeschel, John C., Grant, Charles H., & Hemmick, Rebecca Malcolm. (1995). The Role of Social Support in Adaptation to Cancer and to Survival. *Journal of Psychosocial Oncology, 13*(1-2), 75-95. doi: 10.1300/J077V13N01_05

Blei, David M, Ng, Andrew Y, & Jordan, Michael I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research, 3*, 993-1022.

Borgatti, Stephen P. (2002). NetDraw: Graph visualization software. *Harvard: Analytic Technologies*.

Borgatti, Stephen P, & Everett, Martin G. (1997). Network analysis of 2-mode data. *Social networks, 19*(3), 243-269.

Borgatti, Stephen P, Everett, Martin G, & Freeman, Linton C. (2002). Ucinet for Windows: Software for social network analysis.

Borgatti, Stephen P, Mehra, Ajay, Brass, Daniel J, & Labianca, Giuseppe. (2009). Network analysis in the social sciences. *science, 323*(5916), 892-895.

Bunton, Robin, Baldwin, Steve, Flynn, Darren, & Whitelaw, Sandy. (2000). The 'stages of change'model in health promotion: science and ideology. *Critical Public Health, 10*(1), 55-70.

Burg, Matthew M., & Seeman, Teresa E. (1994). Families and health: The negative side of social ties. *Annals of Behavioral Medicine, 16*(2), 109-115.

Burt, Ronald S. (1987). Social contagion and innovation: Cohesion versus structural equivalence. *American journal of Sociology*, 1287-1335.

Butts, Carter T. (2008). Social network analysis with sna. *J. Stat. Softw, 24*(6), 1-51.

Butts, Carter T. (2010). sna: Tools for social network analysis. *R package version, 2*.

Cacioppo, John T, Fowler, James H, & Christakis, Nicholas A. (2009). Alone in the crowd: the structure and spread of loneliness in a large social network. *Journal of personality and social psychology, 97*(6), 977.

Campbell, Harry, Hotchkiss, Rona, Bradshaw, Nicola, & Porteous, Mary. (1998). Integrated care pathways. *British Medical Journal, 316*, 133-137.

Carleton, Richard A, Lasater, Thomas M, Assaf, Annlouise R, Feldman, Henry A, & McKinlay, S. (1995). The Pawtucket Heart Health Program: community changes in cardiovascular risk factors and projected disease risk. *American Journal of Public Health, 85*(6), 777-785.

CDC, Center for Disease Control and Prevention. (2013). Smoking cessation- fact sheet. Retrieved October 17, 2013, from http://www.cdc.gov/tobacco/data_statistics/fact_sheets/cessation/quitting/

CDC, Centers for Disease Control and Prevention. (2011a). Quitting Smoking Among Adults—United States, 2001-2010. *Morbidity and Mortality Weekly Report.* Retrieved October 17, 2013

CDC, Centers for Disease Control and Prevention. (2011b). Vital Signs: Current Cigarette Smoking Among Adults Aged ≥ 18 Years—United States, 2005–2010. *Morbidity and Mortality Weekly Report.* Retrieved October 17, 2013

Centola, D. (2013). Social Media and the Science of Health Behavior. *Circulation, 127*(21), 2135-2144.

Centola, D. (2010). The spread of behavior in an online social network experiment. *science, 329*(5996), 1194-1197.
Champion, V, Foster, JL, & Menon, U. (1997). Tailoring interventions for health behavior change in breast cancer screening. *Cancer practice, 5*(5), 283.
Chen, G., Warren, J., & Evans, J. (2008). *Automatically generated consumer health metadata using semantic spaces.* Paper presented at the Proceedings of the second Australasian workshop on Health data and knowledge management-Volume 80.
Chen, G., Warren, J., & Evans, J. (2009). 'Qualities' not 'Quality'–Text Analysis Methods to Classify Consumer Health Websites. *electronic Journal of Health Informatics, 4*(1), e5.
Chen, G., Warren, J., & Riddle, P. (2010). Semantic Space models for classification of consumer webpages on metadata attributes. *Journal of Biomedical Informatics, 43*(5), 725-735.
Chew, Cynthia, & Eysenbach, Gunther. (2010). Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PloS one, 5*(11), e14118.
Christakis, N.A., & Fowler, J.H. (2008). The collective dynamics of smoking in a large social network. *New England journal of medicine, 358*(21), 2249-2258.
Christakis, Nicholas A, & Fowler, James H. (2009). *Connected: The surprising power of our social networks and how they shape our lives*: Hachette Digital, Inc.
Christley, RM, Pinchbeck, GL, Bowers, RG, Clancy, D, French, NP, Bennett, R, & Turner, J. (2005). Infection in social networks: using network analysis to identify high-risk individuals. *American journal of epidemiology, 162*(10), 1024-1031.
Chuang, K., & Yang, C. (2012). A study of informational support exchanges in medhelp alcoholism community. *Social Computing, Behavioral-Cultural Modeling and Prediction*, 9-17.
Chuang, K., & Yang, C.C. (2010). Social support in online healthcare social networking.
Cobb, N. K., Graham, A. L., & Abrams, D. B. (2010). Social network structure of a large online community for smoking cessation. *Am J Public Health, 100*(7), 1282-1289. doi: 10.2105/ajph.2009.165449
Cobb, N. K., Graham, A. L., Bock, B. C., Papandonatos, G., & Abrams, D. B. (2005). Initial evaluation of a real-world Internet smoking cessation system. *Nicotine Tob Res, 7*(2), 207-216. doi: 10.1080/14622200500055319
Cobb, N. K., Graham, A. L., Byron, M. J., Niaura, R. S., & Abrams, D. B. (2011). Online social networks and smoking cessation: a scientific research agenda. *J Med Internet Res, 13*(4), e119. doi: 10.2196/jmir.1911
Cohen, T. (2008). Exploring MEDLINE space with random indexing and pathfinder networks. *AMIA Annu Symp Proc*, 126-130.
Cohen, T., Schvaneveldt, R., & Widdows, D. (2010). Reflective Random Indexing and indirect inference: a scalable method for discovery of implicit connections. *J Biomed Inform, 43*(2), 240-256. doi: 10.1016/j.jbi.2009.09.003
Cohen, T., & Widdows, D. (2009). Empirical distributional semantics: methods and biomedical applications. *J Biomed Inform, 42*(2), 390-405. doi: 10.1016/j.jbi.2009.02.002

Cohn, Amy M, Hunter-Reel, Dorian, Hagman, Brett T, & Mitchell, Jessica. (2011). Promoting behavior change from alcohol use through mobile technology: the future of ecological momentary assessment. *Alcoholism: Clinical and Experimental Research, 35*(12), 2209-2215.

Consolvo, S., McDonald, D.W., & Landay, J.A. (2009). *Theory-driven design strategies for technologies that support behavior change in everyday life.* Paper presented at the Proceedings of the 27th international conference on Human factors in computing systems.

Coulson, Neil S, Buchanan, Heather, & Aubeeluck, Aimee. (2007). Social support in cyberspace: a content analysis of communication within a Huntington's disease online support group. *Patient Education and Counseling, 68*(2), 173-178.

DiClemente, Carlo C., & Prochaska, James O. (1998). Toward a comprehensive, transtheoretical model of change: Stages of change and addictive behaviors. In W. R. M. N. Heather (Ed.), *Treating addictive behaviors (2nd ed.)* (pp. 3-24). New York, NY, US: Plenum Press.

Dijkstra, Arie, De Vries, Hein, & Bakker, Martijntje. (1996). Pros and cons of quitting, self-efficacy, and the stages of change in smoking cessation. *Journal of consulting and clinical psychology, 64*(4), 758.

Doing-Harris, Kristina M, & Zeng-Treitler, Qing. (2011). Computer-assisted update of a consumer health vocabulary through mining of social network data. *Journal of medical Internet research, 13*(2).

Doreian, Patrick. (1980a). Linear Models with Spatially Distributed Data Spatial Disturbances or Spatial Effects? *Sociological Methods & Research, 9*(1), 29-60.

Doreian, Patrick. (1980b). On the evolution of group and network structure. *Social Networks*, *2*(3), 235-252.

Doreian, Patrick. (1981). Estimating linear models with spatially distributed data. *Sociological methodology, 12*, 359-388.

Dow, Malcolm M. (2007). Galton's Problem as Multiple Network Autocorrelation Effects Cultural Trait Transmission and Ecological Constraint. *Cross-Cultural Research, 41*(4), 336-363.

Dow, Malcolm M. (2008). Network Autocorrelation Regression With Binary and Ordinal Dependent Variables Galton's Problem. *Cross-Cultural Research, 42*(4), 394-419.

Dow, Malcolm M, Burton, Michael L, White, Douglas R, & Reitz, Karl P. (1984). Galton's problem as network autocorrelation. *American Ethnologist, 11*(4), 754-770.

Dumais, Susan, Platt, John, Heckerman, David, & Sahami, Mehran. (1998). *Inductive learning algorithms and representations for text categorization.* Paper presented at the Proceedings of the seventh international conference on Information and knowledge management.

Dumais, Susan T. (2004). Latent semantic analysis. *Annual review of information science and technology, 38*(1), 188-230.

Eysenbach, Gunther, & Till, James E. (2001). Ethical issues in qualitative research on internet communities. *Bmj, 323*(7321), 1103-1105.

Ezzati, Majid, Lopez, Alan D, Rodgers, Anthony, & Murray, Christopher JL. (2004). Comparative quantification of health risks. *Global and regional burden of disease attributable to selected major risk factors. Geneva: World Health Organization*.

Facebook. (2013). Facts. Retrieved 2013, October 17, from www.newsroom.fb.com/Key-Facts

Fernandes, Ralston, Hatfield, Julie, & Soames Job, RF. (2010). A systematic investigation of the differential predictors for speeding, drink-driving, driving while fatigued, and not wearing a seat belt, among young drivers. *Transportation research part F: traffic psychology and behaviour, 13*(3), 179-196.

Fernandez-Luque, Luis, Karlsen, Randi, & Bonander, Jason. (2011). Review of extracting information from the Social Web for health personalization. *Journal of medical Internet research, 13*(1).

Fiore, M.C. (2000). Treating tobacco use and dependence: an introduction to the US Public Health Service Clinical Practice Guideline. *Respiratory care, 45*(10), 1196-1199.

Fiore, Michael. (2008). *Treating tobacco use and dependence: 2008 update: Clinical practice guideline*: DIANE Publishing.

Fishbein, Martin. (1979). A theory of reasoned action: some applications and implications.

Fisher, William A, Fisher, Jeffrey D, & Rye, Barbara J. (1995). Understanding and promoting AIDS-preventive behavior: Insights from the theory of reasoned action. *Health Psychology, 14*(3), 255.

Fogg, Brian J. (2002). Persuasive technology: using computers to change what we think and do. *Ubiquity, 2002*(December), 5.

Free, Caroline, Whittaker, R, Knight, R, Abramsky, T, Rodgers, A, & Roberts, IG. (2009). Txt2stop: a pilot randomised controlled trial of mobile phone-based smoking cessation support. *Tobacco control, 18*(2), 88-91.

Freedman, Michelle J, Lester, Kristin M, McNamara, Cecelia, Milby, Jesse B, & Schumacher, Joseph E. (2006). Cell phones for ecological momentary assessment with cocaine-addicted homeless patients in treatment. *Journal of Substance Abuse Treatment, 30*(2), 105-111.

Freeman, L.C. (1979). Centrality in social networks conceptual clarification. *Social networks, 1*(3), 215-239.

Fu, W.T., Kannampallil, T.G., & Kang, R. (2009). *A semantic imitation model of social tag choices.* Paper presented at the Computational Science and Engineering, 2009. CSE'09. International Conference on.

Fujimoto, Unger, & Valente. (2012). A network method of measuring affiliation-based peer influence: assessing the influences of teammates' smoking on adolescent smoking. *Child Dev, 83*(2), 442-451. doi: 10.1111/j.1467-8624.2011.01729.x

Fujimoto, & Valente. (2012). Alcohol Peer Influence of Participating in Organized School Activities: A Network Approach. *Health Psychol*. doi: 10.1037/a0029466

Fujimoto, Williams, & Ross. (2013). Venue-based affiliation networks and HIV risk-taking behavior among male sex workers. *Sexually transmitted diseases, 40*(6), 453-458.

Fujimoto, K. (2012). Using Mixed-Mode Networks to Disentangle Multiple Sources of Social Influence. *Social Computing, Behavioral-Cultural Modeling and Prediction*, 214-221.

Fujimoto, K., Chou, C. P., & Valente, T. W. (2011). The Network Autocorrelation Model using Two-mode Data: Affiliation Exposure and Potential Bias in the Autocorrelation Parameter. *Soc Networks, 33*(3), 231-243. doi: 10.1016/j.socnet.2011.06.001

Gibson, D.R. (2005). Taking Turns and Talking Ties: Networks and Conversational Interaction1. *American Journal of Sociology, 110*(6), 1561-1597.

Giles, J.T., Wo, L., & Berry, M.W. (2003). GTP (General Text Parser) software for text mining. *Statistical data mining and knowledge discovery*, 455-471.

Giovino, Gary A., Mirza, Sara A., Samet, Jonathan M., Gupta, Prakash C., Jarvis, Martin J., Bhala, Neeraj, . . . Asma, Samira. (2012). Tobacco use in 3 billion individuals from 16 countries: an analysis of nationally representative cross-sectional household surveys. *The Lancet, 380*(9842), 668-679.

Glanz, K., Rimer, B.K., & Viswanath, K. (2008). *Health behavior and health education: theory, research, and practice*: Jossey-Bass.

Glaser, B.G., & Strauss, A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research*: Aldine de Gruyter.

Graaf, C de, Gaag, M Van der, Kafatos, A, Lennernas, M, & Kearney, JM. (1997). Stages of dietary change among nationally-representative samples of adults in the European Union. *European journal of clinical nutrition, 51*(2), S47.

Graham, Amanda L, Cobb, Nathan K, Raymond, Linda, Sill, Stewart, & Young, Joyce. (2007). Effectiveness of an internet-based worksite smoking cessation intervention at 12 months. *Journal of Occupational and Environmental Medicine, 49*(8), 821-828.

Gruzd, Anatoliy, Wellman, Barry, & Takhteyev, Yuri. (2011). Imagining Twitter as an imagined community. *American Behavioral Scientist, 55*(10), 1294-1318.

Hagan, Joe. (2011). Tweet science. *New York Magazine*(October 2).

Harper, F Maxwell, Moy, Daniel, & Konstan, Joseph A. (2009). *Facts or friends?: distinguishing informational and conversational questions in social Q&A sites.* Paper presented at the Proceedings of the 27th international conference on Human factors in computing systems.

Health, United States. Department of, & Services, Human. (2004). *The health consequences of smoking: A Report of the Surgeon General*: US Department of Health and Human Services.

Heaney, C.A., & Israel, B.A. (2002). Social networks and social support. *Health behavior and health education: Theory, research, and practice, 3*, 185-209.

Hearst, Marti A., Dumais, ST, Osman, E, Platt, John, & Scholkopf, Bernhard. (1998). Support vector machines. *Intelligent Systems and their Applications, IEEE, 13*(4), 18-28.

Heron, K.E., & Smyth, J.M. (2010). Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *British journal of health psychology, 15*(1), 1-39.

Hochbaum, Godfrey, Rosenstock, Irwin, & Kegels, Stephen. (1952). Health belief model. *United States Public Health Service*.
Hofmann, Thomas. (1999). *Probabilistic latent semantic analysis.* Paper presented at the Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence.
House, J.S. (1981). *Work stress and social support*: Addison-Wesley Publishing Company Reading, MA.
Hwang, Kevin O, Ottenbacher, Allison J, Green, Angela P, Cannon-Diehl, M Roseann, Richardson, Oneka, Bernstam, Elmer V, & Thomas, Eric J. (2010). Social support in an Internet weight loss community. *International Journal of Medical Informatics, 79*(1), 5-13.
Jafarinaimi, Nassim, Forlizzi, Jodi, Hurst, Amy, & Zimmerman, John. (2005). *Breakaway: an ambient display designed to change human behavior.* Paper presented at the CHI'05 extended abstracts on Human factors in computing systems.
Janz, Nancy K, & Becker, Marshall H. (1984). The health belief model: A decade later. *Health Education & Behavior, 11*(1), 1-47.
Kale, A., Kolari, P., Java, A., Finin, T., & Joshi, A. (2007). On Modeling Trust in Social Media using Link Polarity1.
Kiekel, Preston A, Cooke, Nancy J, Foltz, Peter W, & Shope, Steven M. (2001). Automating measurement of team cognition through analysis of communication data. *Usability evaluation and interface design*, 1382-1386.
Killen, Joel D, Robinson, Thomas N, Telch, Michael J, Saylor, Keith E, Maron, David J, Rich, Tia, & Bryson, Susan. (1989). The Stanford Adolescent Heart Health Program. *Health Education Quarterly*.
Kreuter, Matthew. (2000). *Tailoring health messages: Customizing communication with computer technology*: Routledge.
Lancaster, T., & Stead, LF. (2005). Individual behavioural counselling for smoking cessation. *Cochrane Database Syst Rev, 2*.
Lancaster, Tim, Stead, Lindsay, Silagy, Chris, & Sowden, Amanda. (2000). Regular review: Effectiveness of interventions to help people stop smoking: findings from the Cochrane Library. *BMJ: British Medical Journal, 321*(7257), 355.
Landauer, T.K. (2002). *Applications of latent semantic analysis.* Paper presented at the Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society, George Mason University, Fairfax, Virginia.
Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review; Psychological Review, 104*(2), 211.
Landauer, T.K., Foltz, P.W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes, 25*(2-3), 259-284.
Landauer, T.K., Laham, D., Rehder, B., & Schreiner, M.E. (1997). *How well can passage meaning be derived without using word order? A comparison of Latent Semantic Analysis and humans.* Paper presented at the Proceedings of the 19th annual meeting of the Cognitive Science Society.

Laumann, E.O., Marsden, P.V., & Prensky, D. (1989). The boundary specification problem in network analysis. *Research methods in social network analysis, 61*, 87.

Lawson, Peter J, & Flocke, Susan A. (2009). Teachable moments for health behavior change: a concept analysis. *Patient education and counseling, 76*(1), 25-30.

Leander, Kevin M, & McKim, Kelly K. (2003). Tracing the everyday'sitings' of adolescents on the internet: A strategic adaptation of ethnography across online and offline spaces. *Education, Communication & Information, 3*(2), 211-240.

Leimeister, Jan Marco, Schweizer, Karin, Leimeister, Stefanie, & Krcmar, Helmut. (2008). Do virtual communities matter for the social support of patients?: Antecedents and effects of virtual relationships in online communities. *Information Technology & People, 21*(4), 350-374.

Lin, James J, Mamykina, Lena, Lindtner, Silvia, Delajoux, Gregory, & Strub, Henry B. (2006). Fish’n’Steps: Encouraging physical activity with an interactive computer game *UbiComp 2006: Ubiquitous Computing* (pp. 261-278): Springer.

Looije, Rosemarijn, Neerincx, Mark A, & Cnossen, Fokie. (2010). Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human-Computer Studies, 68*(6), 386-397.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, 28*(2), 203-208.

Lustria, Mia Liza A, Cortese, Juliann, Noar, Seth M, & Glueckauf, Robert L. (2009). Computer-tailored health interventions delivered over the Web: review and analysis of key components. *Patient education and counseling, 74*(2), 156-173.

Lyons, Russell. (2011). The spread of evidence-poor medicine via flawed social-network analysis. *Statistics, Politics, and Policy, 2*(1).

Mariolis, P, Rock, VJ, Asman, K, Merritt, R, Malarcher, A, Husten, C, & Pechacek, T. (2006). Tobacco use among adults-United States, 2005. *Oncology Times, 28*(22), 42-47.

Marshall, Simon J, & Biddle, Stuart JH. (2001). The transtheoretical model of behavior change: a meta-analysis of applications to physical activity and exercise. *Annals of Behavioral Medicine, 23*(4), 229-246.

Matsuo, Yutaka, Mori, Junichiro, Hamasaki, Masahiro, Nishimura, Takuichi, Takeda, Hideaki, Hasida, Koiti, & Ishizuka, Mitsuru. (2007). POLYPHONET: an advanced social network extraction system from the web. *Web Semantics: Science, Services and Agents on the World Wide Web, 5*(4), 262-278.

McAuley, Edward, Lox, Curt, & Duncan, Terry E. (1993). Long-term maintenance of exercise, self-efficacy, and physiological change in older adults. *Journal of gerontology, 48*(4), P218-P224.

McPherson, M., Smith-Lovin, L., & Cook, J.M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 415-444.

Mizruchi, Mark S, & Neuman, Eric J. (2008). The effect of density on the level of bias in the network autocorrelation model. *Social networks, 30*(3), 190-200.

Mokdad, AH, Marks, JS, Stroup, DF, & Gerberding, JL. (2004). Actual causes of death in the United States, 2000. *Journal of American Medical Association, 291*, 1238-1245.

Morris, M.E. (2005). Social networks as health feedback displays. *Internet Computing, IEEE, 9*(5), 29-37.

Myneni, Sahiti, Cobb, Nathan K, & Cohen, Trevor. (2012). Finding Meaning in Social Media: Content-based Social Network Analysis of QuitNet to Identify New Opportunities for Health Promotion. *Studies in health technology and informatics, 192*, 807-811.

Myneni, Sahiti, Iyengar, Sriram, Cobb, Nathan K, & Cohen, Trevor. (2013). Identifying persuasive qualities of decentralized peer-to-peer online social networks in public health *Persuasive Technology* (pp. 155-160): Springer.

Newman, M.E.J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*(23), 8577-8582.

Nvivo. (2013). NVivo. Retrieved November 12, 2012, from http://www.qsrinternational.com/products_nvivo.aspx

Obregón, R., Chitnis, K., Morry, C., Feek, W., Bates, J., Galway, M., & Ogden, E. (2009). Achieving polio eradication: a review of health communication evidence and lessons learned in India and Pakistan. *Bulletin of the World Health Organization, 87*(8), 624-630.

Oinas-Kukkonen, Harri, & Harjumaa, Marja. (2009). Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems, 24*(1), 28.

Pallonen, Unto E, Velicer, Wayne F, Prochaska, James O, Rossi, Joseph S, Bellis, Jeffrey M, Tsoh, Janice Y, . . . Prokhorov, Alexander V. (1998). Computer-based smoking cessation interventions in adolescents: description, feasibility, and six-month follow-up findings. *Substance use & misuse, 33*(4), 935-965.

Paul, Michael J, & Dredze, Mark. (2011). *You Are What You Tweet: Analyzing Twitter for Public Health.* Paper presented at the ICWSM.

Perry, Cheryl L, Kelder, Steven H, Murray, David M, & Klepp, Knut-Inge. (1992). Communitywide smoking prevention: long-term outcomes of the Minnesota Heart Health Program and the Class of 1989 Study. *American Journal of Public Health, 82*(9), 1210-1216.

Phua, Joe. (2013). Participating in Health Issue-Specific Social Networking Sites to Quit Smoking: How Does Online Social Interconnectedness Influence Smoking Cessation Self-Efficacy? *Journal of Communication*.

Poirier, J., & Cobb, N. K. (2012). Social influence as a driver of engagement in a web-based health intervention. *J Med Internet Res, 14*(1), e36. doi: 10.2196/jmir.1957

Prochaska, Velicer, Wayne F, Redding, Colleen, Rossi, Joseph S, Goldstein, Michael, DePue, Judith, . . . Fava, Joseph L. (2005). Stage-based expert systems to guide a population of primary care patients to quit smoking, eat healthier, prevent skin cancer, and receive regular mammograms. *Preventive medicine, 41*(2), 406-416.

Prochaska, J.J., & Prochaska, J.O. (2011). A review of multiple health behavior change interventions for primary prevention. *American Journal of Lifestyle Medicine, 5*(3), 208-221.
Prochaska, James O., & Velicer, Wayne F. (1997). The Transtheoretical Model of Health Behavior Change. *American Journal of Health Promotion, 12*(1), 38-48. doi: 10.4278/0890-1171-12.1.38
Prokhorov, Alexander V, Kelder, Steven H, Shegog, Ross, Murray, Nancy, Peters, Ronald, Agurcia-Parker, Carolyn, . . . Hudmon, Karen Suchanek. (2008). Impact of A Smoking Prevention Interactive Experience (ASPIRE), an interactive, multimedia smoking prevention and cessation curriculum for culturally diverse high–school students. *Nicotine & Tobacco Research, 10*(9), 1477-1485.
Purpura, Stephen, Schwanda, Victoria, Williams, Kaiton, Stubler, William, & Sengers, Phoebe. (2011). *Fit4life: the design of a persuasive technology promoting healthy behavior and ideal weight.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
Puska, Pekka, Koskela, Kaj, McAlister, A, Mäyränen, H, Smolander, A, Moisio, S, . . . Rogers, EM. (1986). Use of lay opinion leaders to promote diffusion of health innovations in a community programme: lessons learned from the North Karelia project. *Bulletin of the World Health Organization, 64*(3), 437.
Revere, Debra, & Dunbar, Peter J. (2001). Review of Computer-generated Outpatient Health Behavior Interventions Clinical Encounters “in Absentia”. *Journal of the American Medical Informatics Association, 8*(1), 62-79.
Rhodes, Fen, Fishbein, Martin, & Reis, Janet. (1997). Using behavioral theory in computer-based health promotion and appraisal. *Health Education & Behavior, 24*(1), 20-34.
Riemsma, Robert Paul, Pattenden, Jill, Bridle, Christopher, Sowden, Amanda J, Mather, Lisa, Watt, Ian S, & Walker, Anne. (2003). Systematic review of the effectiveness of stage based interventions to promote smoking cessation. *Bmj, 326*(7400), 1175-1177.
Riley, W.T., Rivera, D.E., Atienza, A.A., Nilsen, W., Allison, S.M., & Mermelstein, R. (2011). Health behavior models in the age of mobile interventions: are our theories up to the task? *Translational behavioral medicine, 1*(1), 53-71.
Rodgers, Anthony, Corbett, Tim, Bramley, Dale, Riddell, Tania, Wills, Mary, Lin, Ruey-Bin, & Jones, Mark. (2005). Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. *Tobacco control, 14*(4), 255-261.
Rogers, E.M. (1995). *Diffusion of innovations*: Simon and Schuster.
Rosenquist, J Niels, Murabito, Joanne, Fowler, James H, & Christakis, Nicholas A. (2010). The spread of alcohol consumption behavior in a large social network. *Annals of Internal Medicine, 152*(7), 426-433.
Royce, Jacqueline M, Hymowitz, Norman, Corbett, Kitty, Hartwell, Tyler D, & Orlandi, Mario A. (1993). Smoking cessation factors among African Americans and whites. COMMIT Research Group. *American Journal of Public Health, 83*(2), 220-226.

Schvaneveldt, R.W. (1990). *Pathfinder associative networks: Studies in knowledge organization*: Ablex Publishing.

Shalizi, Cosma Rohilla, & Thomas, Andrew C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research, 40*(2), 211-239.

Shiffman, S., Stone, A.A., & Hufford, M.R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol., 4*, 1-32.

Shiffman, Saul, Balabanis, Mark H, Gwaltney, Chad J, Paty, Jean A, Gnys, Maryann, Kassel, Jon D, . . . Paton, Stephanie M. (2007). Prediction of lapse from associations between smoking and situational antecedents assessed by ecological momentary assessment. *Drug and alcohol dependence, 91*(2), 159-168.

Siegel, Rebecca, Naishadham, Deepa, & Jemal, Ahmedin. (2012). Cancer statistics, 2012. *CA: a cancer journal for clinicians, 62*(1), 10-29.

Skinner, Celette Sugg, & Kreuter, Matthew W. (1997). Using theories in planning interactive computer programs. *Health promotion and interactive technology: Theoretical applications and future directions*, 39-66.

Smith, Catherine Arnott, & Wicks, Paul J. (2008). *PatientsLikeMe: Consumer health vocabulary as a folksonomy.* Paper presented at the AMIA annual symposium proceedings.

Smith, Tony. (2001). Aggregation bias in maximum likelihood estimation of spatial autoregressive processes. *North American Regional Science Association, Charleston*, 15-17.

Smyth, Joshua M, & Stone, Arthur A. (2003). Ecological momentary assessment research in behavioral medicine. *Journal of Happiness Studies, 4*(1), 35-52.

Stead, L.F., & Lancaster, T. (2005). Group behaviour therapy programmes for smoking cessation. *Cochrane Database Syst Rev, 2*.

Steyvers, Mark, & Griffiths, Tom. (2007). Probabilistic topic models. *Handbook of latent semantic analysis, 427*(7), 424-440.

Stone, Arthur A, Shiffman, Saul, Atienza, Audie A, & Nebeling, Linda. (2007). *The science of real-time data capture*: Oxford University Press New York:.

Strauss, A, & Corbin, J. (2008). *Basics of qualitative research: techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage.

Strecher, Victor J, DeVellis, Brenda McEvoy, Becker, Marshall H, & Rosenstock, Irwin M. (1986). The role of self-efficacy in achieving health behavior change. *Health Education & Behavior, 13*(1), 73-92.

Strecher, Victor J, Kreuter, Matthew, Den Boer, Dirk-Jan, Kobrin, Sarah, Hospers, Harm J, & Skinner, Celette S. (1994). The effects of computer-tailored smoking cessation messages in family practice settings. *Journal of Family Practice, 39*(3), 262-270.

Street Jr, Richard L, Makoul, Gregory, Arora, Neeraj K, & Epstein, Ronald M. (2009). How does communication heal? Pathways linking clinician–patient communication to health outcomes. *Patient education and counseling, 74*(3), 295-301.

Tang, X., & Yang, C.C. (2010). *Identifing influential users in an online healthcare social network.* Paper presented at the Intelligence and Security Informatics (ISI), 2010 IEEE International Conference on.

Thoits, Peggy A. (1995). Stress, coping, and social support processes: Where are we? What next? *Journal of health and social behavior*, 53-79.

Thoreau, Henry David. (2002). Applying the transtheoretical model: tailoring interventions to stages of change. *Promoting exercise and behavior change in older adults: Interventions with the transtheoretical model*, 181.

Toscos, Tammy, Faber, Anne, An, Shunying, & Gandhi, Mona Praful. (2006). *Chick clique: persuasive technology to motivate teenage girls to exercise.* Paper presented at the CHI'06 extended abstracts on Human factors in computing systems.

Trends In Telecommunication Reform. (2012). Geneva, Switzerland: International Telecommunication Union.

Umberson, D., & Montez, J.K. (2010). Social Relationships and Health A Flashpoint for Health Policy. *Journal of Health and Social Behavior, 51*(1 suppl), S54-S66.

Valente. (2010). *Social Networks and Health: Models, Methods, and Applications: Models, Methods, and Applications*: Oxford University Press, USA.

Valente, T. W. (2003). Diffusion of innovations. *Genet Med, 5*(2), 69. doi: 10.1097/01.gim.0000061743.67794.c4

Valente, T. W. (2006). Opinion leader interventions in social networks. *BMJ, 333*(7578), 1082-1083. doi: 10.1136/bmj.39042.435984.43

Valente, T. W., & Fujimoto, K. (2010). Bridging: Locating Critical Connectors in a Network. *Soc Networks, 32*(3), 212-220. doi: 10.1016/j.socnet.2010.03.003

Valente, T. W., & Pumpuang, P. (2007). Identifying opinion leaders to promote behavior change. *Health Educ Behav, 34*(6), 881-896. doi: 10.1177/1090198106297855

Valente, Thomas W. (1995). *Network Models of the Diffusion of Innovations (Quantitative Methods in Communication Series)*: Hampton Press (NJ)(January 10, 1995).

Valente, Thomas W. (2012). Network Interventions. *Science, 337*(6090), 49-53. doi: 10.1126/science.1217330

Van De Belt, T.H., Engelen, L.J., Berben, S.A.A., & Schoonhoven, L. (2010). Definition of Health 2.0 and Medicine 2.0: a systematic review. *Journal of medical Internet research, 12*(2).

Vasuki, V., & Cohen, T. (2010). Reflective random indexing for semi-automatic indexing of the biomedical literature. *J Biomed Inform, 43*(5), 694-700. doi: 10.1016/j.jbi.2010.04.001

Velardi, P., Navigli, R., Cucchiarelli, A., & D'Antonio, F. (2008). *A new content-based model for social network analysis.* Paper presented at the Semantic Computing, 2008 IEEE International Conference on.

Voorhees, Carolyn C, Stillman, Frances A, Swank, Robert T, Heagerty, Patrick J, Levine, David M, & Becker, Diane M. (1996). Heart, body, and soul: impact of church-based smoking cessation interventions on readiness to quit. *Preventive medicine, 25*(3), 277-285.

Vuckovic, Nancy, Polen, Michael R, & Hollis, Jack F. (2003). The problem is getting us to stop: what teens say about smoking cessation. *Preventive medicine, 37*(3), 209-218.
Walters, Scott T, Wright, Jo Anne, & Shegog, Ross. (2006). A review of computer and Internet-based interventions for smoking behavior. *Addictive behaviors, 31*(2), 264-277.
Wellman, Barry, Boase, Jeffrey, & Chen, Wenhong. (2002). The networked nature of community: Online and offline. *It & Society, 1*(1), 151-165.
Wenze, Susan J, & Miller, Ivan W. (2010). Use of ecological momentary assessment in mood disorders research. *Clinical psychology review, 30*(6), 794-804.
West, Robert. (2005). Time for a change: putting the Transtheoretical (Stages of Change) Model to rest. *Addiction, 100*(8), 1036-1039.
Whittaker, Robyn, Borland, Ron, Bullen, Chris, Lin, Ruey B, McRobbie, Hayden, & Rodgers, Anthony. (2009). Mobile phone-based interventions for smoking cessation. *Cochrane Database Syst Rev, 4*.
WHO, World Health Organization. (2009a). Global adult tobacco survey (GATS).
WHO, World Health Organization. (2009b). *Global health risks: mortality and burden of disease attributable to selected major risks*: World Health Organization.
WHO, World Health Organization. (2011a). *Global status report on noncommunicable diseases 2010*: World Health Organization.
WHO, World Health Organization. (2011b). WHO report on the global tobacco epidemic, 2011: Warning about the dangers of tobacco. *Geneva, Switzerland*.
Widdows, D., & Cohen, T. (2010). *The semantic vectors package: New algorithms and public tools for distributional semantics.* Paper presented at the Fourth IEEE international conference on semantic computing.
Widdows, D., & Ferraro, K. (2008). Semantic vectors: a scalable open source package and online technology management application. *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, 1183-1190.
Widdows, D., & Peters, S. (2003). Word vectors and quantum logic: Experiments with negation and disjunction. *Mathematics of language, 8*(141-154).
www.QuitNet.com.). Retrieved October 15, 2013, from www.quitnet.com
Yang, C.C., Tang, X., Huang, J., & Unger, J. (2011). *A Study of Social Interactions in Online Health Communities.* Paper presented at the Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom).
Zhao, Wayne Xin, Jiang, Jing, Weng, Jianshu, He, Jing, Lim, Ee-Peng, Yan, Hongfei, & Li, Xiaoming. (2011). Comparing twitter and traditional media using topic models *Advances in Information Retrieval* (pp. 338-349): Springer.

Appendix A

Glossary of Social Network Terms

Ego: The person whose behavior is being analyzed.

Alter: A person connected to the ego

Node: An object that may or may not be connected to other objects in a network.

Tie: A connection between two nodes that can be either one-way (directed) or two-way (bilateral).

Centrality metrics: Describes a node and its relative importance within a network. Applications include how influential a person is within a social network. There are four main measures of centrality: degree, betweenness, closeness, and eigenvector.

D*egree* (or *connectivity*): The degree of a node is defined as the number of edges incident with the node. If the graph is directed, the degree of the node has two components: the number of outgoing links (referred to as the out-degree of the node), and the number of ingoing links (referred to as the in-degree of the node)

Betweenness: This metric quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.

Closeness: This measure emphasizes the distance of a node to all others in the network by focusing on the distance from each respective node to all others.

Eigenvector: This metric represents the influence displayed by a particular node. If a node influences just one other node, who subsequently influences many other nodes (who themselves influence still more others), then the first node in that chain is highly influential. This concept is similar to Google's PageRank algorithm.

Bridge: An individual node whose weak ties fill a structural hole, providing the only link between two individuals or clusters.

Graph Cohesion: A family of concepts characterizing the extent of connectedness of a Graph, in contrast with the centrality measures which characterize a particular node in a network.

Density: The proportion of direct ties in a network relative to the total number possible

Path length: The minimum number of ties required to connect two particular actors, as popularized by Milgram's famous 'six degrees of separation' small-world experiment

Cluster: A group of nodes, each of which is connected to at least one other node in the group.

Modularity: It is defined as the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random

Appendix B

Examples of Stop words used in the study

a	ate	besides	co
about	are	best	cold
above	around	better	come
accordingly	asb	between	coming
across	back	beyond	consequently
after	be	both	contain
afterwards	became	brief	containing
again	because	but	contains
against	become	by	corresponding
all	becomes	c	could
allows	becoming	came	currently
almost	been	can	d
alone	before	cannot	date
along	beforehand	cant	day
already	behind	cause	days
also	being	causes	degrees
although	below	certain	described
always	beside	changes	did
different	every	further	hence
do	everybody	furthermore	her

does	everyone	fun	here
doing	everything	g	hereafter
done	everywhere	get	hereby
down	ex	getting	herein
downwards	example	gets	hereupon
during	except	given	hers
e	f	gives	i
each	far	go	ie
eg	few	going	if
eight	fifth	gone	ignored
either	first	good	im
else	five	got	immediate
elsewhere	followed	great	in
enjoy	following	h	inasmuch
enough	for	had	inc
et	former	hardly	indeed
etc	formerly	has	indicate
even	forth	have	indicated
evening	four	having	indicates
ever	from	he	inner
insofar	lest	more	nine
instead	let	moreover	no

into	life	morning	nobody
inward	like	most	none
is	little	mostly	no one
it	lol	mr	nope
its	long	much	nor
itself	look	must	normally
ive	looked	my	not
j	looks	myself	nothing
job	m	n	novel
just	made	name	now
k	make	namely	nowhere
keep	man	near	null
kept	many	necessary	o
know	may	neck	of
l	me	neither	off
last	meanwhile	never	often
latter	men	nevertheless	oh
latterly	might	new	old
least	mine	next	on
less	month	night	once
one	perhaps	second	snow
ones	placed	secondly	so

only	please	see	some
onto	plus	seem	somebody
or	possible	seemed	somehow
other	post	seeming	someone
others	probably	seems	something
otherwise	provides	self	sometime
ought	q	selves	sometimes
our	que	sensible	somewhat
ours	quit	sent	somewhere
ourselves	quite	serious	specified
out	r	seven	specify
outside	rain	several	specifying
over	rather	shall	state
overall	really	she	still
own	relatively	should	sub
p	respectively	since	such
particular	right	six	sun
particularly	s	smoke	sunny
people	said	smoked	sup
per	same	smoking	t
take	third	two	w
taken	this	u	was

tell	thorough	under	way
than	thoroughly	unless	we
that	those	until	well
the	though	unto	went
their	thought	up	were
theirs	three	upon	what
them	through	us	whatever
themselves	throughout	use	when
then	thru	used	whence
thence	thus	useful	whenever
there	time	uses	where
thereafter	to	using	whereafter
thereby	today	usually	whereas
therefore	together	v	whereby
therein	tomorrow	value	wherein
thereupon	tonight	various	whereupon
these	too	very	wherever
they	toward	via	whether
think	towards	viz	which
thinking	twice	vs	while
whither	yes	way	whoever
who	yesterday	we	whole

whom	x	ya	yourself
whose	y	year	yourselves
why	wont	years	youve
will	work	yet	z
with	world	you	zero
within	would	your	
without	wow	yours	