Enhancing Health Information-Gathering Experiences in Online Health Communities

Albert Park

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Wanda Pratt, Chair

Meliha Yetisgen

Andrea Hartzler

Program Authorized to Offer Degree:

Department of Biomedical Informatics & Medical Education

University of Washington

Abstract

Enhancing Health Information-Gathering Experiences in Online Health Communities

Albert Park

Chair of the Supervisory Committee:
Professor Wanda Pratt
Department of Biomedical Informatics & Medical Education
The Information School

Online health communities can offer a range of diverse personal health expertise and experiences, yet gathering relevant health information is a significant challenge for members and researchers as each party faces different obstacles. For instance, members face challenges that pertain to text based computer-mediated communication (CMC) and information availability as determined by the active participation of other members. Similarly, the challenge of making sense of vast amounts of text prevents researchers from reaping the full benefits of the collective knowledge.

In my dissertation, I examine the challenges of gathering health information from online health communities in two parts according to the respective stakeholders. I first address the challenge that patient members face during their time of interaction with the online community to gather information. Within the context of CMC in online health communities, I focus on issues associated with topic changes–topic drift—and sustainment of active participation–posting

messages to participate in the communities. Next, I address the challenge of processing and making sense of a large amount of collective knowledge shared in online health communities. Within the context of patient-generated text in online cancer communities, I focus on the challenges of automatically understanding patient-generated text using existing natural language processing (NLP) tools.

More specifically, my specific aims for this dissertation are:

1. Understand topic drift and its effect on gathering health information from online health communities
2. Understand the beneficial effects of vocabulary similarity–homophily of vocabulary usage—associated with active participation in online health communities
3. Understand the challenges of automatically processing online community text and automated methods to detect failures

I examine these specific aims using two distinct online health communities: cancerconnect.com (a small, private online health community for patients with a wide range of cancers) and webmd.com (a large, public online health community with many non-cancer, disease-specific sub-communities).

One of the most important findings I discovered is that many members of online communities are willing to go the *extra mile* to help others in similar situations. Yet, many challenges are hindering the experience of gathering health information from online health communities. Although these efforts leave a digital trace that is embedded with diverse personal health expertise and experiences, we still lack the capability to automatically utilize this invaluable information. I contribute to resolving the issues faced by members and researchers; thereby, maximizing the benefits of online health communities and improving the experience of

gathering health information from them. I extend the existing knowledge related to topic drift, sustaining active participation, and processing patient-generated text with respect to the experience of gathering health information from online health communities.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Research is a solitary endeavor. The pleasures of research tend to be most intense for the researcher who is actually doing the research: thinking and rethinking the problem and seeing and understanding gaps that have never been explored. However, the journey to become a researcher and getting a Ph.D. was a collaborative adventure. So with big gratitude and much respect, I would like to thank everyone who helped me to grow as a person, to learn as a researcher, and to enjoy the journey for the past 4 years. I am truly grateful to all.

Thank you Prof. Wanda Pratt, Ph.D., my adviser, who guided me to grow as a researcher and a person. My journey in the University of Washington, Seattle would have never happened, if it wasn't for her. Wanda was always encouraging and understanding and she taught me the necessary skills to do what I really love to do. I am grateful for her guidance and the opportunity to see the value of patient expertise and online health communities.

Thank you Andrea Hartzler, Ph.D., who has been so kind and thorough in my and peermentoring research. She led by example, and always impressed me with her work ethic. I am making a conscious effort to emulate her kindness and professionalism into my future research effort.

Thank you Prof. Jina Huh, Ph.D., who has been my personal inspiration as a fellow 1.5 gen Korean-American walking in the same path. She continues to amaze me with her success, which in turn allows me to "dream big."

Thank you Prof. David W. McDonald, Ph.D., who always lighten the mood of research meetings and has provided countless wisdom in the field social computing. I am grateful to learn the fundamental concepts and skills in social computing as I hope to continue my future research in this direction.

Thank you Prof. Meliha Yetisgen, Ph.D., and Prof. Gary Hsieh, Ph.D., who served in my dissertation committee and helped to shape this dissertation. I am grateful for their insight and to learn different perspectives.

I need to thank even more people who were a part of my personal life during the years of pre-doctoral training. Thank you Jen for being wonderful, Hyunboo for being amazing roommate, and thank you both for making five ER trips with me. I also like to thank BHI, my cohort+/-1 (Alan, Amanda, Hannah, HG, JJ, Leslie, Logan, Meher, Nick, Nikhil, Rebecca, Sho, Ted, Thai, and Wen Wai), Imed, peermentoring, OWRC (Abby, Emily, Will), and friends for the extraordinary and memorable moments as well as for making Seattle home for four years. You made my transition to Seattle and earning Ph.D. easy and enjoyable.

I thank my mother for her endless love, my father for his wisdom, my big sister for her constant support, uncle HC & aunt JM for their care, Ha family for treating me as their son, my cousins and other members of my family. You instilled the principles I live by, the foundation for the wonderful experiences in life, and the belief in myself. I will always act knowing a part of you is in me.

Lastly, I thank the Lord for all wonderfulness in life.

Chapter 1: INTRODUCTION: OVERVIEW OF CHAPTERS

What do you do when you have health-related questions? Some people ask their doctors; some people ask their elders; but more of us now use the Internet to look up health information (Fox and Duggan 2013). With the development of social web technologies, a new practice of health information- gathering has emerged. Social media, such as online health communities, have become popular resources to gather health information (Fox and Rainie 2002; Fox and Duggan 2013). For this research, I am motivated by a number of different challenges that exist in the experience of gathering health information from online health communities for members and researchers as each party faces different obstacles. For instance, members face challenges that pertain to text based computer-mediated communication (CMC) and information availability as determined by the active participation of other members. Similarly, the challenge of making sense of vast amounts of text prevents researchers from reaping the full benefits of utilizing collective knowledge on behalf of community members and similar others in need.

The two sets of people start from the same resource—online health communities and a similar purpose of acquiring health information, but face different challenges. My goal for this thesis is to enhance this experience of gathering health information from online health community, by identifying challenges and providing practical implications. More specifically, my specific aims for this dissertation are:

1. Understand topic drift and its effect on gathering health information from online health communities
2. Understand the beneficial effects of vocabulary similarity–homophily of vocabulary usage—associated with active participation in online health communities
3. Understand the challenges of automatically processing online community text and automated methods to detect failures

I examine these specific aims using two distinct online health communities: cancerconnect.com (a small, private online health community for patients with a wide range of cancers) and webmd.com (a large, public online health community with many non-cancer, disease-specific sub-communities). Details of datasets are described in Chapter 3.

1.1 STUDYING THE CHALLENGES OF GATHERING HEALTH INFORMATION FROM MEMBERS' PERSPECTIVES

Understand topic drift and its effect on gathering health information from online health communities: Despite its increasing popularity, one prominent argument against online health communities is the lack of communication channels in text-based CMC. Though the problem is not exclusive to health, internet-based communication is limited in terms of the lack of nonverbal communication cues, such as gesture, facial expression, and physiological indicators (e.g., blushing and voice tone) compared to face-to-face communication (Sproull and Kiesler 1986; Joseph B. Walther 1994; J. B. Walther 1992). These nonverbal communication cues are crucial in communication, because they stimulate physiological indicators, and the interpretation of the indicators is a part of communication (Frijda 1993). As one consequence, **topic drift** (Hobbs 1990), the change of topic with progression of discussion, occurs frequently in CMC. Topic drift is one frequently occurring phenomena in CMC and has been linked to incoherence (S. Herring 1999) and conflict (Lambiase 2010). In a previous study of social-oriented chat on the Internet, nearly half (47%) of the conversation was considered off-topic (S. C. Herring and Nix 1997) in social chat. Although a minor level of topic drift is common and even expected in any conversation (Dorval 1990), for sensitive topics, such as in health, topic drift could prevent members from getting proper health information with respect to treatment decisions, symptom management, and clinical management. The prevention of acquiring proper health information can pose life-altering implications, yet topic drift has not been studied in these contexts. My first specific aim for this dissertation is to demonstrate the challenges associated with topic drift and to provide automatic methods detecting topic drift (Chapter 4).

Understand the beneficial effects of homophily of vocabulary usage associated with active participation in online health communities: Next, I seek to understand factors correlated with encouraging active participation (i.e., posting messages or replying to messages) in their own threads and sustaining active participation with the community. Active participation plays a significance role in both information gathering and information dissemination. The majority of health information available in online health communities provided by peer members (Gray et al. 1997; Sarasohn-Kahn 2008), thus without active participation of those peer members, health

information could not be disseminated and the community would cease to exist. Because active participation plays a crucial role in disseminating health information in online health communities, my second specific aim for this research is to examine a number of factors, such as the similarity of vocabulary usage that correlates with active participation in online health communities (Chapter 5).

1.2 STUDYING THE CHALLENGES OF GATHERING HEALTH INFORMATION FROM RESEARCHERS' PERSPECTIVES

Understand the challenges of automatically processing online community text and automated methods to detect failures: Reusing the collective experience and knowledge available in online health communities could provide powerful insights and could be a basis for interventions for online health communities. Moreover, related research opportunities are increasing as the prevalence of this type of information is rapidly increasing with the development of social web technologies (Sarasohn-Kahn 2008). Despite the importance and availability of such information, technical and methodological challenges exist in understanding the large amount of patient-generated text that is available in online health communities. My third specific aim for this research is to understand challenges of extracting health information using existing biomedical NLP tools and to provide automated methods to assess the capability of NLP tools in extracting health information from online health communities (Chapter 6).

1.3 SUMMARY AND OVERVIEW OF CHAPTERS

I investigate the problem of gathering health information from online health communities from the perspectives of both members and researchers in this dissertation. I introduce the overarching theme of my dissertation and the three specific aims of this research in this chapter. In chapter 2, I present the prior research on online health communities, CMC, topic drift, online community participation, and various challenges of processing patient-generated text. I then describe the details of the datasets for the specific aims of my research in Chapter 3. In Chapter 4, *Topic drift and its effect on gathering health information in online health communities*, I present the analysis of the first specific aim and provide insight on topic drifts on online health discussion and automatic methods to detect topic drift. I begin this chapter with methodological approaches for

automatically detecting topic drift then move on to quantitative and qualitative analysis. In Chapter 5, *Homophily of vocabulary usage: Beneficial effects of vocabulary similarity in online health communities*, I present the analysis of the second specific aim, quantitative and qualitative analysis on factors correlated with participation in the online communities. I focus on the effects of homophily of vocabulary usage, although I examine other factors as well. I begin this chapter with methodological approaches for measuring homophily of vocabulary usage and then describe the quantitative and qualitative analysis of homophily of vocabulary usage effect. Next, I describe the analysis of the third specific aim, in Chapter 6, *Challenges of automatically processing online community text*. I focus on NLP failures that occur when patient-generated text is processed by biomedical NLP tools as well as automatic detection of those failures. I begin the chapter with methodological approaches for manually identifying failures followed by description of the failures. Then I present methodological approaches for automatically detecting those failures and evaluation of the automated methods. Finally, I summarize contributions from this research and suggest possible implications and future research in the context of health information in online health communities in Chapter 7.

The prevalence and significance of patient-generated information, such as the information that resides in online health communities, is growing (Fox and Duggan 2013). We can enhance the experience of gathering health information and reuse this invaluable knowledge by understanding the challenges that exist in the experience of gathering health information from online health communities. Although the context of this research is scoped to only health-oriented online communities, contributions from this work could hold promise beyond the context of online health communities, such as providing seamless transaction of CMC, maintaining appealing social media platforms for members to share information, and processing patient-generated text.

Chapter 2: MOTIVATION AND LITERATURE REVIEW

I was motivated by the invaluable health information available in online health communities as well as the challenges of gathering and reusing that health information. I start the Chapter 2 by describing the characteristics of health information that members share in online health communities.

2.1. INTRODUCTION TO HEALTH INFORMATION, SUPPORTS, AND BENEFITS FOUND IN ONLINE HEALTH COMMUNITIES

More than a decade ago, websites like WebMD and MedHelp hosted health information online and patients acquired health information in the convenience of their homes. Today, these websites offer much more than hosting health information. As a part of their offerings, these websites now facilitate online health communities that further endorse conversations in which medical experts and non-experts share health information.

Although some communities like WebMD and MedHelp have several medical expert members, non-medical expert members—peer members—provide the majority of health information in online health communities. These information provided by peer members are particularly valued by patients (Gray et al. 1997; Sarasohn-Kahn 2008). For instance, one in four Internet users living with a chronic condition sought information from a peer with a similar condition in 2011 (Fox 2011), and such health information generated by peer members was used in decision making process (Berry et al. 2003). Peer members acquire their knowledge from their experience of managing their conditions (B. L. Paterson, Thorne, and Dewis 1998; B. Paterson and Thorne 2000) and few are more knowledgeable than their doctors on topics like self-management (Petersen 2006).

Improving members' experience of gathering health information depends on solid understanding of the characteristic of health information shared in online health communities. Previous research suggest that instead of acting as "amateur doctors," peer-members provide health information that is unique from what medical experts typically provide (Hartzler and Pratt 2011). According to the study, peer members provided actionable advice to cope day-to-day health issues, such as managing responsibilities and social issues. On the other hand, medical

experts provided fact-oriented information ranging from the health care delivery system and biomedical research.

In the context of social support, such exchange of actionable advice or suggestion is known as *informational support* (Cutrona and Suhr 1994). Cutrona and Suhr (Cutrona and Suhr 1994) identified six different types of social support: (1) informational support (e.g., suggestion/advice), (2) emotional support (e.g., affection or sympathy), (3) esteem support (e.g., compliment), (4) tangible/instrumental aid (e.g., loan), (5) social network support (e.g., companions), and (6) negative support (e.g., disagreement/disapproval). Of the six types of social support, Braithwaite et al. found emotional and informational support are the most frequently offered support; whereas social network support and tangible/instrumental aid were least frequently offered in an online health community for disabled patients (Braithwaite, Waldron, and Finn 1999). Moreover, emotional and informational support are the two most studied forms of social support in online health communities (Y. Wang, Kraut, and Levine 2012). The emotional support in online health communities, although an in-depth discussion of the topic is out of scope for this thesis, is an important aspect to consider when discussing acquisition of information from online health communities. The communication in online health communities is much more than simply exchanging health information—informational support; the communication also consist of emotional support that is expressed through empathy (Jenny Preece 1999). Moreover, informational and emotional support complement each other. For instance, after diagnosis, a patient requires support in the form of emotional needs first, followed by informational support (Jacobson 1986). Similarly, emotional support provided by peer members are particularly valued by patients, much like information support by peers. For instance, a study examining the experience of a breast cancer support community (Gray et al. 1997) showed that the support of peer-patients could be more comforting than that of family and friends. Patients instantly took comfort and connected with other patients because they were experiencing the same struggles.

Health information in online health communities, along with emotional support from peer members provides many benefits to online health community members. Research on the benefits of online health communities highlighted psychosocial benefits, such as reduced depression (C F van Uden-Kraan et al. 2009; Cornelia F van Uden-Kraan et al. 2008; Setoyama, Yamazaki, and Namayama 2011; Bartlett and Coulson 2011; Griffiths, Calear, and Banfield 2009), anxiety (Griffiths, Calear, and Banfield 2009; Setoyama, Yamazaki, and Namayama 2011), stress

(Setoyama, Yamazaki, and Namayama 2011; Bartlett and Coulson 2011), and negative mood (B. Shaw, Hawkins, and McTavish 2006). Also, participation in online health communities is linked to adaptive coping (Mo and Coulson 2012) and patient empowerment (Cornelia F van Uden-Kraan et al. 2008; C F van Uden-Kraan et al. 2009; Bartlett and Coulson 2011; Mo and Coulson 2012).

Another benefit of online health communities is improved health outcomes. A web-based system called CHESS (the Comprehensive Health Enhancement Support System) is a good example. Chess facilitates interactions between experts and peer patients, as well as providing resources and tools like lifestyle assessment (Gustafson et al. 1994). According to Gustafson et al, internet-based interaction improved quality of life and cognitive functioning, while reducing health care costs and hospitalizations for patients facing HIV/AIDS (Gustafson et al. 1994). In another study, Farnham et al. showed that the use of HutchWorld, an online health community, along with access to the Internet, buffered the detrimental effects of stress in times of social isolation (Farnham et al. 2002). *Stress and coping theory* (Lazarus and Folkman 1984) explains the buffering effect and coping. The authors of the theory suggest that coping is an internal change of the perception of the distressing situation. This change of appraisal can emerge during the buffering period resulting in coping. The participants reported no changes in life satisfaction whereas the control group showed a dramatic drop in life satisfaction.

Online health communities provide opportunities for members to be supporters, which had shown to associate with psychosocial benefits and improve quality of life. This phenomenon is also known as *helper therapy principle* (Riessman 1965) and has been reported in a number of studies. A study by Schwartz and Sendor (C. E. Schwartz and Sendor 1999) illustrated that supporters improved their psychosocial role performance (e.g., social activity), adaptability (e.g., self-efficacy function), and well-being (e.g., depression) even more than the supported patients. Similarly, providing support to an informal social network has been linked with feelings of personal control and lower levels of depressive symptoms in a nationwide survey of elderly people (Krause, Herzog, and Baker 1992). Furthermore, a peer support study among abused women suggested that the ability to give support was associated with the supporters' own recovery (Henderson 1995).

Informational and emotional supports are both essential components in improving patient experience and obtaining benefits offered by online health communities. However, we focus on

gathering health information for the following reasons. First, obtaining proper health information can empower patients and have a profound impact on adjustment to new situations (Helgeson and Cohen 1996). For instance, in a study of breast cancer patients, patients who received informational support from peers gained more perceived information competence compared to patients who acquired health information through clinically oriented resources (B. R. Shaw et al. 2007). Second, health information in online health communities can be used in clinical applications, making clinical decisions consistent with patients' preferences for example. Patients heavily relied on the health information by peer members in decision making process (Berry et al. 2003) and planning care based on patient preference was recommended as an effective strategy for improving patient outcome (Ruland 1999). Third, many patients are joining online health communities to gather health information. Newcomers of online health communities typically ask questions in their first posts (Galegher, Sproull, and Kiesler 1998). Furthermore, an intervention supporting information gathering has been suggested to empower patients (Jenny Preece 1999). Fourth, sharing information has been suggested to be a key role of peers. According to Huh and Ackerman (J Huh and Ackerman 2012), members sharing their health experiences, such as illness trajectories, was crucial in helping peer-patients manage their illnesses. Last, emotional support may be tailored towards one individual, but the health information written in online health communities can be used by many in a similar situation.

Although the prior studies delineate the benefits and characteristics of health information shared in online health communities, the efforts and challenges of gathering information are relatively unknown. In my dissertation, I contribute to resolving the issues faced by members and researchers; thereby, maximizing the benefits of online health communities and improving the experience of gathering health information from them. In the following subsections (2.2 – 2.4), I describe the prior work on communication challenges that can hinder patient member's experience of health information gathering in online health communities.

2.2. INTRODUCTION TO CHALLENGES ASSOCIATED WITH CMC IN ONLINE HEALTH COMMUNITIES

The most prominent use of the Internet is textual communication (Fox and Rainie 2002). The number of Americans who use internet-based communication grew more than 4 times from 1995 to 2002 (Fox and Rainie 2002). Computer-mediated communication (CMC) ranges widely from

emailing with friends and family members to online chatting with someone new who are like-minded or in a similar circumstance, peers in online community, for example.

Online health communities allow patients to cope and manage their illnesses through communication while providing means to overcome barriers like geographical isolation, stigma of the disease, and physical challenges. Text-based CMC is typically the basic mechanism for communication in the online health communities. The outcome of health information gathering can be determined by the experience of text-based CMC. Text-based CMC can be sensitive, especially in online health communities where topics are often serious and grim. Moreover, text-based CMC lacks many of face-to-face feedback cues. (Jenny Preece 1999)

Though the problem is not exclusive to health communities, internet-based communication is limited in terms of the lack of nonverbal communication cues, such as gesture, facial expression, and physiological indicators (e.g., blushing and voice tone) compared to face-to-face communication (Sproull and Kiesler 1986; Joseph B. Walther 1994; J. B. Walther 1992; Jenny Preece 1999). These nonverbal communication cues are important, because they stimulate physiological indicators, and the interpretation of the indicators is a part of communication (Frijda 1993). To increase the channels of communication, a new type of nonverbal cues, such as emoticons (Brooks et al. 2013), ASCII graphics, punctuation, and use of space are now being used in Internet-based communication (J Preece and Ghozati 2001).

Despite the lack of traditional nonverbal communication cues, studies found that certain emotions (i.e., nice, mean, happy, or sad) on the internet-based communication can still be interpreted as accurately as in face-to-face communication (Joseph B. Walther, Loh, and Granka 2005; J. T. Hancock, Landrigan, and Silver 2007; Kramer, Guillory, and Hancock 2014). According to *Social Information Processing theory*, writers' attitudes and feelings can be expressed with their word choice, punctuation use, and timing in text-based communication (J. B. Walther 1992). Moreover, emotion, such as sadness and frustration, can spread through internet-based communication as is the case in face-to-face communication (J. Hancock et al. 2008; Kramer, Guillory, and Hancock 2014). The authors suspect that the word choices can influence the reader despite the lack of traditional nonverbal cues. As evidence suggests, internet-based communication in online health communities is quite comparable to traditional face-to-face support group communication while being more accessible (Weinberg and Schmale 1996).

Although text-based CMC has improved in respect to increase nonverbal cues (J Preece and Ghozati 2001; Brooks et al. 2013) which are shown to be as interpretable as in face-to-face communication (Joseph B. Walther, Loh, and Granka 2005; J. T. Hancock, Landrigan, and Silver 2007; Kramer, Guillory, and Hancock 2014), many unresolved challenges remain in CMC of online health communities. In the following subsection 2.3, I describe one particular challenge of CMC, topic drift.

2.3. TOPIC DRIFT: A CHALLENGE OF GATHERING INFORMATION THROUGH CMC

We all have experienced a similar problem. You start a conversation with a colleague about work, and the next thing you know you are talking about weekend plans. This phenomenon is called **topic drift** (Hobbs 1990), in which topics change as a discussion progresses. In a conversation, topics naturally and continuously change (Dorval 1990). However, the topic drift that occurs frequently in CMC can be a source of incoherence (S. Herring 1999) and conflict (Lambiase 2010). For example, in a previous study of social-oriented chat on the Internet, nearly half (47%) of the conversation was considered off-topic (S. C. Herring and Nix 1997). Additionally, keeping the conversation on topic has shown to be difficult even for a highly-focused discussion group. For instance, in less than a month only a third (33%) of the conversation topics in the discussion group dedicated for the Oklahoma City bombing were related to the instance of the bombing or its related subtopics (Lambiase 2010). Although a minor level of topic drift is common and even expected in any conversation (Dorval 1990), for sensitive topics, such as in health, topic drift could prevent members from getting proper health information with respect to treatment decisions, symptom management, and clinical management. The prevention of acquiring proper health information can pose life-altering implications, yet topic drift has not been studied in these contexts.

Maintaining the goal and topic of discussion groups has been identified as a crucial element in CMC (S. Herring 1999). Despite the importance of maintaining the topic, drifts still occur. According to Hobbs, the three conversational devices attributed to topic drift in dialogs are: semantic parallelism, chained explanation, and metatalk (Hobbs 1990). Semantic parallelism occurs when a small portion of a topic gradually changes to other topics with similar and relevant properties. Chained explanation occurs when an explanation seems more interesting than the

topic and becomes the new topic. Metatalk occurs when conversational participants evaluate the drifted topic and change it back to the original goal of the conversation. The first two devices are cases of gradual topic drift, whereas metatalk opposes the drift by explicitly encouraging a return to the main topics.

Different conversational devices of topic drift identified by Hobbs were used in manual analyses (S. C. Herring 2003; Lambiase 2010) to explain how topics were changed and discarded in both synchronous (e.g., chat) and asynchronous (e.g., e-mail, forums) CMC. These studies analyzed topic drift with respect to topic changes at global (i.e., community-wide) or local (i.e., conversation) levels. Most of the topic drift occurred through semantic parallelism (Lambiase 2010) although distinguishing the types and severities of topic drift is inherently subjective (S. C. Herring 2003). Hobbs' topic drift devices can also help us understand how conversational participants contribute to topic drift or gain control over topics. Topic drift and topic control are imperative to analyze the relationship among conversational participants (Fairclough 1992), thus topic drift needs to be considered when analyzing the experience of health information gathering from members' perspectives.

For example, previous research on topic drift focused on whom, more specifically which gender, contributed to topic drift. Herring found that a small cohort of male participants dominated discussion groups through sheer volume of messages and adversarial strategies (S. C. Herring 1993). Similarly, Henley and Kramarae (1994) suggested that male participants induced radical topic drift by ignoring conventional conversational rules.

However, other studies found that dominant participants of both genders controlled discussion topics. Selfe and Meyer found that participants who used powerful and persistent language controlled the topic of conversation while limiting the opinions of others (Selfe and Meyer 1991). Similarly, in an unmoderated electronic discussion group called OKLABOMB, dominant participants of both genders changed the topic of the current conversation to their topics of interest (Lambiase 2010). These few dominant participants controlled the topics of discussion with inflammatory and emotionally aggressive postings while leading the discussion topics away from the original purpose, and forcing other participants to unsubscribe or remain inactive. This destructive behavior can be detrimental to members in online health communities and their experience of gathering health information. One likely reason for such aggressive participants could be the lack of moderation.

Few online communities employ moderators to govern discussion and create an engaging and respectful community culture (Wenger, McDermott, and Snyder 2002). In a moderated community, it is reasonable to assume that moderators will provide structure to keep topics relevant to the goal of the thread and community. However, how moderators can provide structure to prevent topic drift is an unanswered question in the research of topic drift. Also previous studies on topic drift have focused on e-mail-based newsgroups (Lambiase 2010), online discussion communities (Barcellini et al. 2005), and chats on various topics ranging from classical music (Nash 2005) to class meetings on pharmacy (S. C. Herring 2003). Although for some topics, topic drift can be inconsequential or even a natural course of conversation, for other sensitive topics, such as health, topic drift can pose serious consequences. Online health communities provide psychosocial benefits (e.g., adaptive coping) (Mo and Coulson 2012) as well as useful health information (Galegher, Sproull, and Kiesler 1998; Hartzler and Pratt 2011). Although topic drift can hinder these benefits, topic drift in online health communities has not been studied.

Topic drift can have negative impacts in CMC, especially in circumstances where clear goals (e.g., gathering health information) exist in the conversation. Despite the importance of staying on topic, supporting conversational participants in returning back to the original goals and topic of the discussion has received limited attention. Moreover, *who* provides this effort in CMC is unknown. In addition the role of moderators with respect to topic drift is unclear. In particular, although many online health communities focus on well-defined topics that are sensitive in nature and could have life-altering implications, topic drift studies of these communities are limited. This question is important for researchers of CMC as well as the designers and managers of these online communities. Thus, in chapter 5, I present a study of the moderated online health communities to understand topic drift with the aim of bringing new insights to researchers, designers, and managers of online communities. In the following subsection 2.4, I describe another challenge of health information gathering from members' perspectives–active participation of members.

2.4. PARTICIPATION: A CHALLENGE OF SUSTAINING ONLINE HEALTH COMMUNITIES AND INFORMATION DISSEMINATION

Active participation plays a significant role in both information gathering and information dissemination. As previously mentioned in Chapter 2.1, the majority of health information available in online health communities generates from communication among peer members. Therefore without active participation (i.e., interaction with other members by CMC) of members, health information could not be disseminated and the community would cease to exist.

On the two opposite ends of the spectrum, members' two main methods of gathering information are lurking (i.e., reading other people's conversation without posting) (Nonnecke and Preece 2000) and actively participating (Setoyama, Yamazaki, and Namayama 2011) in online health communities. Although lurkers can gain health information by reading other members' conversations, the information is not tailored for them. Moreover, lurkers do nothing to help promote their community (e.g., help other members to gain health information), gain less psychosocial benefits than active participants (Setoyama, Yamazaki, and Namayama 2011), and lose the benefits associated with being a supporter (Henderson 1995; Krause, Herzog, and Baker 1992; Riessman 1965; C. E. Schwartz and Sendor 1999). Although active participation could provide more benefits to members, Nonnecke and Preece (Nonnecke and Preece 2000) noted that lurkers (i.e., members who participate without posting) make up over 46% of online health communities. Even if the members have posted once, the majority will never post again (Joyce and Kraut 2006). Sustaining active participation remains a prominent challenge for online communities in general(Millen and Patterson 2002; Joyce and Kraut 2006; Y. Wang, Kraut, and Levine 2012; Arguello et al. 2006; Nonnecke and Preece 2000) due to issues like lurking and dropouts.

A growing number of studies investigate different aspects of members' participation and sustained active participation in online communities has been shown to positively correlate with a number of different factors. For example, researchers found that getting a reply significantly impacts members' participation in online groups. Joyce and Kraut (Joyce and Kraut 2006) examined newcomers' first post and found that receiving a reply increased the likelihood to post again in the community by 12% from six online groups including one Mozilla User Interface Newsgroup and five Usenet support groups.

On the opposite end of the line, Arguello et al. (Arguello et al. 2006) analyzed messages of eight Usenet newsgroups including two health groups, to learn conversational factors that can generate responses and active participation. They analyzed different characteristics of postings that elicited responses and found various factors influenced response rate. These factors include context (e.g., group identity), prior participation level, content (e.g., topical coherence with respect to community), rhetorical strategy (e.g., request for an advice), and linguistic features (e.g., first-person pronouns). Similar to the result found in a study by Joyce and Kraut (Joyce and Kraut 2006), the new comers were more likely to receive a response if they asked a question in their initiating post.

Furthermore, different types of responses can influence member retention. Wang et al. (Y. Wang, Kraut, and Levine 2012) showed that exposure to emotional support was positively associated with prolonged participation, whereas exposure to informational support was negatively associated with prolonged participation. Similarly, Welbourne et al. (Welbourne, Blanchard, and Boughton 2009) suggested that exchanging emotional support was positively associated with the sense of virtual community (SOVC), a sense of identification and belonging to the group. Furthermore, SOVC was positively related to better health outcomes while serving as a buffer to the detrimental effects of stress. Informational support, however, was negatively associated with SOVC.

To summarize the previous studies, receiving a response to a newcomers' first post (Joyce and Kraut 2006), receiving emotional support (Y. Wang, Kraut, and Levine 2012), obtaining a sense of community (Roberts 1998), and having familiarity with online interactive services (e.g., chat) (Millen and Patterson 2002) have all shown to positively correlate with active participation or degree of effort and time spent with the community. Although these studies provide insight on how to sustain active participation in general online communities, only one study examined online health communities. In contrast to the typical online community, active participation in online health communities could have implications for quality of life due to the purpose of participation: exchanging health information and psychosocial support. In that study of an online health community, participation was measured by sign-ins, which includes passive behaviors like lurking (Nonnecke and Preece 2000).

Homophily, a tendency of individuals to be attracted to others with similar traits such as attitude and behavior mimicry (Granitz, Koernig, and Harich 2008), is an important yet

unexplored principle in studying online community participation. However, homophily is a well-established principle in the context of social network analysis (McPherson, Smith-Lovin, and Cook 2001), and it has been shown to positively correlate with credibility of authors in online health communities (Z. Wang et al. 2008). Moreover, homophily expressed through unconscious behavior-mimicry was correlated with likelihood of liking the respondents (Chartrand and Bargh 1999). Similarly in language, verbal mimicry of function words (i.e., content-free parts of speech) has been shown to positively correlate with liking the respondents (Ireland et al. 2011) and positively functioning social dynamics (Gonzales, Hancock, and Pennebaker 2009). In online health communities, both function words and content words can serve as important cues for measuring homophily expressed in vocabulary usage. The function words are related to unconscious mimicry whereas the content words are related to similarity in health traits. Although homophily measured using all types of vocabulary usage—vocabulary similarity—could have effects on members' active participation, vocabulary similarity has not been studied with respect to active participation in online health communities.

Based upon previous literature, we expect community members to appreciate responses that use a similar vocabulary. Thus, we hypothesized that individuals are likely to sustain active participation if they receive replies written with a similar vocabulary. In Chapter 6, I focus on active participation and replies written with a similar vocabulary to better reflect the benefits of online health communities to members and community sustainability. In the following subsection 2.6, I describe researchers' challenge of automatically gathering and processing a large amount of health information in online health communities.

2.5. UNDERSTANDING THE BENEFITS AND CHALLENGES OF AUTOMATICALLY PROCESSION TEXTUAL COMMUNICATION IN ONLINE HEALTH COMMUNITIES

Members in online health communities openly discuss, seek support, and exchange health information (Fox and Rainie 2002) through CMC. The CMC records contain valuable health information for both patients and researchers alike, such as peer members' personal health management experiences. The continuously growing number of CMC users, the body of peer member-generated text, and inquiries of peer generated health information (Sarasohn-Kahn 2008)

present more opportunities for understanding such resources. However, now we know that topics of conversation are constantly changing due to issues like topic drift (Chapter 2.3) and the communication in online health communities has elements of health information and empathy (Jenny Preece 1999). Reusing collective experience and knowledge available in the online health community can be a powerful tool (J Huh and Ackerman 2012), however, we lack the necessary capability and resource to make sense of the vast amount of data as well as extracting relevant health information.

One scalable approach to processing text-based peer member-generated data is natural language processing (NLP). An increasing number of researchers studying peer member-generated text, such as online health communities, have used statistical methods based on manually annotated datasets (Wen and Rose 2012; Y. Wang, Kraut, and Levine 2012; Chee, Berlin, and Schatz 2011; MacLean and Heer 2013; Maclean et al. 2015). Utilizing statistical methods, researchers extracted cancer patient trajectories from patients' posts (Wen and Rose 2012), estimated the level of social support in an online breast cancer community (Y. Wang, Kraut, and Levine 2012), predicted adverse drug reactions from health and wellness Yahoo! Groups (Chee, Berlin, and Schatz 2011), identified medically relevant terms (MacLean and Heer 2013), classified addiction phases (Maclean et al. 2015), predicted individual at risks for depression (De Choudhury et al. 2013), and discovered patient posts in need of expertise from moderators (Jina Huh, Yetisgen-Yildiz, and Pratt 2013). These methods can be highly effective in a given online community, but they either require tremendous upfront effort to manually annotate or do not provide semantic connections. Furthermore, maintenance and generalizability remain as major challenges for such statistical methods.

Existing biomedical NLP tools have the potential to be used immediately and promise to provide greater generalizability than statistical approaches while providing semantic connections. Researchers have developed various NLP techniques and applications in the biomedical domain. For example, the Clinical Text Analysis and Knowledge Extraction System (cTakes) (Savova et al. 2010) was developed to map concepts to medical ontology from clinical notes. cTakes is specifically trained for clinical domains and consists of NLP components that can be executed in sequence. Also, the National Center for Biomedical Ontology (NCBO) (Musen et al. 2012; Jonquet, Shah, and Musen 2009) annotator identifies a term and maps it to ontological concepts from multiple knowledge resources to allow the utilization of integrated knowledge. Other

applications have been developed primarily for specific uses, such as Medical Language Extraction and Encoding System (MedLEE) (C. Friedman et al. 1995), whose goal pertains to identifying specified conditions in radiology reports. However, MedLEE was later adapted as a decision support system for Columbia-Presbyterian Medical Center (CPMC) (Carol Friedman 1997), and as a phenotypic information extractor (BioMedLEE) (L. Chen and Friedman 2004) from biomedical literature.

One of the most widely regarded NLP applications in biomedicine is MetaMap (Aronson and Lang 2010). The National Library of Medicine (NLM) developed MetaMap which employs various computational linguistic and NLP techniques designed to process a relatively wide range of biomedical and health research literature. MetaMap is a standalone application that is highly configurable but cannot be trained. MetaMap processes text and links text fragments to specific concepts from the Unified Medical Language System (UMLS). The UMLS is a collaborative effort to enable semantic interoperability between systems by providing mapping structures of biomedical vocabularies (Humphreys, Lindberg, and Schoolman 1998). The UMLS consists of more than 1.3 million concepts from more than 100 families of biomedical vocabularies (Y. Chen et al. 2007). Three knowledge sources enable applications to utilize the UMLS: (a) Metathesaurus, (b) Semantic Types and Semantic Network, and (c) SPECIALIST Lexicon. Each concept in the UMLS is classified into one or more semantic types.

However, traditional biomedical NLP tools target biomedical literature and clinical notes in electronic medical records rather than peer member-generated text in online communities. One of the biggest challenges in applying biomedical NLP tools to a different type of text is the difference in vocabulary. For example, Zeng et al. recognize differences in the vocabulary used by patients and clinicians (Q. Zeng et al. 2001), and initiated a movement to capture the consumer health vocabulary used by patients (Q. T. Zeng and Tse 2006). Smith and Wicks manually evaluated patient-generated text from PatientsLikeMe.com and found that over 50% of patient-submitted symptoms did not map to the UMLS due to issues like misspellings and slang (C. A. Smith and Wicks 2008). Although Keselman et al. (Keselman et al. 2008) reported fewer cases of unmapped terms from patient-generated online community posts than Smith and Wicks (C. A. Smith and Wicks 2008), the researchers recognize this remaining challenge as a significant problem.

Recognizing the differences in vocabulary, a number of efforts involving expansion of UMLS to include peer member generated text have been reported (Q. T. Zeng et al. 2006; Keselman et al. 2008; Brennan and Aronson 2003; C. A. Smith, Stavri, and Chapman 2002; Q. T. Zeng and Tse 2006; Q. Zeng et al. 2001). One of the biggest efforts is the open-access, Collaborative Consumer Health Vocabulary Initiative (CHV) (Q. T. Zeng and Tse 2006; Q. Zeng et al. 2001; Keselman et al. 2008). CHV is a collaborative effort to address differences in terminology used by peer members and medical experts. Compared to text generated by medical experts, peer member-generated text tends to include layman-friendly terminology that is familiar to patients (Q. Zeng et al. 2001). CHV seeks to address this issue by expanding patient-oriented vocabularies in the UMLS. Although the terminology difference could theoretically be addressed by expanded vocabularies, it is questionable whether CHV can fully address other issues of patient-generated text, such as misspellings, community nomenclature, and Internet-oriented writing styles. To address this issue, Elhadad et al. applied an unsupervised, semantics based methods to detect community nomenclature including typical misspellings (Elhadad et al. 2014). Although the method is domain-independent, it only accounts for three semantic types.

The effort to process peer member-generated text, such as e-mail (Brennan and Aronson 2003; C. A. Smith, Stavri, and Chapman 2002) and search queries (Q. T. Zeng et al. 2006; Keselman et al. 2008) using biomedical NLP tools also have been reported. For example, Brennan and Aronson processed patient-authored e-mails using MetaMap and showed the potential of processing this patient-generated, informal text to identify UMLS concepts (Brennan and Aronson 2003). However, Brennan and Aronson identified three types of errors: overly granular parsing of phrases into separate terms (e.g., splitting of the phrase *"feeling nauseous"*), inappropriate mappings that are simply nonsensical or incorrect for the context (e.g., a verb *"back"* being mapped to *"body location or region back"*), and mismatches resulting from terms and semantic types having more than one meaning (e.g., confusion between *"spatial concept right"* and *"qualitative concept right"*) (Brennan and Aronson 2003). Zeng et al. have mapped the UMLS concepts to patients' Internet search queries (Q. T. Zeng et al. 2006), and the study highlighted the difference between terminology structures in UMLS and mental model of patients.

Patient-generated text in online health communities has also been studied. Researchers have applied NLP techniques, such as linguistic features (Alpers et al. 2005) and machine learning

classifiers (Wen and Rose 2012; Y. Wang, Kraut, and Levine 2012; Chee, Berlin, and Schatz 2011), to these patient-generated texts in online community posts. In these studies, researchers investigated extracting cancer patient trajectories from patients' posts (Wen and Rose 2012), estimating the level of social support in a breast cancer community (Y. Wang, Kraut, and Levine 2012), and predicting adverse drug reactions from health and wellness Yahoo! Groups (Chee, Berlin, and Schatz 2011). The machine learning techniques presented from these studies have been successful in their respective purposes and communities. However, as previously mentioned these techniques require the tremendous upfront effort of manual coding and building statistical models. Also, generalizability is questionable because statistical models are tailored to specific data. Recognizing limitations of current techniques for processing patient-generated text, MacLean and Heer combined statistical models and crowdsourcing to identify medically relevant terms in patient-generated text and tested in two different online support communities (MacLean and Heer 2013). This combination of statistical models and crowdsourcing methods outperformed two existing NLP applications. Although their method can better identify medically relevant terms, it lacks concept mappings that other NLP applications can offer.

The prior studies show the continuous effort to improve biomedical NLP tools to process peer member generated text. As NLP technologies and source vocabularies continue to evolve, we need easy, low-cost methods to systematically assess the performance of those tools. Traditionally in NLP, evaluations involve a great deal of manual effort. Moreover, a new evaluation for different types of text requires additional annotated datasets; thus, maintenance can often be difficult. Recognizing the potential benefits of performing a low-cost assessment of NLP tools, we explore the feasibility of automated methods to detect failures without producing annotated datasets. Given MetaMap's long history of use in biomedical contexts, its semantic connection, its configurability, and its scalability, I applied our failure detection system to MetaMap in processing peer member-generated text from an online cancer community to demonstrate the feasibility of automatically detecting occurrence of failures in Chapter 7.

2.6. CONCLUSION AND IMPLICATIONS

The use of the Internet is rapidly increasing and has made an impact on everyday life including communication, gathering information, and health care (Fox and Rainie 2014). Although the future for Internet-based schemes in a health care system seems to be optimistic, knowledge gaps

exist in our understanding of health information gathering from perspectives of both patients and researchers. In my dissertation, I examine the challenges of gathering information from online health communities in two parts. The first part addresses the challenge that patients face during their time of interacting with the online community to gather information. Within the context of text based CMC in online health communities, I focus on issues generated from topic changes and community participation. The second part addresses the challenge of gathering collective knowledge shared in online health communities. Within the context of online health community text, I focus on the challenges of automatically understanding vast amounts of patient-generated text using existing NLP.

We focus on online health community and its text, but the insights can impact wider types of online community and consumer-generated text. For instance, social media platforms, such as Facebook, can share similar conversational interactions with online health communities. While at the same time, question-and-answer sites share similar informal asking and answering interactions, as is the style of online health communities. Moreover, these online social communities could experience the same issues with respect to topic drift and participation. Across this range of settings for which this research has implications, it is crucial to fill the knowledge and technical gaps in our understanding.

Another example of an impact of this dissertation is enhancing the performance of systems that use patient-generated text. Google Flu Trends (Cook et al. 2011) gained significant attention as a promising approach for tracking flu epidemics. Yet evaluations demonstrate performance problems when compared to traditional flu surveillance data sources like CDC (Cook et al. 2011; Wagner 2013). Similarly, tweets were used to improve natural disaster and emergency response situations (Starbird and Palen 2011), and an effectiveness of a drug was evaluated using *in the wild* patient-generated data on PatientslikeMe.com (Wicks et al. 2011). However in these studies, the utilization of patient-generated text required manual structuring of patient data. Enhanced NLP tools that are more closely tailored to the unique characteristics of patient-generated text could help to improve/expedite the performance/process-time of such tools. In the next chapter, Chapter 3, I describe two online health communities that I used for this dissertation.

Chapter 3: DATASETS AND STUDIED ONLINE COMMUNITIES

In this dissertation, I examine two distinct online health communities, WebMD.com and cancerconnect.com. Webmd.com is a larger public online health community with many disease specific sub-communities, whereas Cancerconnect.com is a small private online health community for patients with a wide range of cancers. In section 3.1 and 3.2, I describe the details of each of the online communities.

3.1 DATASET 1. WEDMD.COM

More than 50 disease specific sub-communities[3.1] were available on WebMD.com in Aug of 2014. I selected seven communities that vary with respect to disease and illness characteristics to cover wider aspects of health (i.e., biological, psychological, and sociological). The seven communities are (1) attention deficit hyperactivity disorder (ADHD), (2) breast cancer, (3) diabetes, (4) heart disease, (5) multiple sclerosis (MS), (6) pain management, and (7) sexual health. These communities are highly active communities that ranked within the top 20 WebMD forums in total number of threads to eliminate communities that could have member dropouts due to the low-activity of the community. My research group downloaded all publicly available posts from these seven communities to examine the topic drift question of online health communities (Chapter 4).

To investigate the relationship between active participation and homophily expressed in vocabulary usage, vocabulary similarity, I restricted my analysis to five communities from WebMD.com to better understand the roles of the members and moderators (Chapter 5). These communities also employ moderators and medical doctors (MDs) whom have clearly defined roles in discussions than peer members. This allowed us to analyze the relationship among community members with different roles. To narrow down the communities with respect to this goal, first I selected communities with two or more moderators. I then selected communities with sufficient number of activities from members, at least 50 first replies from both members and moderators. After applying these exclusion criteria, five communities remained eligible for subsequent analysis: (1) ADHD, (2) Diabetes, (3) Heart Disease, (4) Pain Management, and (5) Sexual Health communities. All of the quotes in this dissertation have been de-identified. My

team sought review by University of Washington Institutional Review Board (IRB) and the data was exempt from review.

Table 1 shows overall characteristics of the seven WebMD communities that I studied in this dissertation. Each community had an average between 2.86 and 7.78 posts in a thread (depending on the community), and across all communities the average thread length was 6.06 posts.

Table 1. Characteristics of seven WebMD communities studied

	ADHD	Breast Cancer	Diabetes	Heart Disease	MS	Pain Management	Sexual Health
Dates that data was collected	7/ 2005 ~ 6/2012	8/2007 ~ 5/2012	6/2007 ~ 5/2012	3/2008 ~ 5/2012	11/2007 ~ 1/2013	9/2007 ~ 6/2012	9/2007 ~ 1/2013
Total Post	8,704	21,612	64,085	11,874	27,412	27,333	68,136
Total Thread	2,313	3,227	8,242	4,146	4,943	4,656	10,278
# MD	3	3	1	3	2	3	1
# Staff	10	10	15	7	9	9	10
# User	2,984	2,147	4,385	3,815	2,710	5,843	13,624
# Power-user	5	17	36	8	16	17	32
Mean thread length	3.76	6.70	7.78	2.86	5.55	5.87	6.63
Median thread length	3	5	5	2	4	4	4
Max thread length	85	88	97	71	68	97	99

[3.1] ADD/ADHD, Allergies, Alzheimer's, Anxiety & Panic, Asthma, Baby's First Year, Back Pain, Bipolar Disorder, Breast Cancer, Cancer, Colorectal Cancer, Cholesterol Management, Depression, Diabetes, Diet, Digestive Disorders, Ear, Nose & Throat, Epilepsy, Erectile Dysfunction, Eye Health, Fibromyalgia, Fitness & Exercise, Food & Cooking, Heart Disease, Hepatitis, HIV/AIDS, Hypertension, Knee & Hip Replacement, Lupus, Menopause, Men's Health, Migraines/Headaches, Multiple Sclerosis, Oral Health, Osteoarthritis, Osteoporosis, Parenting, Pregnancy, Prostate Cancer, Pain Management, Parkinson's Disease, Pet Health, Raising FIT Kids, Relationships & Coping, Rheumatoid Arthritis, Sexual Health, Sexually Transmitted Diseases, Skin Beauty, Skin Problems & Treatments, Sleep Disorders, Smoking Cessation, Sports Medicine, Stroke, Substance Abuse, Trying to Conceive, and Women's Health

3.2 DATASET 2. CANCERCONNECT.COM

The second dataset consisted of community posts from CancerConnect.com, an online cancer community for cancer patients, their families, friends, and caregivers to exchange support and advice. More than 60 interest-based sub-communities[3.2] were available cancerconnect.com in Jan of 2013, however not all sub-communities are cancer related. Although "Insurance Issues" and "New User Community" are non-cancer sub-communities, these communities could contain valuable health related information that can enhance patient experience. Thus, I included these non-cancer sub-communities in the analysis of processing peer-member text using biomedical NLP (Chapter 7). The dataset consisted of a total of 2,010 unique user-members who posted to the community and 9,657 posts from March of 2010 to Jan of 2013. CancerConnect.com also have 51 moderators, however, I excluded their posts to focus on texts generated by peer members. (Table 2)

Table 2. Characteristics of CancerConnect communities studied

	CancerConnect Communities
Dates that data was collected	3/ 2010 ~ 1/2013
Total Post	9,657
# Moderators	51
# User	2,010

[3.2] Advocacy, ALL, AML and MDS, Anal Cancer, Bereavement, Bladder Cancer, Bone Cancer, Brain Cancer, Breast Cancer, Carcinoid Tumors, Caregiver, Cervical Cancer, Chemotherapy, Chronic Myelogenous Leukemia, Clinical Trials, Colon Cancer, Complementary Medicine, Dana-Farber CancerConnect Community, Dana-Farber Volunteer Connect, Entertainment, Esophageal Cancer, Exercise , Gastric Cancer, Genetic Testing & Cancer Screening, GIST, Head and Neck Cancers, Healthy Living, Hodgkin's Disease, Insurance Issues, Kidney Cancer, Liver and Bile Cancers, Loyola Cancer Connect, Lung Cancer, Management, Maryland Oncology CancerConnect, Melanoma, Michael's Mission CancerConnect, Multiple Myeloma, New User Community, Non-Hodgkin's Lymphoma, Novel and Emerging Therapies, Nutrition, Ovarian Cancer, Pancreatic Cancer, Pediatric Cancer, Prescribed Reading, Prostate Cancer, Radiation Therapy, Rare Cancers, Rectal Cancer, Roswell Park, Roswell Park BMT Community, Sarcoma, Seattle Cancer Care Alliance, Skin cancer, Spirituality, Stem Cell Transplant, Survivorship, Technology, Testicular Cancer, The James Connect, Thyroid Cancer, Unknown Primary Cancer, Uterine Cancer, Vanderbilt-Ingram Cancer Center, and Young Adults with Cancer

3.3 SUMMARY

In Chapter 3, I described the details of two datasets to help readers contextualize the three specific aims of this dissertation. In the next three chapters (Chapter 4, Chapter 5, and Chapter 6), I describe the analytic methods, results, and findings of each aim in a single chapter. I elected to describe methods with results and findings because the few initial results informed and guided the selection of methods for the secondary research questions in Chapter 5 and Chapter 7. Moreover, a separate methods chapter could unnecessary complicate the readability because each of the three specific aims has multiple research questions that employ different methods. Thus, I describe the methods, results, and findings together.

Chapter 4: TOPIC DRIFT AND ITS EFFECT ON GATHERING HEALTH INFORMATION IN ONLINE HEALTH COMMUNITIES

Topic drift, the change of topic with progression of discussion, is a frequently occurring phenomenon that has been linked to incoherence (S. Herring 1999) and frustration in computer-mediated communication (CMC). This incoherence and frustration can prevent members from gathering health information in online health communities that typically employ text-based CMC as their main method of communication. However, topic drift has not been studied in the contexts of online health communities and members' CMC experience, thus we explore topic drift in online health communities in Chapter 4. In the following subsection 4.1, I describe five new research questions pertains to topic drift in online health communities as well as investigation methods for each of the research questions.

4.1. RESEARCH QUESTIONS AND METHODS

Our overarching research goal for this chapter is to understand gradual and severe topic drift that occur at global and local levels in online health communities. We characterized the severity of topic drift into **gradual** and **severe topic drift**, determined by topical relevance of the current topic to the previous topic. We also categorized the types of topic drifts into **global** and **local topic drift**, based on the level in which topic drift occurred. Local topic drift refers to posts unrelated to the respective thread (i.e., when someone brings up a new topic within a thread that doesn't relate to the original post), whereas global topic drift indicates discussion outside of the respective communities' goals (i.e., when someone starts a new thread that doesn't relate to the topic for that community). We examined topic drifts in online health communities with five questions: (1) What are the reactions and meta-discussions towards topic drift in explicitly identified topic drift threads? (2) How does local topic drift occur in threads? (3) Who brings the topic back to the original goal of threads? (4) Can local topic drift be detected automatically? (5) Can the effort to stay on topic be detected automatically?

RQ 1. What are the Reactions and Meta-discussions Towards Topic Drift in Explicitly Identified Topic Drift Threads?

To gain a deep understanding of members' reactions and meta-discussions towards topic drift, we began our analysis using self-identified topic drifted threads. The interpretation of topic drift is inherently subjective and consistently of interpretation has shown to be challenging (S. C. Herring 2003). Thus, we focused on the reactions and meta-discussions when topic drift is apparent to the members. We extracted the threads containing explicitly-identified topic drift with the key terms "*hijack"* or "*off topic."* Although few key terms contain modifiers to indicate lesser degree (e.g., *"may be slightly off topic"*), the topic drifts in these threads are typically considered severe. Within each thread, we manually referred back from those key terms to the preceding posts and reviewed the context of the conversation to ensure that the key terms were used for rhetorical strategies to change topic, gain control over the topic, or indicate off-topic content in the post. In other words, the key terms had to be used to indicate or related to either global or local topic drift. We then qualitatively analyzed these threads with respect to meta-discussion and members' emotional reaction towards topic drift.

RQ 2. How Does Local Topic Drift Occur in Threads?

To understand how gradual and severe topic drift occurs in typical threads (i.e., local level), we first qualitatively analyzed 50 randomly selected threads with at least six posts, which is the average number of posts in all seven communities. We needed enough posts per thread to perform an in-depth, manual analysis of topic drift. In contrast to RQ1, these threads do not explicitly indicate topic drift and may not show elements of topic drift. We identified a number of main topics in each of the posts and examined whether and how many of those main topics changed as threads evolved. Using this information, we categorized into gradual or severe topic drift. We also identified possible sources and the general trend of topic drift following an open coding process (Strauss and Corbin 1990).

RQ 3. Who Brings the Topic Back to the Original Goal of Threads?

To understand who brought the topic back to the original goals and topics of threads at local level, we first qualitatively analyzed threads. This concept of staying on topic is related to one of

Hobbs' conversational devices, metatalk (Hobbs 1990). Metatalk also can be about a discussion regarding their conversation, therefore, for clarity, we did not use the term metatalk. Instead, we defined this effort as the **counteraction** to topic drift.

In addition to the 50 randomly selected threads in RQ2, we purposely sampled additional 20 threads with at least six posts, in which members with defined roles (i.e., moderators and MDs) have participated. This is to understand how these members impact or counteract topic drift. Their participation was relatively limited and we were not able to sample enough threads with their participation using random selection. Then, we anonymized the community members' ID and manually examined who was making a counteraction to topic drift.

RQ 4. Can Local Topic Drift be Detected Automatically?

Many of the studies on topic drift manually analyzed interactions to understand topic changes and participants' roles (S. C. Herring 2003; Lambiase 2010; Nash 2005; Barcellini et al. 2005). The manual method is accurate, but it is labor intensive and limited to small datasets. However, in the field of information retrieval (IR), researchers have long used automated methods to detect and track topic changes. One of the more widely used methods is the relevance measurement of terms in a text segment using thresholds based on term frequency-inverse document frequency (tf-idf) of text segments (Kumaran and Allan 2004). Similarly, we applied cosine-similarity metric and vector space model (VSM) to assess relevance among posts under the same thread to detect both gradual and severe topic changes at a thread level. We created tf-idf at a community level to reflect important terms of the community.

We evaluated this automated topic drift detection technique with severe topic drift and "on-topic" posts. First, we used posts from RQ1 that contain key terms: *hijack* or *off topic* as ***positive cases*** of topic drift that our detection system should recognize as low in relevance measurement, given that members explicitly indicated the off topic nature of the post. To ensure the quality, we manually examined and removed posts from this analysis if (1) the keyword *hijack* literally meant illegally seize or steal (few posts were about the 9-11 tragedy), (2) the keywords were used to describe the definition of an acronym (e.g., *"OT means [...]"*) or community nomenclature (e.g., *"hijacking a thread means [...]"*), (3) the keywords had a modifier to indicate lesser degree (e.g., *"may be slightly off topic"*), (4) the keywords were used to have meta-discussion about off topic discussions, or (5) the keywords were used to start new off topic

threads (e.g., *"OFF TOPIC BUT […]"*). These are stricter criteria than RQ1, because this also removes global topic drifts along with the definition and lesser degreed local topic drifts.

To identify ***negative cases*** of topic drift, we used posts from RQ1 if (1) the posts has negated keyword (e.g., *"this is not off topic"*) or (2) the community members had shown intentions to bring topic back to the original goals (e.g., *"your question got hijacked, I'll try to get it back on track"*). Because such negative cases only amounted to few cases, we then added manually selected additional 70 "on-topic" posts (i.e., little or no topic drift) that the detection system should recognize as high in relevance from RQ3-qualitative analyses.

Using these positive and negative cases as a gold standard, we calculated the F1-score, recall, precision, and accuracy of the automated topic drift detection system by comparing to the average score of the posts in the same position of all threads. The position of the post is also important because we expect the scores to naturally decrease as conversation progresses. We made these selections and adjustments prior to the evaluation process without any information on their relevance scores.

RQ 5. Can the Effort to Stay On Topic be Detected Automatically?

According to Dorval, the topic of conversation is not a static but constantly changing feature (Dorval 1990). Assuming the same natural deviation happens in the online health discussions, we only focus on the irregular increase of relevance scores at a post level to detect counteractions to topic drift. The irregular increase of relevance score indicates that the current post contains more relevant topics compared to the previous post (i.e., threshold), a sign of making a counteraction to topic drift. Similar to RQ4, we applied the cosine-similarity metric and VSM with tf-idf to detect counteraction.

Then, we evaluated the automated counteraction detection technique using 50 new randomly selected posts: 25 posts with natural decrease in relevance score and 25 counteracting posts with increase in relevance score. We anonymized the 50 posts then manually categorized into natural topic drift or counteraction to topic drift, while referring back to initiating and other previous posts to understand the context. Using manual assessment of 50 posts as a gold standard, we then calculated the F1-score, recall, precision, and accuracy of the automated topic drift detection system.

Lastly, we automatically measured who (i.e., which type of member) made counteractions. The unit of analysis was a member and we counted the occurrences of counteractions. Because the most active members can have the highest chance of providing such effort, we normalized each member's counteraction occurrences by dividing it by their total number of replying posts, thus converting the occurrences into percentages. To understand how people in defined roles provide counteractions, we categorized the members into moderators/MDs and users according to their defined roles in the communities. We then further categorized users into original posters, power users, and regular users. The user types were mutually exclusively for individual threads. We categorized **original posters** as users who initiated a thread and came back to the thread for further conversation; **power-users** as users who posted more than the average number of posts by staff moderators and MDs; and the rest as **regular-users**. We then applied the Mann-Whitney U test to examine whether significant difference exists between user types in providing the counteraction to topic drift. Next, I report on the results of these research questions in section 4.2.

4.2. RESULTS

Results for RQ1. What are the Reactions and Meta-discussions Towards Topic Drift in Explicitly Identified Topic Drift Threads?

We found 185 posts: 53 posts in which community members used key terms "*hijack"* and 132 posts in which community members used "*off topic"*. These terms were used by both members enduring and causing topic drifts. After applying the criteria, only 118 posts were considered in this analysis due to terms' usage.

"*Hijack"* were typically associated with local topic drift whereas "*off topic*" were used to indicate both local and global topic drift. The types of topic drift were not mutually exclusive. (Table 3)

Table 3. Usages of the key terms

	Local Only	Global Only	Both
Hijack	34	0	7
Off Topic	33	23	19

Two major themes emerged in the coding, and are presented in their own sections below. First, we found posting culture and reaction towards severe topic drift (i.e., hijacking and off

topic discussions). Second, contrarily to previous research, we found support for having off topic discussions (i.e., global topic drift).

Posting culture with respect to severe topic drift

The following is a canonical example of how a community-member believes these conversations should start and unfold in WebMD community.

> *"It is usually best to start your own discussion if you have questions or are seeking support. Certainly, you can share your own experiences and that is encouraged here. […] Elaborating too much is sometimes considered "hijacking a thread" in internet message board lingo. Many times this happens in these discussions […] Regardless of how a discussion evolves, I always pray that we all can find the answers and relief we need." (Pain Management community)*

As shown in the example post, the member was clearly aware of topic drift and described one of the causes with the term, "*hijacking"*. Hijacking occurs when a member elaborates too much or in other words, dominated the thread with posts. Dominating the conversation was often associated with topic control and topic drift in previous studies (S. C. Herring 1993; Selfe and Meyer 1991; Lambiase 2010), because the dominant participant frequently changed the current topic to their own areas of interest. In the last sentence of the example post, the member indicates how topic drift can affect original posters in getting needed help. Furthermore, the member showed an intuitive understanding that the thread's main goal was to answer or give support to the original poster. Similarly, the original posters showed frustration when the topic drifted away as shown in the example below.

> *"Why do my post always get treated as if I am posting something none [no one] needs to know I do not think I will post here anymore, :angry: [name]" (Diabetes community)*

According to Lambiases (Lambiase 2010), off topic discussions were associated with discontinuation or inactivity by members of the community. Similarly, we observed frustration by original posters when topic drifted in the middle of threads as shown in the example above.

Furthermore, we found apologetic behavior shown by community members who caused the topic drift. The following example post is a response to the example post above. The member apologized to the original poster for changing to the topic after being confronted.

> *"I am sorry I hijacked your thread, [name]. That is a bad habit of mine. Your post IS valuable. […] Truly, [name], I didn't mean to hurt your feelings. I am sorry." (Diabetes community)*

Because WebMD members showed an intuitive understanding of the thread's main goals, members showed effort to stay on topic and even brought the conversation back to the original purpose of the thread as shown in the following example post.

> *"Since your question seems to have gotten hijacked by a debate about the economy and the merits of various forms of education, I'll try to get it back on track […]" (Sexual Health community)*

Moreover, experienced community members knew the sensitivity of certain topics, such as religion, that could easily become the main topics of the conversation through chained explanation.

> *"As for the Christian aspect, I hesitate to go there at all because in my observation of past threads, this tends to hijack the main topic completely […]" (Sexual Health community)*

Example posts showed how community members negatively reacted to local topic drift and topic control. These examples of topic drift often occurred in the middle of threads as the result of how conversations evolved. In contrast, members described starting off-topic discussions with regard to the specific community (i.e., global topic drift) positively.

Useful Purposes of Sharing Off-topic Discussions

Although members reacted negatively towards topic drift in the middle of the thread, they reacted positively to off topic stories shared separately in new threads. The following posts are examples regarding off topic threads written by a moderator and user respectively.

> *"To all re your comments about staying on point.... if this community were being taken over with off-topic and/or "fun" discussions, that would be one thing. But that's not the case in this community or even on this thread. Yes, on any board there are newcomers and lurkers. They get good information and support here. But, to me, a bit of fun can also add to creating a community where someone would like to stay for a while." (Diabetes community moderator)*

> *"Personally, if all that was discussed in any community on WebMD was the main topic, I would cease to be involved. I enjoy sharing with others and getting to know them by discussing what is happening in their lives other than the main health concern." (Diabetes community member)*

Members showed support for having off-topic discussion, because it could build rapport and bring members closer; however members also suggested ways to indicate that the topic of the thread was irrelevant to the specific community. For example adding either *OT* or *off topic* in the title was suggested or practiced in five communities (i.e., ADHD, Breast Cancer, Diabetes, MS, and Sexual Health) as shown in the quote below.

> *" "OT" means "off topic." It lets people know the subject won't be MS. Otherwise, someone will click on it expecting to find MS info, then they may get upset when they find that it's not what they wanted." (MS community)*

Our findings suggest that members' negative reaction to local topic drift and topic control similar to previous studies (S. C. Herring 1993; Selfe and Meyer 1991; Lambiase 2010). However, we extend the literature by identifying benefits of having off-topic discussions (i.e.,

global topic drifts) in a melancholy and serious topic focused community, such as an online health community.

Results for RQ2. How Does Local Topic Drift Occur in Threads?

Severities and sources of topic drift

Our qualitative analysis shows that in most threads, the topic changed gradually. The majority of the **gradual topic drift**—in which most of the topics remained in the discussion while few topics were newly introduced and neglected, fit into semantic parallelism. This level of slight changes occurred in almost every post of the threads; however, they generally stayed within the global frame of the topic for the specific community. In contrast, the following is an example thread from the Heart Disease community that shows a **severe topic drift**—in which the previous topic was replaced with a new topic—and topic control. The thread ended as Poster_C repeatly posted about their personal experience and changed the topic.

> *Poster_A: "I have heard that minutes makes a difference concerning a stroke, could seven hours make a difference with a blood clot beginining in the uppper leg traveling down?"*
> *Poster_B: "I don't know how long it takes for tissue to die, but I would not wait 7 hours. But more important the clot can breakup and go to the lungs."*
> *Poster_C: "My mother died waiting for 7 hours, before she was taken to hospital. She was refused transport by ambulance service, because of misdiagnosed by Paramedic."*
> *MD_Poster_D: "It could - the longer tissues are deprived of blood and oxygen, the greater the risk of having permanent damage. Always better to seek medical attention earlier when there are concerns of a stroke or of other similar types of issues."*
> *Poster_C: "Thanks Dr. [Name], I feel she could have been saved, if she had gotten treatment sooner. The doctors will not say one way or the other, they afraid of being ask to testify in court."*
> *MD_Poster_D: "I'm so sorry to hear about your loss – it's really helpful for other people in this forum to hear about your experiences - so thank you for sharing them with us."*

Poster_C: "Dr. [Name], Thanks for your welcome response. You seem like a caring and knowledgeable Doctor. I would like to talk to you futher about this situation, My email address is [e-mail address]"

We observed that sharing personal experience that pertained to the main topic of the thread was a commonly practiced communication method in WebMD. Although the personal narratives can provide powerful information (Hartzler and Pratt 2011), they can also prompt topic drift as shown in the example thread above.

Another source of severe topic drifts was the MD members of the community. Many community members asked personal questions to MD members in the middle of the threads similar to how Poster_C behaved in the example thread above. Other causes of severe topic drift included jokes or the inability of community members to use the online interface. For example, members started a new conversation or sent a personal message from within the thread, then excused themselves for changing the topic as shown below.

"Hi guys, it maybe kinda off topic. I actually don't know how to post my own topic (I'm new here, sorry.) […]" (Sexual Health Community)

Severe topic drift occurred from multiple sources, including the desire to joke, share personal stories or interact with MD members and inability of community members to use the communities' interface. Although complete prevention of severe topic drift may not be possible, some can be addressed with better design (See 4.3 Discussion).

Results for RQ3. Who Brings the Topic Back to the Original Goal of Threads?

Original posters make the biggest difference

Our qualitative analysis shows that in threads with highly active original posters tended to stay on topic better. Below is an example thread from the Heart Disease community in which the original poster provided counteraction to the topic drift.

Poster_A: "My roommate is not yet 40 and has had to have 3 stints in the last year. Now the Cardiaologists are saying that he needs a pacemaker and most likely was born with

Bradycardia. What exactly is Bradycardia and are we looking at a not so good prognosis for his future? Isn't he somewhat young to be needing a pacemaker and what if the pacemaker does not have the expected result? What is the next step?"

Poster_B: "Bradycardia just means a heart rate of less than 60. That in itself is not a problem. The problem is when it is not beat fast enough to keep up with demand. Here is some information on the causes and treatment. [URLs]"

MD_Poster_C: "Bradycardia means a low heart rate, usually less than 60 beats per minute. A pacemaker can be recommended when bradycardia is symptomatic, or if there is another underlying electrical problem with the heart that increases the risk of the heart slowing even more or even stopping. Pacemakers work very well […]"

Poster_D: "Dear Dr. [Name], My mother is 73 years old, and had a pacemaker placed 2 years ago at the, Mayo Clinic. She is doctoring in her home town now. They are having trouble contoling her comidon levels, it has been 2 weeks now, and still do not have the levels controled. Is this unusual to have it take so long to adjust her levels?"

Poster_A: "Thanks for your reply. One more question. How does all of this associate with the stints and I forgot to mention that my friend has had two heart attacks this past year. Can we possibly look forward to my friend having a long and somewhat healthy life if the pacemaker and his new medication, Coreg, do what they are supposed to do? I realize that I am asking you to look into a crystal ball, but surely you have an educated guess?"

Poster_A started the thread with multiple questions including (1) what *Bradycardia* is and (2) what possible outcomes and expectations are. Both Poster_B and MD_Poster_C focused on the *Bradycardia* and treatments (e.g., pacemakers). Poster_D, however, changes the topic to member's personal question and attempted to engage in a side discussion with the MD_Poster_C. However, the original poster, Poster_A, controlled the topic and changed it back to the unanswered question by elaborating on the situation. In our qualitative analysis, we found that members with defined roles also provided effort to counteract topic drift; however, they also went along with the current topic of conversation. In contrast, active original posters provided the most effective counteractive efforts.

Results for RQ4. Can Local Topic Drift be Detected Automatically?

We automatically measured local (i.e., thread level) topic drift based on a term-based relevance measurement. Figure 1 shows the general trend of relevance scores as threads evolved for all seven communities. The x-axis indicates position of the posts in threads, and the y-axis indicates average relevance scores for posts in that position across the seven communities. We captured average relevance scores if the position had more than 50 data points. We applied logarithmic regression, which resulted in a relatively high r-squared value of 0.93. Individual WebMD community showed a similar trend in which the topic gradually drifted as conversation progressed. This pattern aligns well with an existing manually assessed topic drift study (Lambiase 2010) as well as our qualitative findings of how most topic drift occurred through semantic parallelism.

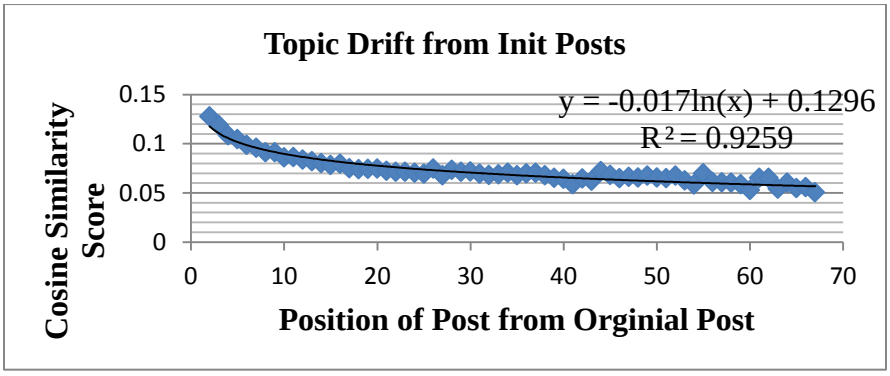

Figure 1. The general trend of topic drift in seven WebMD communities

Next, we evaluated the automated topic drift detection technique. After applying the criteria, the gold standard was furthered narrowed down to 74 positive cases and 73 negative cases. Our evaluation against the gold standard showed promising results as an application to track topic drift as shown in Table 4 Automatically detecting topic drift through relevance measurement achieved an F1-score of 0.71, recall of 0.72, precision of 0.70, and accuracy of 0.70.

Table 4. Confusion matrix of automated topic drift detection technique

		Gold Standard	
		Positive	**Negative**
Relevance Score	**Positive**	53	23
	Negative	21	50

Results for RQ5. Can the Effort to Stay On Topic be Detected Automatically?

By detecting irregular increases in relevance scores in threads (i.e., counteraction), the automated technique achieved an F1-score of 0.73, precision of 0.72, recall of 0.75, and accuracy of 0.74 (Table 5).

Table 5. Confusion matrix of automatically detecting counteraction to topic drift

		Gold Standard	
		Counteraction	**Topic Drift**
Relevance Score	**Counteraction**	18	7
	Topic Drift	6	19

Next, we used the automated technique to determine who provides most counteraction to topic drift. Table 6 summarizes how each type of member provided counteraction to topic drift. For example, 6,233 original posters posted again in their own threads. On average those original posters posted counteracting posts 61% of the times. Their effort to stay on topic exceeded that of any other user, similar to our finding in the qualitative analysis of RQ3.

Table 6. Mean of normalized counts of counteraction, standard deviation, and number of members for different types of member

	Normalized Average Counteraction	Standard Deviation	Num. of members
Original posters	0.61	0.42	6,233
Staff/MD	0.46	0.26	33
Power-User	0.53	0.13	94
Regular-User	0.35	0.46	33,469

Original posters provided significantly higher rates of counteraction compared to the other types of members (Table 7). In contrast, regular users provided significantly lower rates of counteraction compared to other types of members. Lastly, the difference between Staff/MD and Power-user was found not significant (P=0.05001).

Table 7. A pair-wised comparison of different types of members providing counteraction to topic drift using Mann-Whitney-U tests

		U	P-value
VS. Original posters	**Staff/MD**	130,339	0.005
	Power-User	358,110	<0.001
	Regular-User	1.3E+08	<0.001
VS. Staff/MD	**Original posters**	75,350.5	0.005
	Power-User	1,194	0.05
	Regular-User	691,542	0.003
VS. Power-User	**Original posters**	2,058,326	<0.001
	Staff/MD	1,908	0.05
	Regular-User	227,792	<0.001
VS. Regular-User	**Original posters**	1,087,761	<0.001
	Staff/MD	412,935	0.003
	Power-User	73,909,664	<0.001

4.3 DISCUSSION

This chapter sheds light on how members of the WebMD communities react to topic drift, how topic drift unfolds in the communities, and who brings topics back to the original goals of threads. We also address gaps in previous literature by illustrating possible benefits of having off topic discussion, highlighting counteraction provided by different community members, and applying an automated method to detect topic changes at a thread level.

Topic drift occurred in our online community data at two levels: global (i.e., community-wide) and local (i.e., thread-specific) levels. Previous studies associated topic drift from global goals with incoherence (S. Herring 1999) and enforcing conversational participants to stay on global topics as difficult (Kollock and Smith 1996). Moreover, off-topic discussion at a global level has shown tendency to escalate the volume of conversation and force other participants to unsubscribe or remain inactive (Lambiase 2010).

However, in our online health communities, we found topics generally stayed within the global level (i.e., topics related to specific communities) with exception of *OT* or *off topic* titled threads that purposefully discuss off topic materials. These off topic discussions, global level topic drifts, were supported by both power-users and moderators. The off-topic discussion supporters, however, advocated to indicate the off topic nature of threads in the title so that the threads would not interfere with other discussions that pertained to the global goals of the communities. The supporters voiced the opinion that off-topic discussions can build rapport and bring members closer. This may not be representative reactions of all topic drifts, because we only focused on self-identified topic drifted threads. However, reactions in our other qualitative analyses (RQ2 and RQ3) threads were minimal other than the original posters' effort to bring conversation back to the purpose of the thread.

Many off-topic discussions were lively and humorous. In contrast, the tone of many on-topic discussions was melancholy and serious. We found that well-managed off-topic discussions, global level topic drifts, could positively affect online health communities. However, we did not find evidence that regular-users also support off-topic discussions. We suspect that only experienced members (e.g., high level of active participation or defined community roles) know about the culture of having off-topic discussion because we came across posts that asked about having *OT* in the title. Given that many community members asked about the meaning of *OT*, we

suggest that designers and managers of online health communities consider more structured ways to have off-topic discussions even for the new members. An intuitive structure for having off-topic discussions could build rapport and lighten the general mood of the community as well as contribute to sustaining active participation, which has been shown as a prominent challenge for online communities (Millen and Patterson 2002; Nonnecke and Preece 2000).

Although power-users and moderators supported off topic discussions, members in general reacted negatively towards severe local topic drifts. We observed two types of local topic drift (i.e., thread level): gradual topic drift and severe topic drift. Gradual topic drift, in which only a fraction of topics changed through a semantic parallel, occurred most frequently. This change is common and expected in any conversation (Dorval 1990) including CMC (Lambiase 2010). However, members reacted negatively towards severe topic drift—in which previous topics were completely discarded and replaced with different topics. When severe topic drift occurred, original posters showed frustration, and few community members even tried to convert the topic back to the original one. Although complete avoidance of severe topic drift could be difficult, two causes of severe topic drift can be prevented with improved design. For example, in our data, we discovered that few members changed topics completely, perhaps because starting a new thread and sending a private message are not intuitive interactions. An intuitive interface supporting such interactions might reduce severe topic drift.

Severe topic drifts can be further controlled with our automated topic drift detection technique. The automated technique can inform community moderators of severe topic drift and allow a structured system to provide the adequate information and support that original posters seek. Another use of our automated technique would be to alert community members when their posts are off topic. Raising self-awareness could help to control severe topic drifts. As for the community, a similar relevance measurement technique with respect to the goals of the community could be a basis for filtering spam or abusive content, while keeping relevant content available to the community. The automated technique performed reasonably well; however, an extended evaluation using a large dataset could deepen our understanding of the technique in tracking topic drift at the thread level.

We provide evidence that original posters provide more effort for counteracting topic drift than other members including MDs and moderators. We acknowledge that our large sample size could have inflated the significance levels and raises questions to the practical significance of our

findings. However, both qualitative and quantitative analyses showed consistent results in a diverse group of online health communities. This trend could indicate that original posters have a higher stake in keeping the thread on topic. However, this finding could also be due to differences in the responsibilities of moderators and other types of community members. From previous research, we expect moderators to recruit new members, temper discussions, and create an engaging and respectful community culture (Wenger, McDermott, and Snyder 2002). Although we are uncertain of the specific obligations of WebMD moderators and MDs, it is reasonable to assume that they attend to many threads to create an engaging and respectful community culture. Due to demanding responsibilities, the moderators and MDs may not always notice topic drift in threads. Conversely, original posters might be more invested in their own thread, thus investing more effort to keep the topic aligned to their interest to get the needed information or support. Future work using mixed methods, such as surveys and interviews, asking about the responsibility of the WebMD moderators and MDs, could lead to a deeper understanding of this finding.

One solution for the community to regularly provide relevant replies would be automatically offering previously written posts on a similar topic (Ackerman and Malone 1990). Automatically providing relevant information can reduce the responsibilities of moderators while serving members who experience a severe topic drift.

In future work, we plan to test other sophisticated relevance measurements, such as knowledge-based (Leacock and Chodorow 1998) and corpus-based (Turney 2001) approaches. Although the relevance measurement has not been developed to analyze conversations, our study showed consistent results across the seven diverse WebMD communities and a practical application in analyzing CMC. These sophisticated relevance measurements that consider semantic meaning or syntactic organizations of the words could result in a highly accurate model for detecting topic drift.

The significance of this study goes beyond CMC based online health communities. Our finding could generalize to CMC based discussions on other serious topics. One way to further bolster our claims would be to replicate the findings in different types of CMC.

4.4 SUMMARY AND CONCLUSION

We provide new insights on topic drift by illustrating possible benefits of having global topic drift in online health communities, identifying sources of severe local topic drift, highlighting counteraction provided by original posters, and applying an automated method to detect topic drift and counteraction at the thread level. To understand topic drift in health context, we first qualitatively examined topic drift in online health communities on WebMD. Our qualitative analysis revealed that members reacted negatively towards local topic drift in the middle of the thread, but advocated sharing globally off topic stories to build rapport and bring members closer. Also, we observed that confusing online interfaces led to severe local topic drift. Moreover, we observed that original posters, compared to other types of community members, provided more effort to keep their threads on topic. Then, we demonstrated automated techniques to detect both topic drift and its counteracting effort and then quantitatively examined topic drift by measuring similarity between posts. By using an automated method based on term similarity, we achieved F1-score of 0.71 and 0.73 for detecting topic drift and its counteracting effort to stay on topic respectively. These findings suggest that an automated tool could help detect topic drift, support counter efforts to bring the conversation back on topic, and improve communication in these important online communities. Moreover, the findings from this chapter have potential to reduce topic drift and improve online health community members' experience of gathering health information through CMC. Improved CMC has significant potential to improve personal health care management of members who seek essential information and support during times of difficulty.

Chapter 5: HOMOPHILY OF VOCABULARY USAGE: BENEFICIAL EFFECTS OF VOCABULARY SIMILARITY IN ONLINE HEALTH COMMUNITIES

Gathering health information through CMC in online health communities is closely related to other members' active participation (i.e., interaction with other members by CMC) and their effort to disseminate the health information. As previously mentioned in Chapter 2.1, the majority of health information available in online health communities generates from peer members, thus without active participation of those peer members, health information could not be disseminated and the community would cease to exist. In Chapter 5, I describe how homophily encourages active participation, which allows members to reap the benefits of readily available health information. In the following subsection 5.1, I describe three new research questions that pertain to homophily of vocabulary usage in online health communities as well as investigation methods for each of the research questions.

5.1 RESEARCH QUESTIONS AND METHODS

Our overarching research goal of this chapter is to understand participation in online health communities, in particular the relationship between vocabulary similarity of received replies and the member's future interaction in the community. We define **original post** as a post that starts a thread and **original poster** as the author of the original post. Similarly, we define **first reply** as the first post to reply to an original post and **respondent** as the author of the first reply. If the original poster or respondent uses multiple posts consecutively, we considered the accumulation of those posts as the original post or first reply, respectively. For example, original posters and respondents occasionally add comments in subsequent posts before any other member replies. Hence we included any supplementary posts as a part of the original post/first reply. We define **reengagement** as the behavior of original posters returning back to threads they started and having further conversation with the respondent (i.e., by posting a reply). Conversely, we defined **disengagement** as the behavior of original posters not posting a reply to that thread.

In our analysis, we restricted our focus to the first reply for two reasons. First, we wanted to pick the post with the highest chance of reaching the original poster. First replies appear for the

longest time compared to other posts in the thread. Hence original posters have the longest time to read first replies. Second, systematically assessing who is responding to whom is difficult without analyzing the content of each post. For instance, the third person to post (the second replier) could be interacting with the respondent or the original poster. Those posts that are not replying back to the original post could skew the results; thus, we focused our analysis on first replies.

We reviewed common approaches from information retrieval that could be used to quantify **vocabulary similarity** that would represent homophily of vocabulary usage between original posters and respondents. We decided to use a vocabulary-based *cosine similarity* measurement without any feature reductions (e.g., removing common words) to quantify vocabulary similarity score. We chose to use cosine similarity because it is one of the most common and thoroughly studied measures (Singhal 2001). One advantage of cosine similarity over other text similarity measures, such as *Jaccard similarity*, is that cosine similarity normalizes the text length during the comparison. Thus, long first replies would not necessarily be considered to have higher number of shared words. To determine the cosine similarity between original posts and first replies, we first represent each post as vector in N-dimensional space, where N is the number of unique terms across all posts and the value is the frequency with which terms occur in that post. Cosine similarity measures the cosine of the angle between two vectors representing the posts. The resulting similarity score ranges from zero to one. A score of zero indicates no shared terms between the two posts, whereas a score of one indicates all terms and the relative proportion of the terms used are exactly equal.

RQ1. What is the Relationship Between Receiving Replies Written Using a Similar Vocabulary and the Original Posters' Subsequent Thread Engagement?

To examine RQ1, we investigated whether original posters reengaged or disengaged in the threads given the vocabulary similarity score of first replies to original posts. We applied statistical tests (i.e., Pearson's Chi squared test (X^2) and t-tests) to determine whether original posters who received replies with higher similar vocabulary scores reengaged more often. Thus, we compared the mean vocabulary similarity score among original posters who reengaged with the mean vocabulary similarity score among original posters who disengaged.

Next, we used logistic regression to predict the likelihood of original posters reengaging in their threads given vocabulary similarity score. Logistic regression is a statistical technique for predicting dichotomous outcome variables (i.e., engagement) given one or more predictor variables (i.e., vocabulary similarity scores). Logistic regression limits the range of outcome variables from zero to one, satisfying assumptions for dichotomous outcome. Then, we tested the overall effects of vocabulary similarity score using Wald test.

RQ2. What is the Relationship Between Receiving Replies Written Using a Similar Vocabulary in the Early Stage of Joining the Community and the Newcomers' Sustained Community Participation?

We applied survival analysis to examine the relationship between newcomers receiving replies written using a similar vocabulary to their own posts in the early stage of joining the community and the newcomers' sustained participation in the community over time. To identify the newcomers to study, we selected members who contributed at least one original post and received at least one first reply in their newcomer stage. The threshold for the newcomer stage was defined as up to three original posts. This threshold was picked because members with less than three posts were considered lurkers, who are not yet a regularly contributing member, in a different study (Nonnecke and Preece 2000).

Survival analysis is a time duration analysis that models survival time until the failure event occurs. We define the survival object (i.e., "sustained participation") as the period of time in which members continue to participate in the online health community. Defining survival time with respect to online participation can be difficult because the failure event cannot be as clearly defined as in other fields, such as biological and medical sciences where survival analysis has been widely used. In the context of online health communities, members can always return to the community after years of absence as long as the community is active. We adopted a definition of a failure event from Wang et al. (Y. Wang, Kraut, and Levine 2012) to be a period of inactivity of three months without posting to the community. We considered members' first post (i.e., either original post or replying post in threads) as the starting point of their participation in the community and their last post as the end of their participation. However, if members posted within three months of the data collection date, we considered them right censored (i.e., member

who did not experience the failure event) because they might still be actively participating in the community. We calculated the survival time as the days between members' first and last post.

RQ3. What Factors Other than Homophily in Vocabulary Usage are Correlated with Active Participation in Online Health Communities?

We selected a random sample of 100 original-first reply post pairs by selecting 10 pairs that reengaged and 10 pairs that disengaged in each of the five communities. We manually examined these 100 threads to examine other factors related to active participation. We drew on findings from previous studies to guide our content analysis. Previous studies have shown types of social support (Cutrona and Suhr 1994) sought by original posters (e.g., informational or emotional), types of social support that original posters received (Y. Wang, Kraut, and Levine 2012), length of original posts and first replies (Adamic et al. 2008; Agichtein et al. 2008), and rhetorical elements (i.e., asking questions (Arguello et al. 2006)) are associated with participation.

In addition, we considered *coverage of information*—whether replies address all of the concerns expressed in original posts. In information retrieval, cosine similarity is used to measure the similarity of two documents with respect to their subject (Singhal 2001). Because cosine similarity can calculate homophily of vocabulary and similarity of two documents—a proxy for coverage of information—we investigate a possible correlation between information coverage and active participation to have a deeper understanding of the effects of homophily in vocabulary.

We blindly examined the effect of these factors on future interactions with respondents. Furthermore, we explored the purpose of original posters' reengagement. The review of types of emotional support and the purpose of original posters' reengagement followed an open coding process (Strauss and Corbin 1990), which is a method used to elicit unknown, emerging themes grounded in data. For informational support, we assessed whether the original posters were seeking information.

5.2 RESULTS

First reply distribution

First replies in our data set were posted within a day (mean of 21 hours), and 99% were posted within a week. Although replies posted later have an increased chance of not reaching the

original posters, the reengagement rate for late replies that came after a week (9%) was comparable to the entire dataset's reengagement rate (17%). Thus, all first replies were included in the following analysis.

Results for RQ1. What is the Relationship Between Receiving Replies Written Using a Similar Vocabulary and the Original Posters' Subsequent Thread Engagement?

As is the case in most online health communities, our data is not normally distributed. As shown in Table 8, we observed several instances in which vocabulary similarity scores were zero. Zero vocabulary similarity scores often resulted from the limited terms posted by respondents in first replies (e.g., *"Find another doctor"* or *"no comment"*). Original post and first reply pairs with zero vocabulary similarity scores had mean of 14 terms (Standard Deviation (*SD)*=29) in the first replies whereas the pairs with non-zero vocabulary similarity scores had significantly higher mean of 129 terms (*SD*=134; $t(800)=70.21$, $p<0.0001$) in the first replies. Because short generic replies are common in online communities, it was important that we keep the posts with zero similarity scores. We solved the high number of zero vocabulary similarity scores problem by fitting the data into a two-part model (i.e., zero-inflated continuous data model). We then analyzed original post and first reply pairs with zero and non-zero vocabulary similarity scores separately.

Table 8. Proportions of original post-fist reply pairs with zero and above zero vocabulary similarity score by type of engagement by the respondent

	# zero scores (%)	**# non-zero scores**	**Total # of pairs**
Reengagement	51(1%)	4,330	4,381
Disengagement	336 (2%)	15,782	16,118

In the first part of the two-part model, we compared reengagement associated with vocabulary similarity score of zero versus non-zero. We compared the reengagement rate between the zero data set and the non-zero data set, which was significant ($X^2(1, N=20,499) = 15.27$, $p<0.0001$). Thus, having any vocabulary similarity score is associated with significantly higher reengagement.

In the second part of the model, we applied t-tests to the non-zero portion of data to compare reengagement and disengagement by the original poster. We applied these parametric tests because the non-zero portion of data is normally distributed.

Overall, t-tests show significantly higher vocabulary similarity score for reengagement compared to disengagement by the original poster in all communities except for Heart Disease (Table 9). Table 9 also shows the percentage of times that original posters reengage later in the thread after multiple posters have posted. This data shows a full picture of original posters' engagement behaviors. We suspect that no difference was found in the Heart Disease community because of its overall higher rate of disengagement. In the Heart Disease community, original posters disengaged 81% of threads, which is 20% higher than the average disengagement rate of 60% in the other four communities.

Table 9. Comparison of mean similarity and by type of engagement, and percentage of reengagement by original posters, disengagement by original posters, and reengagement by original posters later in the thread after multiple posters have posted

	Mean (SD) vocabulary similarity for reengagement	**Mean of vocabulary similarity for disengagement**	**Comparison of vocabulary similarity score: Reengagement vs. disengagement**	**% of reengaging**	**% of disengaging**	**% of reengaging later in thread**
All five communities	0.38(0.15)	0.35(0.15)	$T(6,637)=13.45$ $p<2.2e-16$	17%	63%	20%
ADHD	0.42(0.16)	0.38(0.15)	$t(370)=4.07$ $p=5.8e-05$	15%	76%	9%
Diabetes	0.34(0.15)	0.32(0.15)	$t(2,186)=5.58$ $p=2.687e-08$	19%	52%	30%
Heart Disease	0.38(0.13)	0.37(0.12)	$t(680)=1.58$ $p=0.11$	14%	81%	5%
Pain Management	0.43(0.15)	0.37(0.15)	$t(1,242)=9.61$ $p < 2.2e-16$	19%	63%	18%
Sexual Health	0.39(0.16)	0.34(0.15)	$t(2,351)=10.80$ $p < 2.2e-16$	17%	62%	20%

Next, we used logistic regression with one predicator variable—vocabulary similarity score—to predict the likelihood of original posters reengaging in their threads. Figure 2 shows the plot of the predicted probability with 95% confidence intervals of reengagement given the vocabulary similarity score between original post and first reply. This regression model was significant ($X^2(1)=91.43$, $p=1.55e-43$). Using the vocabulary similarity score, we predicted future

participation with 79% accuracy in a 10-fold cross validation. The Wald test indicates that for a one unit increase in vocabulary similarity score, the odds of original posters reengaging increases by a factor of 4.7 ($X^2(1)=188.60$, p=6.43e-43).

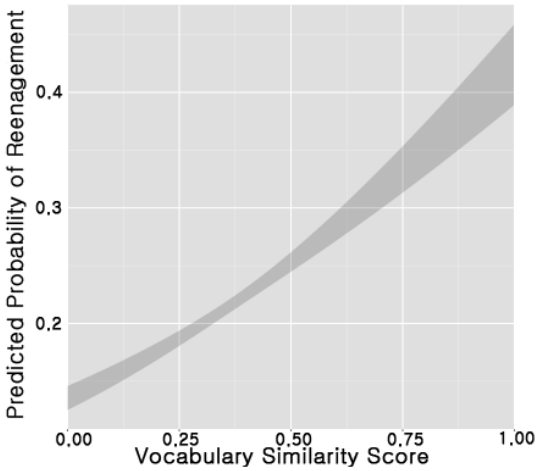

Figure 2. Predicted probabilities of reengagement graph with 95% confidence intervals.

Results for RQ2. What is the Relationship Between Receiving Replies Written Using a Similar Vocabulary in the Early Stage of Joining the Community and the Newcomers' Sustained Community Participation?

We applied survival analysis to test the effect of receiving replies written using a similar vocabulary on members' sustained participation in the community. We partitioned members into three equally sized groups corresponding to members exposed to replies with a "High," "Medium," and "Low" vocabulary similarity score. For members with more than one original and corresponding first reply, we took the average vocabulary similarity score among the first three original and corresponding first replies. Low vocabulary similarity scores ranged from 0 to 0.28, Medium scores ranged from greater than 0.28 to 0.41; and high scores ranged from greater than 0.41 to 0.83, which was the highest vocabulary similarity score in our dataset.

Figure 3 illustrates the effect of receiving replies written using a similar vocabulary on members' sustained active participation. Members in the High group were most likely to stay active in the community, followed by members in the Medium group, followed by members in the Low group as least likely to stay active. These differences were sustained between the high and low groups beyond 300 days.

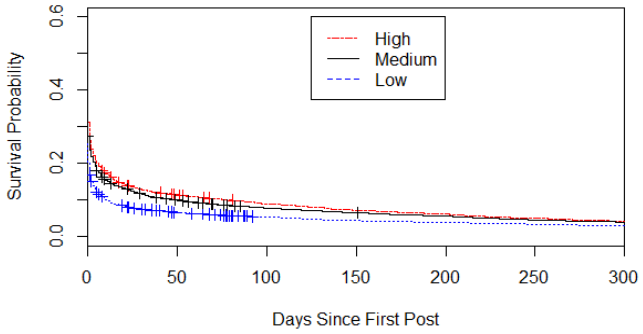

Figure 3. Survival curves for members exposed to high, medium, and low levels of vocabulary similarity in replies

Results of two survival models are shown in Table 10. Model 1 reports the effects of the covariates. For instance, the hazard ratio of 0.75 for the total number of original posts indicates that those who initiate threads one standard deviation more have a 34% (i.e., (1/0.75) – 100%) higher survival rate. Similarly, Model 2 shows that members who received replies with a vocabulary similarity score of one standard deviation higher have a 5% (i.e., (1/0.95) – 100%) higher survival rate when controlling for covariates.

The hazard ratio indicates the odds of members dropping out of the community (encountering the failure event). We also considered a number of covariates and their relationship to sustained participation in two survival models. We selected covariates representative of each member's intrinsic characteristics (e.g., sociable) as well as the amount of participation in the community. These variables include the total number of posts, total number of first replies provided, total number of first replies received, and total number of original threads. We normalized variables (i.e., (observation – mean)/standard deviation) to show predicted change in odds for a unit increase in the predictor.

Table 10. Survival analysis showing influence of covariates in two models

	Model 1		Model 2	
Covariates	**Hazard Ratio**	**Standard Error**	**Hazard Ratio**	**Standard Error**
Total number of posts	0.91**	0.034	0.92**	0.034
Total number of first replies provided	0.92**	0.032	0.92**	0.031
Total number of first replies received	0.83*	0.091	0.84	0.091
Total number of original threads	0.75**	0.095	0.74**	0.094
Vocabulary similarity scores			0.95***	0.008

*****: p<0.001, **: p<0.01, *: p<0.05**

Results for RQ3. What Factors Other Than Homophily in Vocabulary Usage are Correlated with Active Participation in Online Health Communities?

Without any knowledge of their vocabulary similarity scores or reengagement status, we manually categorized original post and first reply pairs into three groups: high, medium, and low coverage groups. We categorized pairs with first replies that addressed all of the concerns expressed in original posts as **high coverage**, first replies that addressed some concerns as **medium coverage**; and first replies that did not address any concerns as **low coverage**. We then examined how well the vocabulary similarity measures performed compared to manual categorization. High coverage group compared to low coverage show significantly higher vocabulary similarity scores ($t(19)=2.58$, $p=0.02$) (Table 11). However, the difference between high coverage group and medium coverage group ($t(20)=0.34$, $p=0.74$) as well as the difference between medium coverage group and low coverage group ($t(29)=1.63$, $p=0.11$) were not found significant.

Table 11. A comparison among subjective and vocabulary similarity scores

	Mean of high coverage (*SD*)	**Mean of medium coverage (*SD*)**	**Mean of low coverage (*SD*)**
Vocabulary similarity scores	0.40 (0.14)	0.39 (0.18)	0.29 (0.16)

In the 50 reengaging pairs, we found that 86% provided high coverage, 8% provided medium coverage, and 6% provided low coverage. In contrast, of the 50 disengaging pairs, 52% provided high coverage, 24% provided medium coverage, and 24% provided low coverage. We applied Pearson's Chi squared test (X^2) to compare the equality of information coverage proportions between reengaging and disengaging. We found reengaging had a significantly higher (X^2(1, N=100)=11.97, p=0.0005) proportion of high coverage replies. As well, disengaging had a significantly higher (X^2(1, N=100)=5.02, p=0.03) proportion of low coverage replies However, equalities of medium coverage proportions was not found significant (X^2=3.65, DF=1, p=0.06).

Similarly, when we compared pairs associated with reengagement and with disengagement, we found a noticeable difference in the length of the posts measured by number of words. The pairs associated with reengagement used more words in both original posts and first replies. On average in the reengaging pairs, original posters used 198 words (SD=152) while respondents used 181 words (SD=150). In contrast, in the disengaging pairs, original posters used 122 words (SD=116) while respondents used 137 words (SD=142). The difference between reengaging and disengaging pairs was significant in original posts (t(92)=2.78, p=0.006) but not significant in first replies (t(98)=1.48, p=0.14).

We also found differences in emotional support: whether respondents indicated an aspect of empathizing with or helping original posters. We found two themes of emotional support: acknowledgement of members' experience of difficulty (e.g., *"I know exactly how you feel […] I'm in the same boat"*) and encouragement for the current situation (e.g., *"You've got a great*

attitude, and will do well"). In reengaging pairs, 30% of respondents acknowledged the difficulty of the original poster's situation and 34% of respondents encouraged original posters. Conversely, in disengaging pairs, 18% and 28% of respondents acknowledged their difficulty and provided encouragement, respectively. However, differences in these propositions were not significant (Acknowledgement: $X^2(1, N=100)=1.37$, $p=0.24$; Encouragement: $X^2(1, N=100)=0.19$, $p=0.67$).

The overall exchange of emotional support was less frequent than informational support; this fits well with what original posters were seeking. Overall, 89% of the original posters asked for new information, while only 24% of the original posters showed any signs of requesting emotional support (e.g., *"can't take more"* or *"desperate!"*). Exchanges of emotional and informational support were not mutually exclusive as some original posters sought both.

We also examined the effects of respondents asking questions to original posters in first replies. In reengagement, 32% of respondents asked a question in their reply, while only 20% of respondents asked a question in disengagement, however, difference in the proposition were not significant ($X^2(1, N=100)=1.30$, $p=0.25$). Similarly difference in proportion of providing informational support was not significant ($X^2(1, N=100)=0.07$, $p=0.79$) between reengagement and disengagement. In reengagement, 84% provided informational support, while 80% provided informational support in disengagement.

In our qualitative analysis, we identified four themes for the purpose of original posters' reengagement: (1) providing more information on their situation (82%), (2) thanking the original posters (62%), (3) asking more questions (20%), and (4) getting defensive (8%). These behaviors were not mutually exclusive.

In our sample, reengaging and disengaging pairs varied in the degree of information coverage. First replies that covered more aspects of the original post received higher vocabulary similarity scores than those that covered fewer aspects. The following is an example of a reengaging pair that received a relatively high vocabulary similarity score of 0.66.

> ***Original poster_A****: "I'd much rather ask others that have a condition I may have... I've been pretty hyper all my life […] and have been bouncing around the idea in my head that I may have some sort of ADHD. […]I've noticed I can focus more when I am very tired. […]"*

Respondent_to_A*: "You are describing Adult ADHD. You should make an appointment with a psychiatrist who was experience with adult ADHD. […] My mind is usually all over the place bouncing from topic to topic […] Things aren't like that now. […] Go. Get tested. […] If you do have ADHD, he may or may not prescribe medication. […]You have nothing to loose as long as your honest with your doctors. Good luck. […]"*

Conversely, the following is an example of a disengaging pair that received relatively low vocabulary similarity score of 0.14.

Original poster_B*: "been having muscle ache between upper left chest (near armpit) through shoulder […] Did I strain a muscle/group of muscles, and what's best - heat, NSAIDS,chiropratic?"*
Respondent_to_B*: "Hello. No way for any of us on an internet message board to know what is causing your symptoms. You will need to see your doctor for an evaluation and treatment plan […]"*

In both examples, the respondents advocate seeking professional help. However, only in the first example, did respondent_to_A provide their experience and perspective. During this process, respondent_to_A covered a number of issues raised by original poster_A while using more shared vocabulary. This conversational pair resulted in a relatively high vocabulary similarity score and elicited original poster_A to reengage in further conversation. In the second example, respondent_to_B does not address the concerns original poster_B raised. This resulted in using less shared vocabulary with original poster_B and a relatively low vocabulary similarity score.

As mentioned earlier, other factors also correlate with original posters' engagement. Respondent_to_A acknowledges the difficulty and encourages original poster_A while using more words (521 words) than respondent_to_B (34 words). Conversely respondent_to_B is succinct and does not provide any emotional support while neither respondent asked questions. Although a combination of many factors can influence engagement of the original poster, in our manual analysis receiving replies with more shared vocabulary appears to be associated with reengagement, which supports our quantitative analysis.

5.3 DISCUSSION AND FUTURE WORK

Prior research shows that sustaining active participation presents a prominent challenge for online communities (Millen and Patterson 2002; Joyce and Kraut 2006; Y. Wang, Kraut, and Levine 2012; Arguello et al. 2006; Nonnecke and Preece 2000). In this paper, we showed the importance of homophily expressed through shared vocabulary associated with members' ongoing engagement in online health communities. Members who received replies that contained more shared vocabulary with their own posts tended to continue their conversations with respondents.

Additionally, we created prediction models that estimate the likelihood that original posters will reengage with respondents. Although many factors can contribute to members disengaging from conversation, our logistic regressions showed that vocabulary similarity predicts future participation with the respondents. Similar prediction models could be one solution for the challenge of sustaining active participation. For instance, our prediction model could allow moderators to identify members who are most likely to disengage. This added knowledge could enable moderators to provide replies that encourage reengagement. Moreover, we discovered that receiving replies written in a similar vocabulary in the early stages of joining the community predicts long-term active participation within the community. One solution for sustaining newcomers' participation in the community is to encourage community members to abide by a set of guidelines, which reflect ideal community member interactions with newcomers. Furthermore, measuring vocabulary similarity can be a basis for automatically alerting members when they deviate from posting replies that encourage reengagement. In contrast with targeting specific members to encourage reengagement, the environment of the community as a whole could be improved by filtering spam or abusive content through comparing terms with the common vocabulary of the community.

In our manual analysis, we found that a combination of many different factors could influence original posters' engagement. Other factors, such as exposure to higher degree of information coverage and higher word counts by original poster(Adamic et al. 2008; Agichtein et al. 2008) were associated with reengagement in our data. All other factors had higher occurrences in reengagement, but they are not found significant. We suspect this is due to small sample size of 100 in our manual analysis. Investigating which factors had the biggest impact on

original posters' engagement is an important question for future work. However, measuring vocabulary similarity of first reply to the original post is a relatively easy and robust technique that seems to measure an important marker for member reengagement.

Furthermore, the vocabulary similarity of the first reply might not be the sole factor of member engagement with respondents. For instance, other components of life can influence online health community participation. Original posters could have gained the needed information through other sources and did not check back with the community. A serious medical crisis could prevent original posters from checking back with the community too, for example. Still, the consistent statistical results from examining the relationship between vocabulary similarity and participation show that homophily of vocabulary provides an important marker for member reengagement.

Our survival analysis examines members' early stages of joining the community and uses a threshold of the first three replying posts that members received based on previous literature on lurking (Nonnecke and Preece 2000). Understanding the transition point of members' participation can provide a deeper understanding of how to sustain active participation in online communities. Also, our analysis does not answer whether inactive members remained lurkers or completely dropped from the community. An analysis of these two different types of inactive members could further extend our findings.

Although overall our analyses showed consistent results in a diverse group of online health communities, we acknowledge that our large sample size could have inflated the significance levels and raises questions to the practical significance of our findings. Also, how much vocabulary similarity is needed for a meaningful increase in reengagement remains an open question. Further investigation, such as surveys or interviews, asking about members' satisfaction could further explore the significance of our findings.

In future work, we plan to investigate the correlation between actual users' perceived qualities of replies and participation to further examine the challenge of sustaining participation in online health communities. Understanding these relationships could provide a more complete view of how to sustain participation in online health communities. The significance of this study goes beyond predicting members' behavior in online health communities. For instance, our findings could generalize in non-health online communities and our automatic approach to analyze computer-mediated communication (CMC) could be applied to other CMC studies.

Furthermore, our study showed a potential method to elucidate the process of forming social bonds through CMC in online communities.

5.4 SUMMARY AND CONCLUSION

Arguably a successful health information gathering depends on other members' active participation and community's capability to sustain active participation. This chapter provides new insights regarding sustaining online health community active participation through systematic analyses of five WebMD online health communities. Our findings suggest that homophily—the vocabulary similarity between members' posts—plays a crucial role in sustained engagement in online health communities. We provide new insights into how vocabulary similarity affects active participation in online communities. Furthermore, vocabulary similarity calculated with cosine similarity shows promising results in measuring the coverage of information in replies. Based on these insights, moderators, online community creators, and online community participants could tailor replies to encourage sustained, active participation by members. Findings from this study can improve member experience in difficult situations when online health communities provide essential support. In Chapter 6, I describe the challenges of extracting health information using existing NLP tools and provide an automatic failure detection system to examine the application of NLP to extract health information in online health community.

Chapter 6: CHALLENGES OF AUTOMATICALLY PROCESSING ONLINE COMMUNITY TEXT

Members in online health communities openly discuss, seek support, and exchange health information (Fox and Rainie 2002) through CMC. The CMC records contain valuable health information for both patients and researchers alike, such as peer members' personal health management experiences. Reusing collective experience and knowledge available in the online health community can be powerful tool (J Huh and Ackerman 2012). For instance, the collective experience and knowledge can be a basis for patient-focused interventions that can empower and guide best decision-making practice. Moreover, collective knowledge can open new opportunities for related research, such as research that aims to predict and tailor individual needs that draw from similar others. Despite these benefits, we lack the necessary capability for reusing collective wisdom: making sense of the vast amount of data as well as extracting relevant health information. Existing biomedical natural language processing (NLP) tools are appealing, most were developed to process clinician or researcher-generated text, such as clinical notes or journal articles. In addition to being constructed for different types of text, other challenges of using existing NLP include constantly developing technologies, source vocabularies, and characteristics of text. These continuously evolving challenges warrant the need for applying low-cost systematic assessment. The primarily accepted evaluation method in NLP, manual annotation, however, requires tremendous effort and time. The primary objective of this chapter is to explore an alternative approach—using low-cost, automated methods to detect when processing patient-generated text with existing biomedical NLP tools. In the following subsections, I first characterize common failures that NLP tools can make in processing online community text. I then demonstrate the feasibility of our automated approach in detecting these common failures using one of the most popular biomedical NLP tools, MetaMap. Finally, I evaluate the automated approach.

6.1 METHODS FOR CHARACTERIZING FAILURES

To characterize the types of failures, we assessed MetaMap's output collaboratively through iterative rounds of manual review among my research team members. We reviewed the output

following an open coding process (Strauss and Corbin 1990) to identify emerging themes grounded in data. Because we did not know all possible failure types, we chose to use an inductive coding process, rather than a structured, reductive content analysis approach. In each iteration, we processed different patient-generated posts and analyzed the mapped terms by examining the corresponding Unified Medical Language System (UMLS) concept definitions and semantic types. We manually evaluated the accuracy of each mapping in each of the iterations. Based on the list of inaccurate mappings, we grouped each inaccurate mapping into failure types and went on to identify potential causes within each failure type through the open coding process. This second step addressed the gaps in previous literature by identifying a number of causes of the failure types and providing information needed to detect these failures automatically.

6.2 RESULTS FOR CHARACTERIZING FAILURES

From our manual review, we characterized three types of failure: (1) boundary failures, (2) missed term failures, and (3) word ambiguity failures. A *boundary failure* occurred when a single coherent term was incorrectly parsed into two or more incomplete terms. A *missed term failure* occurred when a relevant term had not been identified. A *word sense ambiguity failure* occurred when a relevant term was mapped to a wrong concept. Within these three failure types, we discovered 12 causes of failures. In the sections below, we describe each type of failure and go on to identify potential causes within each failure type.

Boundary failures

Boundary failures, in which a single coherent term is incorrectly parsed into two or multiple terms, are well-documented in biomedical NLP literature (Divita, Tse, and Roth 2004; Pratt and Yetisgen-Yildiz 2003; Kang et al. 2012; Brennan and Aronson 2003; C. A. Smith and Wicks 2008). In this literature, boundary failures are referred to as *overly granular parsing* (Brennan and Aronson 2003) or *split phrasing* (Divita, Tse, and Roth 2004). Our analysis expands our understanding with boundary failures associated with patient-generated text.

Our patient-generated text contained extensive descriptive phrases (e.g., ‘feeling great’) and colloquial language (e.g., ‘chemo brain’), contrasting with typical biomedical text that usually contained concepts from standard terminologies. Theoretically, boundary failures can result from

standard medical terminologies. However, descriptive phrases and colloquial language highlight the parsing problem of biomedical NLP because colloquial language and descriptive phrases that patients use in online health communities cannot all be included in the UMLS. For instance, UMLS included 'feeling sick' as a synonym of a concept, although a similar descriptive phrase 'feeling great' was not included in the UMLS. Consequently in our analysis, 'feeling sick' was recognized as one concept, while 'feeling great' was parsed into two separate terms 'Emotions' and 'Large' delivering different interpretations than intended.

Boundary failures also occurred even when proper concepts were available in the UMLS. For instance, a colloquial term 'chemo brain' was commonly used to describe the single concept of cognitive deterioration of cancer patients after chemotherapy. In our analysis, the term was recognized as two UMLS concepts –'chemotherapy' and 'brain-body part'—even though UMLS contained a concept for 'chemo brain'. From our experience, we inferred that the lack of colloquial language and descriptive phrases concepts in the UMLS as well as standard medical terminologies parser were causing boundary failure when processing patient-generated text.

Missed term failures

Missed term failures occurred when a relevant term was not identified (Divita, Tse, and Roth 2004; C. A. Smith and Wicks 2008). We extended the literature by identifying two causes of missed term failures associated with patient-generated text: (1) community-specific nomenclature and (2) misspellings.

Community-specific nomenclature

Community-specific nomenclature refers to members of a community using terms that either are commonly used in a different way elsewhere or not commonly used at all. In online communities, members frequently create their own nomenclature that, over time, can become vernacular that is well understood in the community (Danescu-Niculescu-Mizil et al. 2013). Community nomenclature poses unique challenges and opportunities for NLP.

In particular, community nomenclature regularly referred to relevant health-related content but resulted in three major challenges. First, many of the community-specific terms were not found in the UMLS. For instance, 'PC' referred to 'Prostate Cancer'; however, this acronym was not contained in UMLS. Second, community nomenclature was typically context and community-specific. For instance, the acronym 'BC' was used for 'before cancer', 'blood count',

or ‘breast cancer’ depending on the context. This type of ambiguous usage was also seen with commonly accepted abbreviations. For instance, ‘rad’ was a common abbreviation for ‘radiation therapy’ in the cancer community, but ‘rad’ could also be used for ‘radiation absorbed dose’, ‘reactive airway disease’, ‘reactive attachment disorder’, or ‘RRAD gene’ depending on the community. Third, novel abbreviations and acronyms constantly showed up in our data, similar to what researchers of online communities found (Danescu-Niculescu-Mizil et al. 2013). For instance, our dataset included newly emerged acronyms that were not included in the UMLS, such as ‘LLS’ and ‘PALS’ for ‘Leukemia and Lymphoma Society’ and ‘Patient Advice and Liaison Service’ respectively.

Misspellings

Previous research showed that patients made more medically related misspellings at a significantly higher rate compared to clinicians (Q. Zeng et al. 2001) . Misspellings in our dataset included typographical errors (e.g., “*docotor*”), phonetic errors that could be associated with lack of familiarity with medical terms (e.g., “*byopsi*” and “*methastasis*”), and colloquial language errors (e.g. “*hooooooot flash*”). Biomedical NLP techniques were typically developed using the correct spelling in training models, thus relevant but misspelled terms were often unrecognized. These unrecognized terms comprised a type of missed exact match (Divita, Tse, and Roth 2004) that consequently become false negatives—terms that should have been recognized but were missed. Although previous research in health information query investigated methods to address misspellings of patient generated medical terms (McCray et al. 2004; McCray and Ide 2000), those methods had limitations because they required (1) correctly spelled medical terms in the database and (2) manual selection of terms among recommended terms.

Word sense ambiguity failures

The most prevalent failure was word sense ambiguity, which occurred when a term was mapped to the wrong concept because the two concepts are spelled the same way, share the same acronym (e.g., ‘apt’, an acronym used for appointment was mapped to organic chemical ‘4-azido-7-phenylpyrazolo-(1,5a)-1,3,5-triazine’) or were spelled the same as one of their acronyms (e.g., a verb ‘aids’ was mapped to ‘Acquired Immunodeficiency Syndrome’). This failure had been identified in previous research (Divita, Tse, and Roth 2004; C. A. Smith and Wicks 2008; Brennan and Aronson 2003; Pratt and Yetisgen-Yildiz 2003; Kang et al. 2012), but these studies

did not examine the causes of this failure. From our analysis, we identified nine causes of failure associated with processing patient-generated text: (1) abbreviations and contractions, (2) colloquial language, (3) numbers, (4) e-mail addresses and Uniform Resource Locators (URL)s, (5) Internet slang and Short Message Service (SMS) language, (6) names, (7) the narrative style pronoun 'I', (8) mismapped verbs, and (9) inconsistent mappings (by word sense disambiguation feature). In the following sections, we describe each cause of word sense ambiguity failures in detail and identify associated semantic types where applicable.

Abbreviations and contractions

Frequent use of standard abbreviations and contractions was common in our online health communities. Online community members frequently used contractions such as 'I'd' or abbreviations such as 'i.e.' in their text. Although the use of these shortened forms was common in informal text, it could be a source of errors for many NLP tools. For example, MetaMap maps 'I'd' to 'Incision and drainage' and mapped 'i.e.' to 'Internal-External Locus of Control Scale' due to partial matches with synonyms. Also, MetaMap was inconsistent with some of its correct mappings for abbreviations. For instance abbreviations for some U.S. States were mapped correctly (e.g., 'AK' and 'WA'), whereas others were often missed even though they were in the UMLS (e.g., CA and FL); and some were mismapped (e.g., Virginia was mapped to 'Alveolar gas volume' when written as 'V.A.').

Colloquial language

Colloquial language, such as 'hi' was prevalent in our dataset and caused many failures. Although these terms are obvious to human readers, we found they were often mapped to incorrect terms in the UMLS. For instance, our previous example 'hi,' rather than being left unmapped, was mapped to 'Hawaii', 'ABCC8 gene', or 'AKAP4 gene', because 'hi' was a synonym for all three concepts. In our analysis, this failure was found with many semantic types, however, terms mapped to the semantic type of 'Gene or Genome' were particularly troublesome because of their unusual naming conventions.

Numbers: dates, times, and other numbers that do not indicate disease status

Our online community posts often contained numbers that convey important information, such as a patient's disease status (e.g., 'stage 3 breast cancer'). Other times, numbers conveyed more logistical information, such as time of day and dates, which were be misinterpreted. For instance,

in the phrase, "*I got there at 4:12pm*" '12pm' was mapped to 'Maxillary left first premolar mesial prosthesis', because it was a complete match for one of its synonyms in the UMLS. Numbers that were used to convey diagnostic information was crucial for the identity of many community members, and such information was often included in an automated signature line (e.g., 'stage 2 grade 3 triple negative breast cancer') at the end of posts. Numbers indicating dates and times often resulted in false positives, whereas health status numbers often resulted in a different failure type (i.e., boundary failure caused by splitting a phrase). We saw this type of failure across many different semantic types, including 'Amino Acid, Peptide, or Protein', 'Finding', 'Gene or Genome', 'Intellectual Product', 'Medical Device', 'Quantitative Concept', and 'Research Activity'.

E-mail addresses and URLs

Online community members frequently mentioned URLs and e-mail addresses in our dataset. They often pointed to websites that they found useful and gave out e-mail addresses to start private conversations. Parts of e-mail addresses and URLs were incorrectly mapped in our analysis. For instance, 'net' at the end of an e-mail address was often mapped to the 'SPINK5 gene', because one of its synonyms was 'nets'. Also, 'en', a language code that referred the English language in URLs, incorrectly mapped to 'NT5E gene', because one of its synonyms was 'eN'.

Internet slang and SMS language

Internet slang and SMS language, such as 'LOL' (i.e., 'laugh out loud' or 'lots of love') or 'XOXO' (i.e., hugs and kisses) are highly prevalent in online community text but not in typical biomedical texts. Although these terms should be obvious to human readers, our analysis showed that Internet slang and SMS language were often mapped to incorrect biomedical terms in the UMLS. In particular, Internet slang and SMS language were often mistaken for gene names, such as the mapping of 'LOL' to the LOX1 gene and 'XO' to the XDH gene. To manage the different variations of concepts, the UMLS comprised many synonyms of terms. Varieties of these synonyms overlapped with commonly used Internet slang and SMS language resulting in word sense ambiguity failure.

Names: first, last, and community handles

The use of names is also prevalent in online community posts, particularly when posts address specific individuals. Community members also often include their first names in a signature line and call out other members by first names or community handles. In our analysis, common first names were often mistaken for UMLS concepts, such as ‘Meg’ being mistaken for ‘megestrol’, ‘Rebecca’ for ‘becatecarin’, ‘Don’ for ‘Diazooxonorleucine’, and ‘Candy’ for ‘candy dosage form’. Each individual name was a complete match for one of the UMLS concepts. We identified these mismatches across multiple semantic types, including ‘Antibiotic’, ‘Biomedical or Dental Material’, ‘Clinical Attribute’, ‘Diagnostic Procedure’, ‘Disease or Syndrome’, ‘Finding’, ‘Hormone’, ‘Injury or Poisoning’, ‘Laboratory Procedure’, ‘Mental Process’, ‘Pathologic Function’, ‘Pharmacologic Substance’, and ‘Sign of Symptom’.

Narrative style of pronoun ‘I’

Patients share a wide variety of personal experiences in narrative form in online health communities. Thus, the use of the pronoun ‘I’ is prevalent in community posts but is a source of misinterpretation. For example, over the course of the study we discovered that ‘I’ is typically mapped to either ‘Blood group antibody I’ or ‘Iodides’ which belong to ‘Amino Acid, Peptide, or Protein’, ‘Immunologic Factor’, or ‘Inorganic Chemical’ semantic types.

Mismapped verbs

One of the most fundamental components of NLP tools is a POS tagger, which marks up words with their corresponding POS (e.g., verb, noun, preposition) in a phrase, sentence, or paragraph. POS taggers are commonly used in NLP and have many different applications, such as phrase parsers. In our analysis, we discovered that MetaMap uses a POS tagger called *MedPost SKR* (Semantic Knowledge Representation) (L. Smith, Rindflesch, and Wilbur 2004) to split text into phrases. However, it did not use the resulting POS information when mapping to the UMLS. Such POS failures could have been overlooked in previous studies using biomedical text due to the fact that words like ‘said’ or ‘saw’ were less prevalent in biomedical literature or even in clinical notes. For our online community dataset, MetaMap improperly mapped terms without discriminating between verbs and nouns. For instance, simple verbs used in past tense, like ‘said’ and ‘saw’, were mapped as the acronym, ‘SAID’ (i.e., Simian Acquired Immunodeficiency Syndrome) and ‘saw’ (i.e., a medical device). Verbs in the present tense were also problematic.

For instance, ‘bow’ and ‘snap’ were mapped to ‘Genu varum’ and ‘Snap brand of resin’, respectively. We observed this type of failure across different semantic types, including semantic types where verbs were unexpected, such as ‘Antibiotic’, ‘Biomedical or Dental Material’, and ‘Pharmacologic Substance’.

Inconsistent mappings

Two great strengths of the UMLS are its broad coverage of concepts and its capacity to distinguish among concepts in fine detail. This ability to provide the precise meaning of concepts is valuable for many applications. However this feature also became a source for inconsistent mappings despite similar usage of terms in our analysis. For instance, the term ‘stage’ was mapped to multiple concepts in our dataset. Community members often employed the term ‘stage’ to describe their cancer status (e.g., ‘stage 4 ER+ breast cancer’). Despite the seemingly similar sentence structures and the usage of the term in the sentence, our findings showed that MetaMap inconsistently mapped ‘stage’ to different UMLS concepts. Six different semantic types were identified for the UMLS concepts mapping to ‘stage’ (Table 12). This is a known failure of MetaMap (Divita, Tse, and Roth 2004), however the severity of the failures shows that addressing word sense disambiguation in patient-generated text may require particular attention.

Table 12. Word sense ambiguity failures: inconsistent mappings of *stage* by MetaMap. The mapped terms are in bold.

Sample Sentence	Mapped Term	UMLS Concept	Concept Unique Identifiers (CUI)	UMLS Semantic Type
*"My father was diagnosed with **stage 2b** pancreatic cancer"*	stage 2b	Stage 2B	C0441769	Classification
*"I'm **stage** 4 SLL and stage 2 CLL"*	stage	Tumor stage	C1300072	Clinical attribute
*"I was dx last year at age 46 with **Stage 1**"*	Stage 1	Stage level 1	C0441766	Intellectual Product
*"Almost seven years ago I was diagnosed with **stage** 1 **breast cancer** at age 36 ½"*	Stage breast cancer	malignant neoplasm of breast staging	C2216702	Neoplastic process
*"My friend was just diagnosed with **Stage** IV cancer"*	stage	Stage	C1306673	Qualitative Concept
*"My mom was diagnosed 11/07 with **stage** IV inoperable EC"*	stage	Phase	C0205390	Temporal Concept

6.4 METHODS FOR AUTOMATED FAILURE DETECTION

To explore automated methods for detecting the three types of failures we identified, we created a tool that applies combinations of dictionary-based matching (Jansen and James 2002; Bureau 1990; Bamford et al. 2004) and NLP techniques (Toutanova et al. 2003; de Marneffe, MacCartney, and Manning 2006; A. S. Schwartz and Hearst 2003; Chapman et al. 2001). We describe this detailed automatic detection process in the following sections.

Detecting Boundary Failures

The following section describes automatic detection of boundary failure due to splitting phrases.

Detecting split phrases

Our tool detected failures caused by incorrectly splitting a phrase through a comparison of MetaMap's *MedPost SKR parser* (L. Smith, Rindflesch, and Wilbur 2004), a biomedical text parser, and the *Stanford Parser* (de Marneffe, MacCartney, and Manning 2006), a general-purpose parser. First, we collected all adjacent terms that MetaMap mapped but MedPost SKR had parsed separately. Second, we employed the Stanford Parser to determine whether adjacently mapped terms were part of the same phrase. If adjacently mapped terms were a part of the same phrase, the combined term could deliver a more precise meaning, while individually they often deliver different meanings (Brennan and Aronson 2003; Divita, Tse, and Roth 2004). We found this especially problematic if the combined term was a valid UMLS concept. For instance, we would collect 'chemo brain', as a boundary failure caused by splitting a phrase. 'Chemo' and 'brain' were terms that appeared adjacent to one another in a sentence, and their combination—'chemo brain'—was a valid UMLS concept, but MetaMap split it into two separate terms. However, we also collected combined terms that were not in the UMLS, because they were also cases of improperly splitting terms. Furthermore, the missing combined terms could provide valuable insight to completeness of the UMLS. For instance, both 'double mastectomy' and 'chemo curls' are important concepts that are frequently used by patients; however, these concepts are missing from the UMLS as shown in

Table 13. The aforementioned steps to compare MetaMap's MedPost SKR parser with the Stanford Parser can detect these important but missing terms. In our detection, we used the shortest possible phrase identified by the Stanford Parser. The Stanford Parser parsed phrases as structure trees to indicate grammatical relations. In the structure tree, a shorter phrase was often a part of a longer phrase and delivered more coherent meanings compared to a longer phrase.

Table 13. Examples of splitting a phrase failure. The split phrases are in bold.

Sample Sentence	Ideally Mapped UMLS concept	First Mapped Term (UMLS concept name)	Second Mapped Term (UMLS concept name)
"*My mom had unknown primary and it was a **PET scan** that helped them find the primary.*"	PET/CT scan	PET (Pet Animal)	Scan (Radionuclide Imaging)
*"It was removed and I have had **stereotactic treatment** along with 6 rounds of Taxol/Carbo completed in January 2012."* [sic]	Stereotactic Radiation Treatment	Stereotactic (Stereotactic)	Treatment (Therapeutic Aspects)
"*Had 25 internal **rad treatments** (along with cisplatin on day 1 and 25).*" [sic]	Therapeutic Radiology Procedure	Rad (Radiation Absorbed Dose)	Treatments (Therapeutic Procedure)
"*I am **Triple Negative BC** and there are no follow-up treatments for us TN's.*"	Triple Negative Breast Neoplasms	Triple (Triplicate)	Negative (Negative)
"*My doc thinks I will probably end up having a **double mastectomy***"	*None available*	Double (Double Value Type)	Mastectomy (Mastectomy)
"*I thought after 9 months my hair would be back but I have grown some type of hair that I am told is "**chemo curls**." "*	*None available*	Chemo (Chemotherapy Regimen)	Curls (Early Endosome)

Detecting Missed Term Failures

We identified two causes of missed term failures associated with processing patient-generated text. The following sections describe automatic detection of missed terms, specifically due to community-specific nomenclature and misspellings.

Detecting community-specific nomenclature

Our tool detected missed terms due to abbreviations and acronyms in four steps. First, it ran MetaMap on the original text and then counted the total number of mappings. Second, it

extracted common abbreviations and acronyms and their definitions using a simple rule-based algorithm (A. S. Schwartz and Hearst 2003), but where we manually verified the extracted terms. Third, it ran MetaMap again after replacing the extracted abbreviations and acronyms with their corresponding fully expanded terms. Finally, it calculated the difference in the total mappings between the original text and the updated text.

The simple rule-based algorithm by Schwartz and Hearst (A. S. Schwartz and Hearst 2003) has performed well in finding abbreviations and acronyms in documents (Baumgartner Jr et al. 2008; Jimeno-Yepes, Berlanga-Llavori, and Rebholz-Schuhmann 2010). We modified the algorithm to reflect typical writing styles of online community posts. The algorithm by Schwartz and Hearst uses: (1) order of characters, (2) distance between abbreviations/acronyms and their definitions, and (3) presence of parentheses to find candidates for abbreviations/acronyms and their definitions. Our tool first identified completely capitalized words (with an exception of the last character due to pluralization) as candidate abbreviations/acronyms and then applied the algorithm to find its fully expanded form. Because online community members adopted community's abbreviations/acronyms, we saved this information and applied to other posts written in the same community even when the definition was not available. For instance, in the sentence, "*My mother was diagnosed with Stage 3 Esophageal cancer (EC) earlier this year - EC also counts smoking and alcohol as two major aggravating factors and is an aggressive cancer,*" the poster defined EC once and then continued to use the acronym in place of esophageal cancer. MetaMap could map esophageal cancer but not EC. Our tool used this algorithm to detect EC and its fully expanded form, esophageal cancer, then replaced EC with 'esophageal cancer' to ensure the concept could be identified by MetaMap.

Detecting misspellings

Our tool detected the prevalence of missed terms due to misspelling using three steps. First, it ran MetaMap on the original text and counted the total number of mappings. Second, it ran MetaMap again after correcting possible misspellings using Google's query suggestion service ("Query Suggestion Service"). Finally, it calculated the difference in the mappings between the original text and the corrected text.

Detecting Word Sense Ambiguity Failure

We identified nine causes of word sense ambiguity failure associated with processing patient-generated text. In the following sections, we describe how to automatically detect the word sense ambiguity failures.

Detecting abbreviations and contractions

To detect word sense ambiguity failures due to abbreviations and contractions, we used an NLP tool called the *Stanford POS Tagger* (Toutanova et al. 2003), which assigns POS to terms in text. Our tool processed the data using the *Stanford POS Tagger* to count cases where a single mapped term was tagged with multiple POS. For instance, the *Stanford POS Tagger* would accurately tag 'I'd' with two different POS, the personal pronoun and modal.

Detecting colloquial language

Detecting word sense ambiguity failure caused by colloquial language is particularly challenging. We identified many of these failures by narrowing our focus to consider only the 'gene or genome' semantic type because colloquial language failures were frequently mapped to this semantic type. Our tool automatically detected improperly mapped colloquial language by using an existing cancer gene dictionary—a list of genes known to be associated with cancer (Bamford et al. 2004)—and counting the number of terms categorized as a 'gene or genome' semantic type that were not in the cancer gene dictionary.

Detecting numbers: dates, times, and other numbers that do not indicate disease status

To automatically detect improperly mapped dates and times, we implemented a number of rule-based regular expressions to detect times and dates that were not mapped as 'Quantitative Concept' semantic type concepts. 'Quantitative Concept' is the most appropriate semantic type based on how patients typically used numbers in our dataset. This resulted in counting the numbers mapped to 'Amino Acid, Peptide, or Protein', 'Finding', 'Gene or Genome', 'Intellectual Product', 'Medical Device', and 'Research Activity'.

In our approach, we recognized two types of date or time expression that are problematic for MetaMap. The first type was a time expression containing the term 'pm'. The second type was a string of numbers that has been typically used to describe age, date, or time duration. For instance, '3/4' indicating March fourth was mapped to a concept describing distance vision (CUI:

C0442757). We used specific regular expressions that focused on numbers with ‘am’ or ‘pm’, as well as a string of numbers with or without non-alphanumeric characters in between numbers to identify dates, times and other number that do not indicate disease status.

Detecting e-mail addresses and URLs

Our detection process for E-mail addresses and URLs was completed in one step. First, we used a regular expression to identify all the email addresses and URLs, and then we counted the number of terms that were mapped from e-mail addresses or URLs. In our approach, we used specific regular expressions matching ‘@’ and a typical structure of domain name (i.e., a dot character followed by two to six alphabetic or dot characters) for identifying e-mail addresses and ‘http’ or typical structure of domain name for identifying URLs.

Detecting Internet slang and SMS language

We detected improperly mapped Internet slang and SMS language using a three-step process. First, we identified an Internet dictionary with a list of chat acronyms and text shorthand (Jansen and James 2002). Second, we manually reviewed the list to remove terms that were also medical acronyms. In this process, we identified only three medical acronyms ‘AML’, ‘CMF’, and ‘RX’ and removed them from the list. Third, our tool automatically identified the terms in the text by matching them with the Internet slang/SMS language list.

Detecting names: first, last, and community handles

To identify improperly mapped names, we first combined a number of name dictionaries that consist of first names (Bureau 1990) with a list of community handles from our online community, CancerConnect.com. Then, our tool counted the number of mapped terms that matched one of the names in the combined list.

Detecting narrative style of pronoun ‘I’

We identified a number of cases where the pronoun ‘I’ was improperly assumed to be an abbreviation, such as for Iodine, because the NLP tool did not consider the contextual knowledge from the term’s POS. One of the most fundamental components of NLP tools is a POS tagger, which marks up words with their corresponding POS (e.g., verb, noun, preposition) in a phrase, sentence, or paragraph. ‘I’ as an abbreviation for Iodine should be recognized as a noun whereas

'I' meaning the individual should be recognized as a pronoun by a POS tagger. Our tool used data derived from the *Stanford POS Tagger* (Toutanova et al. 2003) to count cases where the pronoun 'I' was mapped to either the 'Blood group antibody I' or 'Iodides' concepts. We noticed that the pronoun 'I' was sometimes tagged as a foreign word. We included those cases in our counts, because it was a failure of the *Stanford POS Tagger*.

Detecting mismapped verbs

To identify the improperly mapped terms without discriminating between verbs and nouns, we used POS information from the *Stanford POS Tagger* (Toutanova et al. 2003) to count cases where a mapped verb term belonged to a semantic type that did not contain verbs. The 34 semantic types (e.g., 'Activity' and 'Behavior') listed under the 'Event' tree of the UMLS ontology could contain verbs; thus, we excluded verbs from these semantic types from our analysis. We considered all verbs in the 'Entity' tree of the UMLS ontology as incorrect mappings. The 'Entity' portion includes semantic types, such as 'Biomedical or Dental Material', 'Disease or Syndrome', 'Gene or Genome', 'Medical Device', 'Pharmacologic Substance', for which we do not expect verbs. Thus, our tool detected cases where verbs were associated with the 'Entity' tree of the UMLS ontology.

Detecting inconsistent mappings

Detecting word sense ambiguity failures leading up to this section consisted with cases where terms were consistently mapped improperly. However, for other word sense ambiguity failures, MetaMap inconsistently mapped terms, both correctly and incorrectly. The inconsistency was the result of poor performance by MetaMap's word sense disambiguation feature that was designed to select the best matching concepts out of many candidate concepts available in the UMLS. We detected inconsistent mappings by (1) assuming that patients used terms consistently (2) MetaMap accurately selected the best matching concepts majority of the time. For instance, in our online cancer community dataset, we assumed that patients always used the term, 'blood test', to convey the 'Hematologic Tests' concept (CUI: C0018941), which was how MetaMap interpreted this term two thirds of the time, rather than the less frequent mapping to the 'Blood test device' concept (CUI: C0994779). Based on these assumptions, we detected inconsistent mappings in two steps. First, we created a term frequency table based on a term's spelling and its

CUI. Second, assuming the most frequently mapped CUI was the correct concept, we counted the number of cases where the term was mapped to less frequent CUIs.

6.5 RESULTS FOR AUTOMATED FAILURE DETECTION

The automated methods detected that at least 49% of MetaMap's 383,572 mappings for our dataset were problematic. Word sense ambiguity failures were the most widely occurring, comprising 82% among the total detected failures. Boundary failures were the second most frequent, amounting to 16% among the total detected failures, while missed term failures were the least common, making up 2% of the detected failures.

Table 14 summarizes the identified
failures as well as their causes and prevalence for automatic detection of MetaMap's failure on processing patient-generated text. Our process showed the feasibility of automated failure detection; hence showing the types of failures that our tool could identify in similar datasets processed with MetaMap.

We found that word sense ambiguity failures were not mutually exclusive, and several cases had multiple causes; thus, in

Table 14, the sum of percentages for individual failures exceeded 100%. For instance, an acronym ‘OMG’ used for ‘Oh My God’ was incorrectly mapped to ‘OMG gene’. This particular failure was detected as both colloquial language as well as Internet slang and SMS language failures. To avoid redundant counts, we detected 154,904 unique counts of word sense ambiguity failure, making up 82% of failures. In

Table 14, we show both individual counts/percentages as well as the total unique counts/percentages to provide a precise overview of word sense ambiguity failures. In Figure 4, we illustrate the prevalence of these failure types as well as correct mappings. Of the 383,572 total mappings from our entire dataset, nearly half were incorrect. Of those incorrect mappings word sense ambiguity failures were the most predominant whereas missed term failures were the least predominant. Although these failures were recognized in prior studies on MetaMap (Divita, Tse, and Roth 2004; C. A. Smith and Wicks 2008; Brennan and Aronson 2003; Pratt and Yetisgen-Yildiz 2003; Kang et al. 2012), the studies had not presented automated methods for detecting these failures.

Table 14. Detecting MetaMap's failures on processing patient-generated text

Failure Type	Causes of Failure	Count	Percentage of Failure (in %)
1.Boundary Failures	1.1 Splitting a phrase	29,965	15.9
2. Missed Term Failures	2.1 Community specific nomenclatures	1,167	0.6
	2.2 Misspellings	2,375	1.3
3. Word Sense Ambiguity Failures	3.1 Abbreviations and contractions	416	0.2
	3.2 Colloquial language	4,162	2.2
	3.3 Numbers	143	0.1
	3.4 E-mail addresses and URLs	1,448	0.8
	3.5 Internet slang and SMS language	3,442	1.8
	3.6 Names	10,061	5.3
	3.7 Narrative style of pronoun 'I'	61,119	32.4
	3.8 Mismapped verbs	51,193	27.2
	3.9 Inconsistent mappings	29,308	15.6
	Total number of unique word sense ambiguity failures	154,904	82.2
Total Number of Unique Failures		188,411	

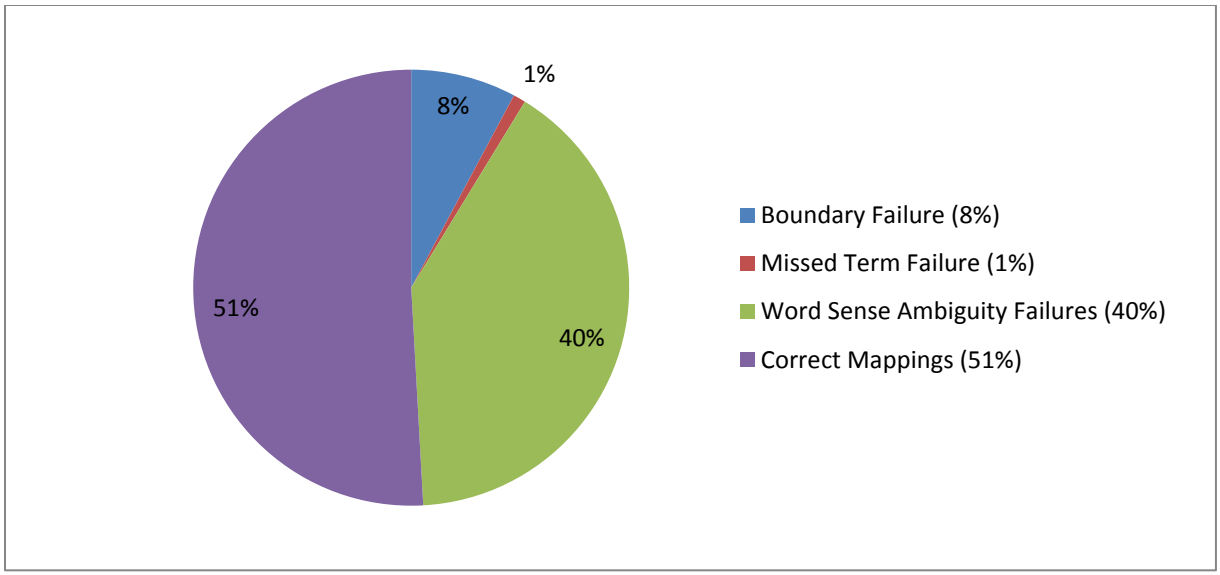

Figure 4. Prevalence of failures from the application of MetaMap to online community text.

6.6 PERFORMANCE EVALUATION OF AUTOMATED FAILURE DETECTION

We manually evaluated the performance of our failure detection tool in two parts: overall performance evaluation and individual component level performance evaluation.

Methods for Evaluation

We randomly selected 50 cases (i.e., mappings) that our tool identified as incorrect mappings from each of the 12 causes of failures, totaling 600 cases that served as positive cases. We then randomly selected another 600 cases from the rest of the mappings not detected as incorrect mappings according to our tool to serve as the negative cases. We then mixed up the selected 1200 cases, and manually assessed the accuracy of mappings through a blind procedure.

We also measured individual performance on each of the 12 detection techniques. We used the previously selected 600 negative cases and individual technique's 50 positive cases to assess the performance. For boundary failure, we examined whether the mapped terms could deliver precise conceptual meaning independent of additional phrases. For missed term failure, we

investigated whether the tool had accurately corrected the spellings and verified the results of the new mappings. For word sense ambiguity failures, we examined whether MetaMap appropriately mapped terms based on the rest of context. The unit of analysis was a single mapping and we evaluated our results using precision, recall, accuracy, and F1 score.

Evaluation Results

Table 15 shows the performance of the automatic failure detection tool. The failure detection tool achieved overall precision, recall, accuracy, and F1 score of 83%, 93%, 88%, and 88%, respectively. At the individual component level, methods using dictionary-based matching or regular expression matching performed more accurately than methods using existing NLP techniques. In the following sections, we discuss findings of individual component of the automatic failure detection tool and its performance.

Table 15. Performance (in %) of automatic failure detection and its individual component

Failure Type	Causes of Failure	Precision	Recall	Accuracy	F1 Score
1.Boundary Failures	1.1 Splitting a phrase	82.0	78.8	96.8	80.4
2. Missed Term Failures	2.1 Community specific nomenclatures	88.0	100.0	99.0	93.6
	2.2 Misspellings	80.0	93.0	97.9	86.0
3. Word Sense Ambiguity Failures	3.1 Abbreviations and contractions	82.0	95.3	98.2	88.2
	3.2 Colloquial language	100.0	100.0	100.0	100.0
	3.3 Numbers	100.0	100.0	100.0	100.0
	3.4 E-mail addresses and URLs	100.0	100.0	100.0	100.0
	3.5 Internet slang and SMS language	100.0	100.0	100.0	100.0
	3.6 Names	66.0	100.0	97.2	79.5
	3.7 Narrative style of pronoun 'I'	100.0	100.0	100.0	100.0
	3.8 Mismapped verbs	32.0	100.0	94.4	48.5
	3.9 Inconsistent mappings	66.0	53.2	92.8	58.9
Total		**83.0**	**92.6**	**88.2**	**87.5**

Result for detecting split phrases

Our automatic failure detection tool identified 16% of the total failures due to splitting a phrase. The performance evaluation of this task achieved precision, recall, accuracy, and F1 score of 82%, 79%, 97%, and 80%, respectively. It is important to note that a single concept can produce multiple split phrase failures. For instance, the phrase 'stage 4 Melanoma' was mapped to three concepts: 'stage,' '4,' and 'Melanoma'. Two boundary failures occurred in this phrase. The first failure occurred between 'stage' and '4'; the second failure occurred between '4' and 'Melanoma'. By focusing on a pair of mapped terms at a time, we correctly identified two failures that occurred in the phrase 'stage 4 Melanoma'. We only considered adjacent paired mappings because splitting a single coherent phrase into two or more UMLS concepts was

clearly a more significant problem. However, split phrase failures could occur in non-paired mappings as well, and we are underestimating the prevalence of split phrases.

Result for detecting community-specific nomenclature

Less than 1% of failures were due to community-specific nomenclature, and automatic detection system achieved precision, recall, accuracy, and F1 score of 88%, 100%, 99%, and 94%, respectively. It should be noted that we underestimated the number of missed terms because the algorithm (A. S. Schwartz and Hearst 2003) can only identify abbreviations or acronyms if they were previously defined by members at some point. In addition, we would not recognize cases where MetaMap still missed the fully expanded term.

Result for detecting misspellings

We automatically assessed that misspellings were responsible for 1% of failures. However, we observed few cases of incorrect assessment due to failures of Google's query suggestion service. For instance, some medications were incorrectly recommended. 'Donesaub', a misspelling of 'Denosumab' was mapped to 'dinosaur'. Furthermore, even with correct recommendation, MetaMap did not always map to the right concept. For instance, 'Wsihng' was correctly recommended to be 'Wishing' but MetaMap mapped it to 'NCKIPSD gene'. Despite a few cases of incorrect assessment, the misspelling component performed relatively well, achieving precision, recall, accuracy, and F1 score of 80%, 93%, 98%, and 86%, respectively.

Result for detecting abbreviations and contractions

Improperly mapped abbreviations comprised less than 1% of failures. Although this was seldom, the automatic detection system performed relatively well, achieving precision, recall, accuracy, and F1 score of 82%, 95%, 98%, and 88%, respectively.

Result for detecting colloquial language

Incorrectly mapped 'gene or genome' semantic types comprised 2% of failures, and automatic detection system achieved precision, recall, accuracy, and F1 score of 100%, 100%, 100%, and 100%, respectively. With this process, we also detected terms like 'lord' and 'wish' that may not be perceived as colloquial language. Nevertheless, they were improperly mapped as 'gene or genome' semantic type. It is also important to note that different disease-specific communities should utilize different gene dictionaries.

Result for detecting numbers: dates, times, and other numbers that do not indicate disease status

Our automatic failure detection tool identified less than 1% of failures as improperly mapped numbers. The performance evaluation of this task achieved precision, recall, accuracy, and F1 score of 100%, 100%, 100%, and 100%, respectively. However, we are underestimating this failure prevalence because MetaMap improperly mapped more than half of the 'Quantitative Concept' semantic type concepts in our dataset. We did not include this semantic type and underestimated this particular failure because few cases were correctly mapped.

Result for detecting e-mail addresses and URLs

Improperly mapped e-mail addresses or URLs comprised less than 1% of failures, and automatic detection system achieved precision, recall, accuracy, and F1 score of 100%, 100%, 100%, and 100%, respectively. It is important to remind that the basis for our manual assessments was how patients had intended to use the term. For instance, MetaMap mapped 'org' at the end of an URL to 'Professional Organization or Group' concept. Although assessment of such cases can be subjective, we followed the basic rule of reflecting patients' intentions.

Result for detecting Internet slang and SMS language

Nearly 2% of failures resulted from Internet slang and SMS language terms. Like other dictionary-based matching techniques, our automatic detection system performed relatively well, accomplishing precision, recall, accuracy, and F1 score of 100%, 100%, 100%, and 100%, respectively.

Result for detecting names: first, last, and community handles

We automatically assessed that names accounted for nearly 5% of failures, however, the name dictionary matching did not perform as well as other dictionary-based matching components. We discovered that unique but popular names, such as 'Sunday', 'Faith', and 'Hope' were incorrectly mapped when used as nouns in a sentence. The name dictionary component achieved precision, recall, accuracy, and F1 score of 66%, 100%, 97%, and 80%, respectively.

Result for detecting narrative style of pronoun 'I'

Over 32% of failures resulted from pronoun 'I'. Although the use of the pronoun 'I' could be considered a part of colloquial language, we noted it as a different cause of failure due to its high frequency. The automatic detection system accomplished precision, recall, accuracy, and F1 score of 100%, 100%, 100%, and 100%, respectively.

Result for detecting mismapped verbs

We automatically assessed that POS accounted for 27% of failures, however, the POS component performed poorly, achieving precision, recall, accuracy, and F1 score of 32%, 100%, 94%, and 49%, respectively. We discovered that although POS has identified verbs correctly, we made the false assumption that verbs did not belong to the entity part of the UMLS ontology. However, verbs like 'lost' and 'wait' belong to the 'Functional Concept' semantic type, which is under the entity part of the UMLS tree. Thus, the POS component of our automatic failure detection tool incorrectly identified such verbs as failures.

Result for detecting inconsistent mappings

Our automatic failure detection tool identified 16% of the total failures due to inconsistent mappings. The performance evaluation of this task achieved precision, recall, accuracy, and F1 score of 66%, 53%, 93%, and 59%, respectively. We found two reasons for the relatively low precision. First, we did not account for cases where the most commonly mapped concept is not the correct mapping. For instance, in our dataset 'radiation' was mapped to 'radiotherapy research' (CUI: C1524021) two-thirds of the time when community members actually meant 'therapeutic radiology procedure' (CUI: C1522449). We incorrectly assessed if less frequent mappings were accurate. Second, we missed cases when correct mappings do not exist. For instance, the verb 'go' was incorrectly but consistently mapped as 'GORAB gene'. In our automated failure detection analysis, our tool overlooked terms like 'go' that was consistently mismapped.

6.7 DISCUSSION

We first discuss challenges of using out-of-the-box biomedical NLP tools to process patient-generated text. We then discuss the contributions and wider implications of this study for research activities that needs to manage constant change and overwhelming amount of patient-generated data. Lastly, we end with summarizing our contributions to the medical internet research community.

In Figure 5, we illustrate challenges of processing patient-generated online health community text and the common failures biomedical NLP tools can produce. In an example sentence, *"Hi Meg, I wish my docotor would haven't said I'd have chemo brain. It's 12PM and*

I'm signing off! LOL Don," MetaMap produced 12 mappings, all of which were incorrect, and overlooked one misspelled term, 'docotor'; producing 13 failures.

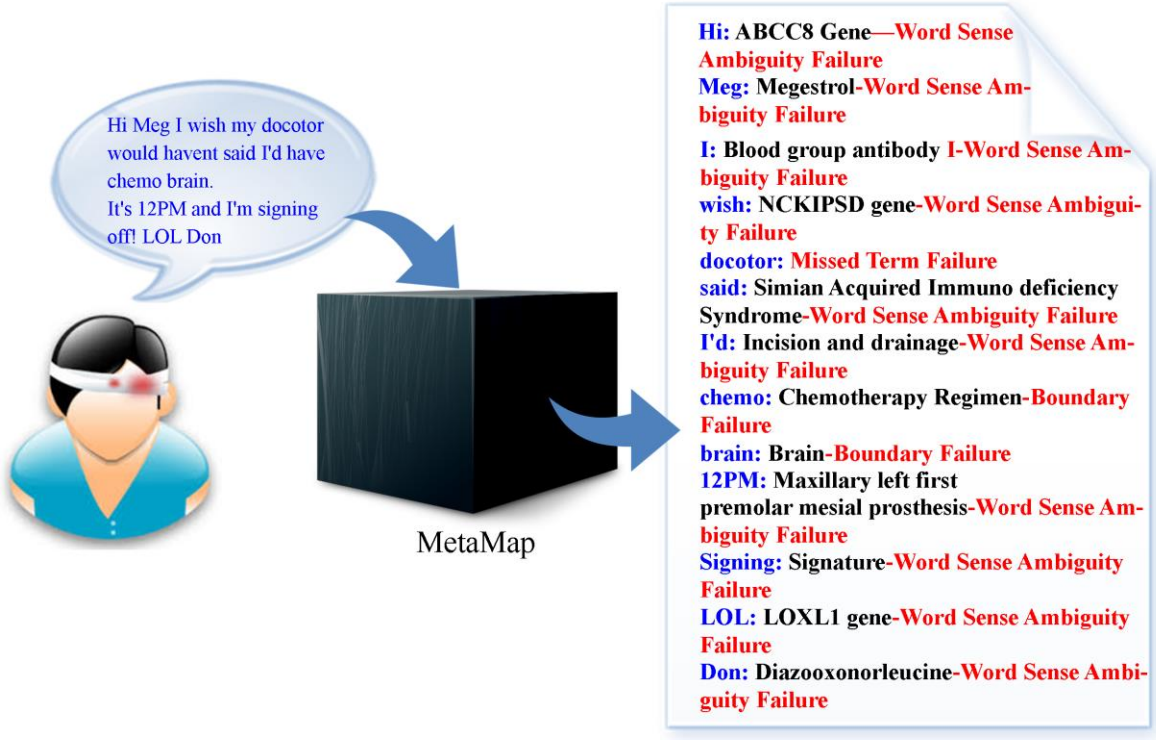

Figure 5. Example failures that resulted from the application of MetaMap to process patient-generated text in an online health community. Blue terms represent patient-generated text; black terms represent MetaMap's interpretation; and red terms represent failure type.

The failures we described in this paper are known, problematic failures of MetaMap (Divita, Tse, and Roth 2004; C. A. Smith and Wicks 2008; Brennan and Aronson 2003; Pratt and Yetisgen-Yildiz 2003; Kang et al. 2012). Our findings extend prior work by identifying the causes for each failure type. Leveraging our understanding of those causes, we developed automated techniques that identified these previously highlighted failures effectively without having to produce manually annotated datasets. In demonstrating the feasibility of our automated failure detection tool, we delineated the use of easily accessible NLP techniques and dictionaries. These techniques can independently examine each failure type. We provided a detailed demonstration of our failure detection tool to allow researchers to select the parts of our

approach that meet the focus of their NLP tool assessment. Additionally, our detection approach can be modified and be used to rectify failures in NLP tools.

Although our study focused on online health community text, the insights inform efforts to apply NLP tools to process various types of patient-generated text, including blogs or online journals, which share similar narrative writing styles and colloquial language. Moreover, Facebook and email provide conversational interactions similar to the interaction in online health communities. Tweets about emergency responses (Starbird and Palen 2011), public health trends (Dredze 2012), or clinical notes from electronic medical records (EMR) could contain a host of abbreviations that NLP tools could incorrectly map. Thus, our failure detection techniques could be applied in these other contexts to assess the capability of processing different types of patient-generated text.

We focused our research on MetaMap, however, findings from our study can apply to other NLP tools in a similar manner. Few failure causes, such as inconsistency of word sense disambiguation feature, may pertain more to MetaMap. However, any NLP tools that provide semantic connections require a similar word sense ambiguity feature. Moreover, different NLP tools could excel in different areas, and our automated failure detection can cost-effectively highlight problematic areas. Similarly, our techniques for detecting failures could strengthen the performance of other NLP tools to process patient-generated text and more traditional types of text. For instance, the word sense ambiguity failure caused by neglecting POS information can also be problematic in different types of text, including biomedical literature. That failure might surface less frequently due to differences in sentence structure between the biomedical literature and patient-generated text. Nevertheless, it is a significant problem that applies to both types of text. Applying such POS information when mapping a term could increase the accuracy of the mappings from a variety of texts. Another example is the missed term failure caused by community nomenclature. MetaMap or other NLP tools will miss terms if particular synonyms are missing from the vocabulary source. Researchers could use the algorithm by Schwartz and Hearst (A. S. Schwartz and Hearst 2003) to collect various synonyms that are used in different domains and frequently update the vocabulary sources, such as UMLS. Furthermore, researchers could use the splitting-a-phrase detection technique to not only prevent boundary failures, but to collect new medical jargon (e.g., ‘chemo curls’) and identify important concepts missing in the UMLS (e.g., ‘double mastectomy’).

The dictionary-based matching and NLP techniques used in our detection process were evaluated in previous studies (Toutanova et al. 2003; de Marneffe, MacCartney, and Manning 2006; A. S. Schwartz and Hearst 2003). However, these studies were conducted in different domains and have been shown to produce errors. Moreover, these tools were not evaluated for patient-generated text. In addition, the automated detection techniques are generally limited to the coverage of the UMLS and MetaMap's capability to map when accurate and full spellings were provided. To strengthen our findings, we evaluated each detection method as well as the overall performance (Table 15). However, our findings could be biased towards cancer community text, and could be further strengthened by generalizing our results in different platforms or patient groups. It is also plausible that we have not encountered all failure types or causes for other patient-generated health data contexts.

Moreover, a number of updates were made for both the UMLS and MetaMap ("MetaMap Updates") since the beginning of our study. To maintain consistency, we continued to use the same versions of the UMLS and MetaMap. However, we used the latest version of MetaMap (2013) and the UMLS (2013AB) to process a sample of 39 posts that were illustrated here, and then compared the results to our findings. The majority (33 out of 39) of the outcomes remained unchanged. Because technologies, source vocabularies, and characteristics of text continue to update in the field of NLP, the need for low-cost automated methods to assess the updates will continuously increase. We demonstrate the feasibility of such automated approaches in detecting common failures using MetaMap and patient-generated text.

Our findings have implications across a range of settings from processing EMR data and popular social media platforms, such as Facebook, to maintaining vocabulary sources like UMLS. Applying existing NLP tools to new situations and types of text requires an in-depth understanding of the benefits and limitations of the tool. Our findings show a number of automated methods to detect biomedical NLP failures at a low cost.

6.8 SUMMARY AND CONCLUSION

Processing patient-generated text provides unique opportunities. However, this process is fraught with challenges. We identified three types of failures that biomedical NLP tools could produce when processing patient-generated text from an online health community. We further identified causes for each failure type, which became the basis for applying automated failure detection

methods using pre-validated NLP and dictionary-based techniques. Using these techniques, we showed the feasibility of identifying common failures occurred in processing patient-generated health text at a low-cost. The value of our approach lies in helping researchers and developers to quickly assess the capability of NLP tools for processing patient-generated text.

Chapter 7: CONCLUSION

In this last chapter, I summarize the key findings of the specific aims, the contributions of my work, and the limitations of my studies. I end this dissertation with a discussion of future work in the field of patient-centered informatics, more specifically in CMC, online community participation, and automatic processing of patient-generated text.

7.1 CHALLENGES OF GATHERING HEALTH INFORMATION FROM ONLINE HEALTH COMMUNITIES

Gathering health information from online health communities can occur in many different ways. In my first two specific aims, I investigate two main challenges of gathering health information from the perspective of community members.

7.1.1 Challenges from Members' Perspectives

Topic drift and its effect on gathering health information in online health communities: In Chapter 4, I addressed this first specific aim by detailing the challenges associated with topic drift and then by describing automatic methods to detect topic drift. I investigated how members of the WebMD communities react to topic drift, how topic drift unfolds in the communities, and who brings topics back to the original goals of threads.

Through qualitative analysis, I found a number of sources of severe local and severe topic drift, including, the desire to talk to the physician members of the communities and the lack of intuitive features supporting private messages as well as starting a thread. Most conversations had an element of gradual topic drift through semantic parallelism (Hobbs 1990) when a small portion of a topic gradually changes to other topics with similar and relevant properties. However, few topics, such as religion often became the new topic of conversation, resulting in a topic drift by means of chained explanation (Hobbs 1990). Members were generally frustrated when their own thread experienced topic drift, which is evidenced by my finding that the original posters provided the most effort to bring the topic back to the original goals of the thread.

The findings of this specific aim suggest that members of online health communities face challenges, such as topic drift, during conversations meant to gather information. Although

previous literature described such a phenomenon as a natural part of conversation (Dorval 1990), for sensitive topics, such as health, topic drift could pose life-altering implications.

Homophily of vocabulary usage: Beneficial effects of vocabulary similarity in online health communities: In Chapter 5, I investigated a number of factors correlated to encouraging active participations in their own threads and sustaining their active participations with the community as the second specific aim. Active participation plays a significant role in both information gathering and information dissemination. The majority of health information available in online health communities was provided by peer members (Gray et al. 1997; Sarasohn-Kahn 2008), thus without active participation of those peer members, health information could not be disseminated, would not be available, and the communities would cease to exist.

For my second research aim, I found that in online health communities, a number of factors, such as the similarity of vocabulary usage–homophily of vocabulary usage–that correlate with active participation in online health communities. I focused this aim within the notion of homophily of vocabulary usage, although other factors have been subjectively examined. Furthermore, I demonstrated automatic calculation of homophily of vocabulary usage by measuring cosine similarity of posts.

The empirical findings suggest that homophily of vocabulary usage between members' posts plays a crucial role in sustained engagement for online health communities. Members who received replies that used a similar vocabulary with their own posts tended to continue their conversations with respondents. Also, when the members are exposed to higher homophily of vocabulary usage in the early stage of joining the community, they are more likely to continue participating in the online health community over the long-term.

For my last specific aim, I investigated the main challenges of gathering health information from the perspective of researchers.

Challenges of automatically processing online community text: Reusing the collective knowledge of online health community members can be beneficial to researchers because it expands our knowledge on patients' treatment decisions, symptom management, and clinical management. This knowledge can be a basis for online community interventions or related future research. The first step and challenge for utilizing collective knowledge in online health

communities is making sense of the vast amounts of text in online communities. As the third specific aim of this thesis, I investigated a number of common failures when extracting health information using existing biomedical NLP tools, such as MetaMap. Then I provided an automatic failure detection system to examine the application of NLP to extract health information from online health community text in Chapter 6. Existing biomedical NLP tools have the potential to process online community text, but they are known to have problems with this type of text (Divita, Tse, and Roth 2004; C. A. Smith and Wicks 2008; Brennan and Aronson 2003; Pratt and Yetisgen-Yildiz 2003; Kang et al. 2012).

To better understand the problem, I manually reviewed MetaMap's commonly occurring failures, grouped the inaccurate mappings into failure types, and then identified causes of the failures through iterative rounds of open coding (Strauss and Corbin 1990). Using 9,657 posts from an online cancer community, I characterized three types of failures: (1) boundary failures, (2) missed terms failures, and (3) word sense ambiguity failures.

The analysis suggests that MetaMap is still problematic when processing online health community text. The current literature also suggests the continuous effort to improve NLP to process patient-generated text, such as text in online health communities. These continuously evolving changes warrant the need for applying a low-cost systematic assessment. The primarily accepted evaluation method in NLP, manual annotation, however, requires tremendous effort and time; thus, I provided an automated method to detect the most common failures.

To automatically detect these failure types, I then explored combinations of existing NLP techniques and dictionary-based matching for each failure cause. The automatic methods suggest that at least 49% of 383,572 MetaMap's mappings as problematic. Word sense ambiguity failure was the most widely occurring, comprising 82% of failures. Boundary failure was the second most frequent, amounting to 16% of failures, while missed term failures were the least common, making up less than 2% of failures. To evaluate the automated failure detection methods, I manually evaluated a randomly selected 1200 mappings. The manual evaluation showed that the automated failure detection achieved precision, recall, accuracy, and F1 score of 83%, 93%, 88%, and 88%, respectively.

7.2 SUMMARY OF CONTRIBUTIONS AND IMPLICATIONS

In this dissertation, my main contribution is to the field of health informatics, more specifically to the field of patient-centered informatics. I provide implications for enhancing the experience of gathering health information from online health communities, more specifically by improving CMC, online community participation, and automatic processing of patient generated text.

7.2.1 Contributions and Implications for Members

Contributions and implications with respect to topic drift: Although online health communities promise better utilization of personal health expertise and experiences, we first need to address knowledge gaps with respect to facilitation of richer interactions as well as the benefits and disadvantages of phenomena like topic drift. Topic drift is a widely studied phenomenon, however, I addressed gaps in previous literature by illustrating possible benefits of having off topic discussion (i.e., global topic drift threads).

Topic drift occurred in our online community data at two levels: global (i.e., community-wide) and local (i.e., thread-specific) levels. Previous studies on topic drift described global topic drift as a source of incoherence (S. Herring 1999) and conflict (Lambiase 2010). However, in WebMD online health communities, I found off-topic discussions (i.e., global-level topic drifts), were supported by both members (i.e., power-users) and moderators. The off-topic discussion supporters voiced the opinion that off-topic discussions can build rapport and bring members closer. However, they advocated to indicate the off-topic nature of threads in the title so that the threads would not interfere with other discussions that pertained to the global goals of the communities.

I found that well-managed off-topic discussions and global-level topic drifts, could positively affect online health communities. For instance, many off-topic discussions were lively and humorous, in contrast to the melancholy and serious tone of many on-topic discussions. However, I did not find evidence that regular-users also support off-topic discussions. I suspect that only experienced members (e.g., those who have a high level of active participation or defined community roles) know about the culture of having off-topic discussions because I came across posts that asked about the meaning of *OT* in the title. Given that many community members asked about the meaning of *OT*, I suggest that designers and managers of online health communities consider more structured ways to have off-topic discussions even for the new members. An intuitive structure for having off-topic discussions could build rapport and lighten

the general mood of the community as well as contribute to sustaining active participation, which has been shown to be a prominent challenge for online communities (Millen and Patterson 2002; Nonnecke and Preece 2000).

Although power-users and moderators supported off-topic discussions, members in general reacted negatively towards severe local topic drifts. I identified two types of local topic drift (i.e., thread level): gradual topic drift and severe topic drift. Gradual topic drift, in which only a fraction of topics changed through a semantic parallel, occurred most frequently. This change is common and expected in any conversation (Dorval 1990), including CMC (Lambiase 2010). However, members reacted negatively towards severe topic drift—in which previous topics were completely discarded and replaced with different topics. When severe topic drift occurred, original posters showed frustration, and few community members even tried to convert the topic back to the original one. Although complete avoidance of severe topic drift could be difficult, two causes of severe topic drift could be prevented with improved design. For example, in the WebMD, I discovered that fewer members changed topics completely, because they did not know how to start their own thread or send a private message. An intuitive interface supporting such interactions might reduce severe topic drift.

I provided an automatic method to detect both gradual and severe topic changes at the thread level. This method could be used to identify severe topic drift. The automated technique can inform community moderators of severe topic drift and allow a structured system to provide the adequate information and support that original posters seek. Another use of the automated technique would be to alert community members when their posts are off topic. Raising self-awareness could help to reduce severe topic drift. As for the community, a similar relevance measurement technique with respect to the goals of the community could be a basis for filtering spam or abusive content, while keeping relevant content available to the community. The automated technique performed reasonably well; however, an extended evaluation using a large dataset could deepen the understanding of the technique in tracking topic drift at the thread level.

Moreover, the automated method could be the basis for automatically offering previously written posts on a similar topic (Ackerman and Malone 1990). Automatically providing relevant information could reduce the responsibilities of moderators while serving members who experience severe topic drift to the threads they started. Using this automated method, I also provided empirical evidence that original posters provide more effort for counteracting topic drift

than other members including MDs and moderators. This trend could indicate that original posters have a higher stake in keeping the thread on topic.

Based on these insights, online community creators, and online community members can better manage topic drifts that are inevitable in conversation. The significance of this study goes beyond CMC-based online health communities. The insights could generalize to CMC based discussions on other serious topics. Findings from this aim could improve member experience of CMC.

Contributions and implication with respect to participation in online health communities: Arguably successful health information gathering depends on other members' active participation and the community's ability to sustain active participation. However, sustaining active participation in online communities is a well-known prominent challenge (Millen and Patterson 2002; Joyce and Kraut 2006; Y. Wang, Kraut, and Levine 2012; Arguello et al. 2006; Nonnecke and Preece 2000). I addressed this issue by showing the importance of homophily expressed through shared vocabulary.

Utilizing the homophily principle, I first created prediction models that estimate the likelihood that original posters will reengage with respondents. Although many factors can contribute to members disengaging from conversation, my logistic regressions showed that vocabulary similarity predicts future participation with the respondents.

Similar prediction models could be one solution to the challenge of sustaining active participation. For instance, my prediction model could allow moderators to identify members who are most likely to disengage. This added knowledge could enable moderators to provide replies that encourage reengagement.

Moreover, I discovered that receiving replies written in a similar vocabulary in the early stages of joining the community predicts long-term active participation within the community. One solution for sustaining newcomers' participation in the community is to encourage community members to abide by a set of guidelines, which reflect ideal community member interactions with newcomers.

Furthermore, measuring vocabulary similarity can be a basis for automatically alerting members when they deviate from posting replies that encourage reengagement. In contrast with targeting specific members to encourage reengagement, the environment of the community as a

whole could be improved by filtering spam or abusive content through comparing terms with the common vocabulary of the community.

Based on these insights, moderators, online community creators, and online community participants could tailor replies to encourage sustained, active participation by members. Findings from this study can improve member experience in difficult situations when online health communities provide essential support.

7.2.2 Contributions and Implications for Researchers of Online Community Text

Contributions and implication with respect to using biomedical NLP tools to process online community text: Although online health communities have invaluable health information shared by members of the community, the technology to process vast amount of patient-generated text is fraught with challenges.

Although some of the failures I described in this aim are known, problematic failures of MetaMap (Divita, Tse, and Roth 2004; C. A. Smith and Wicks 2008; Brennan and Aronson 2003; Pratt and Yetisgen-Yildiz 2003; Kang et al. 2012), I extended prior work by identifying the causes for each failure type. Also, by leveraging the insight, I developed automated techniques that identified these previously highlighted failures without needing to create manually annotated datasets. In demonstrating the feasibility of our automated failure detection tool, I delineated the use of easily accessible NLP techniques and dictionaries. These techniques can independently examine each failure type. I provided a detailed demonstration of our failure detection tool to allow researchers to select the parts of the approach that meet the focus of their NLP tool assessment. Additionally, the detection approach can be modified and be used to rectify failures in NLP tools.

Although I focused on online health community text, the insights inform efforts to apply NLP tools to process various types of patient-generated text, including blogs or online journals, which share similar narrative writing styles and colloquial language. Moreover, Facebook and email provide conversational interactions similar to the interaction in online health communities. Tweets about emergency responses (Starbird and Palen 2011), public health trends (Dredze 2012), or clinical notes from electronic medical records (EMR) could contain a host of abbreviations that NLP tools could incorrectly map. Thus, the failure detection techniques could

be applied in these other contexts to assess the capability of processing different types of patient-generated text.

I focused my research on MetaMap, however, findings from my research can apply to other NLP tools in a similar manner. For example, different NLP tools could excel in different areas, and our automated failure detection can cost-effectively highlight problematic areas. Similarly, our techniques for detecting failures could be used to strengthen the performance of other NLP tools to process patient-generated text and more traditional types of text. For instance, applying POS information when mapping a term could increase the accuracy of the mappings from a variety of texts.

Another use of the detection system is expanding vocabulary source similar to the algorithm used to identify community nomenclature. Researchers could use a similar algorithm to collect various synonyms that are used in different domains and frequently update the vocabulary sources, such as UMLS. Furthermore, researchers could use the splitting-a-phrase detection technique to not only prevent boundary failures, but to collect new medical jargon (e.g., 'chemo curls') and identify important concepts missing in the UMLS (e.g., 'double mastectomy').

The findings have implications across a range of settings from processing EMR data and popular social media platforms, such as Facebook, to maintaining vocabulary sources like UMLS. Applying existing NLP tools to new situations and types of text requires an in-depth understanding of the benefits and limitations of the tool. The insight from this specific aim of the thesis shows a number of automated methods to detect biomedical NLP failures at a low cost.

7.3 LIMITATIONS AND FUTURE WORK

In this section, I describe limitations of each specific aim. Each specific aim has its own limitations and corresponding future work.

In the study investigating topic drift (Chapter 4), I focused on threads that members self-identified as possessing topic drift, which might not be representative reactions of all topic drifts. This method could bias towards sampling a single type of reactions, since members are behaving one way towards topic drift (i.e., explicitly expressing topic drift). However, I observed both positive and negative reactions towards topic drift.

Another limitation is that the large sample size could have inflated the significance levels. However, both qualitative and quantitative analyses showed consistent results in a diverse group of online health communities.

Lastly, I am uncertain of the specific goals of WebMD moderators and MDs. I made an assumption of their goals—creating an engaging and respectful community culture—based on previous literature (Wenger, McDermott, and Snyder 2002). Future work using mixed methods, such as surveys and interviews, asking about the responsibility of the WebMD moderators and MDs, could lead to a deeper understanding of this finding.

Another future work includes examining sophisticated relevance measurements, such as knowledge-based (Leacock and Chodorow 1998) and corpus-based (Turney 2001) approaches. These sophisticated relevance measurements that consider semantic meaning or syntactic organizations of the words could result in a highly accurate model for detecting topic drift.

In the study investigating participation (Chapter 5), one limitation is that I used small sample size of 100 to investigate influence of factors other than homophily of vocabulary that can have influence over original posters' engagement. These factors were chosen from previous studies (Y. Wang, Kraut, and Levine 2012; Adamic et al. 2008; Agichtein et al. 2008; Arguello et al. 2006) and had higher occurrences in reengagement, but they are not significant. I suspect this result is due to the small sample size. Learning the strongest influencing factors on the original posters' engagement using a larger sample suggests an important future work.

Furthermore, the vocabulary similarity of the first reply might not be the sole factor of member engagement with respondents. For instance, other components of life can influence online health community participation. Original posters could have gained the needed information through other sources and did not check back with the community. A serious medical crisis could prevent original posters from checking back with the community too, for example. Still, the consistent statistical results from examining the relationship between vocabulary similarity and participation show that homophily of vocabulary provides an important marker for member reengagement.

The survival analysis examines members' early stages of joining the community and uses a threshold of the first three replying posts that members received based on previous literature on lurking (Nonnecke and Preece 2000). Understanding the transition point of members' participation can provide a deeper understanding of how to sustain active participation in online

communities. Also, my analysis does not answer whether inactive members remained lurkers or completely dropped from the community. An analysis of these two different types of inactive members could further extend our understanding.

Although overall the analyses showed consistent results in a diverse group of online health communities, I acknowledge that our large sample size could have inflated the significance levels and raises questions to the practical significance of our findings. Also, how much vocabulary similarity is needed for a meaningful increase in reengagement remains an open question. Further investigation, such as surveys or interviews, asking about members' satisfaction could further explore the significance of the findings.

In future work, I plan to investigate the correlation between actual users' perceived qualities of replies and participation to further examine the challenge of sustaining participation in online health communities. Understanding these relationships could provide a more complete view of how to sustain participation in online health communities.

In a study investigating biomedical NLP tools to process online health community text (Chapter 6), I found two limitations. The first limitation is that the NLP techniques (Toutanova et al. 2003; de Marneffe, MacCartney, and Manning 2006; A. S. Schwartz and Hearst 2003) that I used in the detection process. Although these techniques were shown to be effective when applied to their own domain of texts, they were not evaluated with patient-generated text. Moreover, despite high performance of these techniques, errors still persist even in their own domain of text. The second limitation is that the automated detection techniques are generally limited to the coverage of the UMLS and MetaMap's capability to map when accurate and full spellings were provided.

To strengthen the findings, I evaluated each detection method as well as the overall performance. However, the findings could be biased towards cancer community text, and could be further strengthened by generalizing our results in different platforms or patient groups in future work. It is also plausible that I have not encountered all failure types or causes for other patient-generated health data contexts.

7.4 CONCLUDING REMARK

One of the most important findings I learned is that many members of online communities are willing to go to the *extra mile* to help others in similar situations. Yet, many challenges are

hindering the experience of gathering health information from online health communities. Although these efforts leave a conversation record that is embedded with diverse personal health expertise and experiences, we still lack the capability to automatically utilize this invaluable information. As exemplified in this dissertation, we still have many technological barriers to overcome to improve community members' experiences. In conclusion, we should be enthusiastic about increasing use and development of social media platforms that can help share valuable information. To reach their full potential, we, as researchers, should continue to enhance the members' experience of gathering health information from online health communities and other social media platforms, provide seamless CMC, maintain appealing platforms for members to share health information, and illuminate methods to capture previous members' effort and knowledge to help those in need of information. We can then reuse online health community members' collective knowledge on patients' experiences and maximize the benefits of online health communities for patients and researchers.

BIBLIOGRAPHY

Ackerman, M. S., and T. W. Malone. 1990. “Answer Garden: A Tool for Growing Organizational Memory.” In *Proceedings of the ACM Conference on Office Information Systems*, 11:31–39. New York, New York, USA: ACM Press. doi:10.1145/91474.91485.

Adamic, Lada A., Jun Zhang, Eytan Bakshy, and Mark S. Ackerman. 2008. “Knowledge Sharing and Yahoo Answers: Everyone Knows Something.” In *Proceedings of the 17th International Conference on World Wide Web*, 665–674. ACM.

Agichtein, Eugene, Carlos Castillo, Debora Donato, Aristides Gionis, and Gilad Mishne. 2008. “Finding High-Quality Content in Social Media.” In *Proceedings of the International Conference on Web Search and Web Data Mining - WSDM ’08*, 183. New York, New York, USA: ACM Press. doi:10.1145/1341531.1341557.

Alpers, Georg W., Andrew J. Winzelberg, Catherine Classen, Heidi Roberts, Parvati Dev, Cheryl Koopman, and C. Barr Taylor. 2005. “Evaluation of Computerized Text Analysis in an Internet Breast Cancer Support Group.” *Computers in Human Behavior* 21 (2) (March): 361–376. doi:10.1016/j.chb.2004.02.008.

Arguello, Jaime, Brain Butler, Elisabeth Joyce, Robert Kraut, Kimberly S. Ling, Carolyn Rose, and Xiaoqing Wang. 2006. “Talk to Me: Foundations for Successful Individual-Group Interactions in Online Communities.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 959–968. ACM.

Aronson, Alan R, and François-Michel Lang. 2010. “An Overview of MetaMap: Historical Perspective and Recent Advances.” *Journal of the American Medical Informatics Association* 17 (3): 229–36. doi:10.1136/jamia.2009.002733.

Bales, Robert F. 1950. *Interaction Process Analysis.* Cambridge, Mass.

Bamford, S, E Dawson, S Forbes, J Clements, R Pettett, a Dogan, a Flanagan, et al. 2004. “The COSMIC (Catalogue of Somatic Mutations in Cancer) Database and Website.” *British Journal of Cancer* 91 (2) (July 19): 355–8. doi:10.1038/sj.bjc.6601894.

Barcellini, Flore, Françoise Détienne, Jean-Marie Burkhardt, and Warren Sack. 2005. “A Study of Online Discussions in an Open-Source Software Community.” In *Communities and Technologies 2005*, 301–320. Springer Netherlands.

Bartlett, Yvonne Kiera, and Neil S Coulson. 2011. “An Investigation into the Empowerment Effects of Using Online Support Groups and How This Affects Health Professional/patient Communication.” *Patient Education and Counseling* 83 (1) (April): 113–119. doi:10.1016/j.pec.2010.05.029.

Baumgartner Jr, William A., Zhiyong Lu, Helen L. Johnson, J. Gregory Caporaso, Jesse Paquette, Anna Lindemann, Elizabeth K. White, Olga Medvedeva, K. Bretonnel Cohen, and Lawrence

Hunter. 2008. “Concept Recognition for Extracting Protein Interaction Relations from Biomedical Text.” *Genome Biology* 9 (Suppl 2): S9. doi:10.1186/gb-2008-9-S2-S9.

Benne, Kenneth D., and Paul Sheats. 1948. “Functional Roles of Group Members.” *Journal of Social Issues* 4 (2): 41–49.

Berry, Donna L., William J. Ellis, Nancy Fugate Woods, Christina Schwien, Kristin H. Mullen, and Claire Yang. 2003. “Treatment Decision-Making by Men with Localized Prostate Cancer: The Influence of Personal Factors.” *Urologic Oncology: Seminars and Original Investigations* 21 (2): 93–100. doi:10.1016/S1078-1439(02)00209-0.

Braithwaite, Dawn O, Vincent R Waldron, and Jerry Finn. 1999. “Communication of Social Support in Computer Mediated Groups for People With Disabilities.” *Health Communication* 11 (2): 123–151. doi:10.1207/s15327027hc1102.

Brennan, Patricia Flatley, and Alan R. Aronson. 2003. “Towards Linking Patients and Clinical Information: Detecting UMLS Concepts in E-Mail.” *Journal of Biomedical Informatics* 36 (4-5) (August): 334–341. doi:10.1016/j.jbi.2003.09.017.

Brooks, Michael, Cecilia R. Aragon, Katie Kuksenok, Megan K. Torkildson, Daniel Perry, John J. Robinson, Taylor J. Scott, Ona Anicello, Ariana Zukowski, and Paul Harris. 2013. “Statistical Affect Detection in Collaborative Chat.” In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work - CSCW ’13*, 317. New York, New York, USA: ACM Press. doi:10.1145/2441776.2441813.

Bureau, US census. 1990. “Frequently Occurring Names in the U.S.” http://www.webcitation.org/6XwJ9Enlw.

Chapman, W W, W Bridewell, P Hanbury, G F Cooper, and B G Buchanan. 2001. “A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries.” *Journal of Biomedical Informatics* 34 (5) (October): 301–10. doi:10.1006/jbin.2001.1029.

Chartrand, TL, and JA Bargh. 1999. “The Chameleon Effect: The Perception–behavior Link and Social Interaction.” *Journal of Personality and Social Psychology* 76 (6): 893.

Chee, Brant W, Richard Berlin, and Bruce Schatz. 2011. “Predicting Adverse Drug Events from Personal Health Messages.” In *AMIA Annu Symp Proc*, 217–226. American Medical Informatics Association. doi:22195073.

Chen, Lifeng, and Carol Friedman. 2004. “Extracting Phenotypic Information from the Literature via Natural Language Processing.” *Studies in Health Technology and Informatics* 107 (Pt 2) (January): 758–62.

Chen, Yan, Yehoshua Perl, James Geller, and James J. Cimino. 2007. “Analysis of a Study of the Users, Uses, and Future Agenda of the UMLS.” *Journal of the American Medical Informatics Association* 14 (2): 221–231. doi:10.1197/jamia.M2202.Introduction.

Cook, Samantha, Corrie Conrad, Ashley L Fowlkes, and Matthew H Mohebbi. 2011. “Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1) Pandemic.” *PloS One* 6 (8) (January): e23610. doi:10.1371/journal.pone.0023610.

Cutrona, Carolyn E., and Julie A. Suhr. 1994. *Social Support Communication in the Context of Marriage: An Analysis of Couples’ Supportive Interactions.*

Danescu-Niculescu-Mizil, Cristian, Robert West, Dan Jurafsky, and Christopher Potts. 2013. “No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities.” *Proceedings of the 22nd International Conference on World Wide Web*: 307–317.

De Choudhury, Munmun, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. “Predicting Depression via Social Media.” In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

De Marneffe, Marie-Catherine, Bill MacCartney, and Christopher D. Manning. 2006. “Generating Typed Dependency Parses from Phrase Structure Parses.” In *Proceedings of LREC*, 449–454.

Divita, Guy, Tse Tse, and Laura Roth. 2004. “Failure Analysis of MetaMap Transfer (MMTx).” In *Medinfo 2004*, 107:763–7.

Dorval, Bruce. 1990. *Conversational Organization and Its Development*. Norwood, NJ: Ablex Publishing Corporation.

Dredze, Mark. 2012. “How Social Media Will Change Public Health.” *Intelligent Systems, IEEE* 27 (4): 81–84.

Elhadad, Noémie, Shaodian Zhang, Patricia Driscoll, and Samuel Brody. 2014. “Characterizing the Sublanguage of Online Breast Cancer Forums for Medications, Symptoms, and Emotions.” In *Proc AMIA Annual Fall Symposium*.

Fairclough, Norman. 1992. “Discourse and Text: Linguistic and Intertextual Analysis within Discourse Analysis.” *Discourse & Society* 3 (2): 193–217.

Farnham, Shelly, Lili Cheng, Linda Stone, Melora Zaner-Godsey, Christopher Hibbeln, Karen Syrjala, Ann Marie Clark, and Janet Abrams. 2002. “HutchWorld: Clinical Study of Computer-Mediated Social Support for Cancer Patients and Their Caregivers.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 375–382. ACM.

Fox, Susannah. 2011. “Peer-to-Peer Healthcare: Many People - Espeically Those Living with Chronic or Rare Diseases - Use Online Connections to Supplement Professional Medical Advice.” *Pew InternePeer-to-Peer Healthcare: Many People - Espeically Those Living with Chronic or Rare Diseases - Use Online Connections to Supplement Professional Medical Advicet & American Life Project*. http://www.webcitation.org/6Y5JjmduI.

Fox, Susannah, and M Duggan. 2013. “Health Online 2013: 35% of U.S. Adults Have Gone Online to Figure out a Medical Condition; of These, Half Followed up with a Visit to a Medical Professional.” *Health*: 1–55.

Fox, Susannah, and Lee Rainie. 2002. “Vital Decisions: How Internet Users Decide What Information to Trust When They or Their Loved Ones Are Sick.” *Pew Internet & American Life Project*. January.

———. 2014. “The Web at 25 in the U.S. : The Overall Verdict: The Internet Has Been a plus for Society and an Especially Good Thing for Individual Users.” *Pew Internet & American Life Project*. http://www.webcitation.org/6Y5L8D2oE.

Friedman, C., G. Hripcsak, W. DuMouchel, S. B. Johnson, and P. D. Clayton. 1995. “Natural Language Processing in an Operational Clinical Information System.” *Natural Language Engineering* 1 (01): 83–108.

Friedman, Carol. 1997. “Towards a Comprehensive Medical Language Processing System: Methods and Issues.” In *Proceedings of the AMIA Annual Fall Symposium*, 595–599. American Medical Informatics Association.

Frijda, Nico H. 1993. “Moods, Emotion Episodes, and Emotions.” In *Handbook of Emotions*, 381–403.

Galegher, J., L. Sproull, and S. Kiesler. 1998. “Legitimacy, Authority, and Community in Electronic Support Groups.” *Written Communication* 15 (4) (October 1): 493–530. doi:10.1177/0741088398015004003.

Garla, Vijay, Vincent Lo Re, Zachariah Dorey-Stein, Farah Kidwai, Matthew Scotch, Julie Womack, Amy Justice, and Cynthia Brandt. 2011. “The Yale cTAKES Extensions for Document Classification: Architecture and Application.” *Journal of the American Medical Informatics Association : JAMIA* 18 (5): 614–20. doi:10.1136/amiajnl-2011-000093.

Gonzales, a. L., J. T. Hancock, and J. W. Pennebaker. 2009. “Language Style Matching as a Predictor of Social Dynamics in Small Groups.” *Communication Research* 37 (1) (November 4): 3–19. doi:10.1177/0093650209351468.

Granitz, N. a., S. K. Koernig, and K. R. Harich. 2008. “Now It’s Personal: Antecedents and Outcomes of Rapport Between Business Faculty and Their Students.” *Journal of Marketing Education* 31: 52–65. doi:10.1177/0273475308326408.

Gray, Ross, Margaret Fitch, Christine Davis, and Catherine Phillips. 1997. “A QUALITATIVE STUDY OF BREAST CANCER SELF-HELP GROUPS.” *Psycho-Oncology* 6 (4): 279–289.

Griffiths, Kathleen M, Alison L Calear, and Michelle Banfield. 2009. “Systematic Review on Internet Support Groups (ISGs) and Depression (1): Do ISGs Reduce Depressive Symptoms?” *Journal of Medical Internet Research* 11 (3) (January): e40. doi:10.2196/jmir.1270.

Gustafson, D H, R P Hawkins, E W Boberg, E Bricker, S Pingree, and C L Chan. 1994. “The Use and Impact of a Computer-Based Support System for People Living with AIDS and HIV Infection.” In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 00:604–608. American Medical Informatics Association.

Hancock, Jeffrey T., Christopher Landrigan, and Courtney Silver. 2007. “Expressing Emotion in Text-Based Communication.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’07*, 929. New York, New York, USA: ACM Press. doi:10.1145/1240624.1240764.

Hancock, JT, Kailyn Gee, Kevin Ciaccio, and JMH Lin. 2008. “I’m Sad You're Sad: Emotional Contagion in CMC.” In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, 295–298. ACM.

Hartzler, Andrea, and Wanda Pratt. 2011. “Managing the Personal Side of Health: How Patient Expertise Differs from the Expertise of Clinicians.” *Journal of Medical Internet Research* 13 (3) (January): e62. doi:10.2196/jmir.1728.

Helgeson, Vicki S., and Sheldon Cohen. 1996. “Social Support and Adjustment to Cancer: Reconciling Descriptive, Correlational, and Intervention Research.” *Health Psychology* 15 (2): 135–148. doi:10.1037/0278-6133.15.2.135.

Henderson, A. 1995. “Abused Women and Peer-Provided Social Supp: The Nature and Dynamics of Reciprocity in a Crisis Setting.” *Issues in Mental Health Nursing* 16 (2): 117–128.

Henley, Nancy M., and Cheris Kramarae. 1994. “Gender, Power, and Miscommunication.” In *C. Roman, S. Juhasz, & C. Miller (Eds.), The Towmen and Language Debate*, 383–406. New Brunswick: Rutgers University Press.

Herring, Susan. 1999. “Interactional Coherence in CMC.” *Journal of Computer-Mediated Communication* 4 (4) (June 23). doi:10.1111/j.1083-6101.1999.tb00106.x.

Herring, Susan C. 1993. “Gender and Democracy in Computer-Mediated Communication. [On-Line].” *Computer-Mediated Communication, Special Issue of the Electronic Journal of Communication* 3 (2).

———. 2003. “Dynamic Topic Analysis of Synchronous Chat.” In *New Research for New Media: Innovative Research Methodologies Symposium Working Papers and Readings*.

Herring, Susan C., and Carole Nix. 1997. “Is ‘serious Chat’an Oxymoron? Academic vs. Social Uses of Internet Relay Chat.” In *American Association of Applied Linguistics*. Orlando, FL.

Hobbs, Jerry R. 1990. “Topic Drift.” *Conversational Organization and Its Development* 38: 3–22.

Huh, J, and MS Ackerman. 2012. “Collaborative Help in Chronic Disease Management: Supporting Individualized Problems.” In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 853–862. ACM.

Huh, Jina, M Yetisgen-Yildiz, and Wanda Pratt. 2013. “Text Classification for Assisting Moderators in Online Health Communities.” *Journal of Biomedical Informatics* 46 (6): 998–1005.

Humphreys, BL, DAB Lindberg, and HM Schoolman. 1998. “The Unified Medical Language System: An Informatics Research Collaboration.” *Journal of the American Medical Informatics Association* 5 (1).

Ireland, Molly E, Richard B Slatcher, Paul W Eastwick, Lauren E Scissors, Eli J Finkel, and James W Pennebaker. 2011. “Language Style Matching Predicts Relationship Initiation and Stability.” *Psychological Science* 22 (1) (January): 39–44. doi:10.1177/0956797610392928.

Jacobson, David E. 1986. “Types and Timing of Social Support.” *Journal of Health and Social Behavior* 27 (3): 250–264. doi:10.2307/2136745.

Jansen, Eric, and Vincent James. 2002. *NetLingo: The Internet Dictionary*.

Jimeno-Yepes, Antonio, Rafael Berlanga-Llavori, and Dietrich Rebholz-Schuhmann. 2010. “Ontology Refinement for Improved Information Retrieval.” *Information Processing & Management* 46 (4) (July): 426–435. doi:10.1016/j.ipm.2009.05.008.

Jonquet, Clement, NH Shah, and MA Musen. 2009. “The Open Biomedical Annotator.” *Summits on Translational Bioinformatics*: 56–60.

Joyce, Elisabeth, and Robert E. Kraut. 2006. “Predicting Continued Participation in Newsgroups.” *Journal of Computer-Mediated Communication* 11 (3) (April): 723–747. doi:10.1111/j.1083-6101.2006.00033.x.

Kang, Ning, Bharat Singh, Zubair Afzal, Erik M van Mulligen, and Jan a Kors. 2012. “Using Rule-Based Natural Language Processing to Improve Disease Normalization in Biomedical Text.” *JAMIA* (October 6): 1–6. doi:10.1136/amiajnl-2012-001173.

Keselman, Alla, Catherine Arnott Smith, Guy Divita, Hyeoneui Kim, Allen C Browne, Gondy Leroy, and Qing Zeng-Treitler. 2008. “Consumer Health Concepts That Do Not Map to the UMLS : Where Do They Fit ?” *JAMIA* 15 (4): 496–505. doi:10.1197/jamia.M2599.Introduction.

Kollock, Peter, and Marc Smith. 1996. “Managing the Virtual Commons: Cooperation and Conflict in Computer Communities.” In *S.C. Herring (Ed.), Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives*, 109–128. Philadelphia: John Benjamins.

Kramer, Adam DI, Jamie E. Guillory, and Jeffrey T. Hancock. 2014. “Editorial Expression of Concern: Experimental Evidence of Massivescale Emotional Contagion through Social Networks.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (24): 8788–8790. doi:10.1073/pnas.1412469111.

Krause, N, a R Herzog, and E Baker. 1992. “Providing Support to Others and Weil-Being in Later Life.” *Journal of Gerontology* 47 (5) (September): P300–311.

Kumaran, Giridhar, and James Allan. 2004. “Text Classification and Named Entities for New Event Detection.” In *Proceedings of the 27th Annual International Conference on Research and Development in Information Retrieval - SIGIR ’04*, 297–304. New York, New York, USA: ACM Press. doi:10.1145/1008992.1009044.

Lambiase, Jacqueline J. 2010. “Hanging by a Thread: Topic Development and Death in an Online Discussion of Breaking News.” *Language@ Internet* 7.

Lazarus, Richard Stanley, and Susan Folkman. 1984. *Stress, Appraisal, and Coping.* Springer Publishing Company.

Leacock, Claudia, and Martin Chodorow. 1998. “Combining Local Context and WordNet Similarity for Word Sense Identification.” In *WordNet: An Electronic Lexical Database*, 265–283. The MIT Press.

Maclean, Diana, Sonal Gupta, Anna Lembke, Christopher Manning, and Jeffrey Heer. 2015. “Forum77 : An Analysis of an Online Health Forum Dedicated to Addiction Recovery.” In *Proceedings of the Companion Publication of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*.

MacLean, Diana Lynn, and Jeffrey Heer. 2013. “Identifying Medical Terms in Patient-Authored Text: A Crowdsourcing-Based Approach.” *Journal of the American Medical Informatics Association : JAMIA* (May 5): 1–10. doi:10.1136/amiajnl-2012-001110.

Maloney-Krichmar, Diane, and Jenny Preece. 2005. “A Multilevel Analysis of Sociability, Usability, and Community Dynamics in an Online Health Community.” *ACM Transactions on Computer-Human Interaction* 12 (2) (June 1): 201–232. doi:10.1145/1067860.1067864.

McCray, Alexa T, Nicholas C Ide, Russell R Loane, and Tony Tse. 2004. *Strategies for Supporting Consumer Health Information Seeking. MedInfo*. Vol. Pt 2.

McCray, Alexa T., and Nicholas C. Ide. 2000. “Design and Implementation of a National Clinical Trials Registry.” *Journal of the American Medical Informatics Association* 7 (3): 313–323. doi:10.1136/jamia.2000.0070313.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. “Birds of a Feather: Homophily in Social Networks.” *Annual Review of Sociology* 27: 415–444.

“MetaMap Updates.” http://www.webcitation.org/6T6e318kp.

Millen, DR, and JF Patterson. 2002. “Stimulating Social Engagement in a Community Network.” In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, 306–313. ACM.

Mo, Phoenix K H, and Neil S Coulson. 2012. “Developing a Model for Online Support Group Use, Empowering Processes and Psychosocial Outcomes for Individuals Living with HIV/AIDS.” *Psychology & Health* 27 (4) (January): 445–459. doi:10.1080/08870446.2011.592981.

Musen, Mark A, Natalya F Noy, Nigam H Shah, Patricia L Whetzel, Christopher G Chute, Margaret-Anne Story, and Barry Smith. 2012. "The National Center for Biomedical Ontology." *Journal of the American Medical Informatics Association* 19 (2): 190–5. doi:10.1136/amiajnl-2011-000523.

Nash, Carlos M. 2005. "Cohesion and Reference in English Chatroom Discourse." In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 00:1–10. IEEE.

Nonnecke, B, and J Preece. 2000. "Lurker Demographics: Counting the Silent." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2:73–80. ACM.

Paterson, Barbara L, Sally Thorne, and Marilyn Dewis. 1998. "Adapting to and Managing Diabetes." *The Journal of Nursing Scholarship* 30 (1): 57–62. doi:10.1111/j.1547-5069.2001.00021.x.

Paterson, Barbara, and Sally Thorne. 2000. "Developmental Evolution of Expertise in Diabetes Self-Management." *Clinical Nursing Research* 9 (4): 402–419.

Petersen, Alan. 2006. "The Best Experts: The Narratives of Those Who Have a Genetic Condition." *Social Science and Medicine* 63 (1): 32–42. doi:10.1016/j.socscimed.2005.11.068.

Pratt, Wanda, and Meliha Yetisgen-Yildiz. 2003. "A Study of Biomedical Concept Identification: MetaMap vs. People." In *AMIA Annual Symposium Proceedings*, 529–33. American Medical Informatics Association.

Preece, J, and K Ghozati. 2001. "Experiencing Empathy Online." In *The Internet and Health Communication: Experience and Expectations*, 233–256.

Preece, Jenny. 1999. "Empathic Communities: Balancing Emotional and Factual Communication." *Interacting with Computers* 12 (1): 63–77. doi:10.1016/S0953-5438(98)00056-3.

———. 2000. *Online Communities: Designing Usability and Supporting Socialbilty*. John Wiley & Sons, Inc.

"Query Suggestion Service." http://www.webcitation.org/6TIeuLgUB.

Riessman, Frank. 1965. *The "Helper" Therapy Principle.* Social Work.

Roberts, Teresa L. 1998. "Are Newsgroups Virtual Communities?" In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '98*, 360–367. ACM Press/Addison-Wesley Publishing Co. doi:10.1145/274644.274694.

Rodgers, Shelly, and Qimei Chen. 2005. "Internet Community Group Participation: Psychosocial Benefits for Women with Breast Cancer." *Journal of Computer-Mediated Communication* 10 (4) (June 23). doi:10.1111/j.1083-6101.2005.tb00268.x.

Ruland, Cornelia M. 1999. “Decision Support for Patient Preference-Based Care Planning Effects on Nursing Care and Patient Outcomes.” *Journal of the American Medical Informatics Association* 6 (4): 304–312.

Sarasohn-Kahn, Jane. 2008. “The Wisdom of Patients: Health Care Meets Online Social Media.” *California Healthcare Foundation*.

Savova, Guergana K, James J Masanz, Philip V Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C Kipper-Schuler, and Christopher G Chute. 2010. “Mayo Clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, Component Evaluation and Applications.” *Journal of the American Medical Informatics Association : JAMIA* 17 (5): 507–13. doi:10.1136/jamia.2009.001560.

Schwartz, Ariel S., and Marti A. Hearst. 2003. “A Simple Algorithm for Identifying Abbreviation Definitions in Biomedical Text.” In , 451–62.

Schwartz, Carolyn E, and Rabbi Meir Sendor. 1999. “Helping Others Helps Oneself : Response Shift E Ects in Peer Support.” *Social Science & Medicine* 48 (11): 1563–1575.

Selfe, Cynthia L, and Paul R. Meyer. 1991. “Testing Claims for on-Line Conferences.” *Written Communication* 8 (2): 163 – 192.

Setoyama, Yoko, Yoshihiko Yamazaki, and Kazuhiro Namayama. 2011. “Benefits of Peer Support in Online Japanese Breast Cancer Communities: Differences between Lurkers and Posters.” *Journal of Medical Internet Research* 13 (4) (January): e122. doi:10.2196/jmir.1696.

Shaw, B R, F McTavish, R Hawkins, D H Gustafson, and S Pingree. 2000. “Experiences of Women with Breast Cancer: Exchanging Social Support over the CHESS Computer Network.” *Journal of Health Communication* 5 (2): 135–159. doi:10.1080/108107300406866.

Shaw, BR, Robert Hawkins, and F McTavish. 2006. “Effects of Insightful Disclosure within Computer Mediated Support Groups on Women with Breast Cancer.” *Health Communication* 19 (2): 133–142. doi:10.1207/s15327027hc1902.

Shaw, Bret R., Jeong Yeob Han, Timothy Baker, Jeffre Witherly, Robert P. Hawkins, Fiona McTavish, and David H. Gustafson. 2007. “How Women with Breast Cancer Learn Using Interactive Cancer Communication Systems.” *Health Education Research* 22 (1): 108–119. doi:10.1093/her/cyl051.

Singhal, Amit. 2001. “Modern Information Retrieval: A Brief Overview.” *IEEE Data Eng. Bull.*: 1–9.

Smith, Catherine Arnott, P Zoe Stavri, and Wendy Webber Chapman. 2002. “In Their Own Words ? A Terminological Analysis of E-Mail to a Cancer Information Service.” In *AMIA Annu Symp Proc*, 697–701. American Medical Informatics Association.

Smith, Catherine Arnott, and Paul J Wicks. 2008. “PatientsLikeMe: Consumer Health Vocabulary as a Folksonomy.” In *AMIA Annual Symposium Proceedings*, 682–6. American Medical Informatics Association.

Smith, L, T Rindflesch, and W J Wilbur. 2004. “MedPost: A Part-of-Speech Tagger for bioMedical Text.” *Bioinformatics* 20 (14): 2320–2321.

Sproull, L, and S Kiesler. 1986. “Reducing Social Context Cues: Electronic Mail in Organizational Communication.” *Management Science* 32 (11): 1492–1512.

Starbird, Kate, and L Palen. 2011. “‘Voluntweeters’: Self-Organizing by Digital Volunteers in Times of Crisis.” In *Proceedings of the 2011 ACM CHI Conference on Human Factors in Computing Systems*, 1071 – 1080. ACM.

Strauss, Anselm Leonard, and Juliet M. Corbin. 1990. *Basics of Qualitative Research*. Newbury Park, CA: Sage Publications.

Toutanova, Kristina, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. “Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network.” In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, 173–180. Association for Computational Linguistics.

Turney, Peter D. 2001. “Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL.” In *Proceedings of the Twelfth European Confer- Ence on Machine Learning (ECML-2001)*.

Van Uden-Kraan, C F, C H C Drossaert, E Taal, E R Seydel, and M a F J van de Laar. 2009. “Participation in Online Patient Support Groups Endorses Patients’ Empowerment.” *Patient Education and Counseling* 74 (1) (January): 61–69. doi:10.1016/j.pec.2008.07.044.

Van Uden-Kraan, Cornelia F, Constance H C Drossaert, Erik Taal, Bret R Shaw, Erwin R Seydel, and Mart a F J van de Laar. 2008. “Empowering Processes and Outcomes of Participation in Online Support Groups for Patients with Breast Cancer, Arthritis, or Fibromyalgia.” *Qualitative Health Research* 18 (3) (March): 405–417. doi:10.1177/1049732307313429.

Wagner, David. 2013. “Google Flu Trends Wildly Overestimated This Year’s Flu Outbreak.” *The Atlantic Wire*. http://www.theatlanticwire.com/technology/2013/02/google-flu-trends-wildly-overestimated-years-flu-outbreak/62113/.

Walther, J. B. 1992. “Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective.” *Communication Research* 19 (1) (February 1): 52–90. doi:10.1177/009365092019001003.

Walther, Joseph B. 1994. “Anticipated Ongoing Interaction Versus Channel Effects on Relational Communication in Computer-Mediated Interaction.” *Human Communication Research* 20 (4) (June): 473–501. doi:10.1111/j.1468-2958.1994.tb00332.x.

Walther, Joseph B., Tracy Loh, and Laura Granka. 2005. “Let Me Count the Ways: The Interchange of Verbal and Nonverbal Cues in Computer-Mediated and Face-to-Face Affinity.”

Journal of Language and Social Psychology 24 (1) (March 1): 36–65. doi:10.1177/0261927X04273036.

Wang, Yi-chia, Robert Kraut, and John M Levine. 2012. "To Stay or Leave ? The Relationship of Emotional and Informational Support to Commitment in Online Health Support Groups." In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. ACM.

Wang, Zuoming, Joseph B Walther, Suzanne Pingree, and Robert P Hawkins. 2008. "Health Information, Credibility, Homophily, and Influence via the Internet: Web Sites versus Discussion Groups." *Health Communication* 23 (4): 358–368. doi:10.1080/10410230802229738.

Weinberg, Nancy, and John Schmale. 1996. "Online Help: Cancer Patients Participate in a Computer-Mediated Support Group." *Health & Social Work* 21 (1): 24 – 29.

Welbourne, Jennifer L., Anita L. Blanchard, and Marla D. Boughton. 2009. "Supportive Communication, Sense of Virtual Community and Health Outcomes in Online Infertility Groups." *Proceedings of the Fourth International Conference on Communities and Technologies - C&T '09*: 31. doi:10.1145/1556460.1556466.

Wen, M, and CP Rose. 2012. "Understanding Participant Behavior Trajectories in Online Health Support Groups Using Automatic Extraction Methods." In *Proceedings of the 17th ACM International Conference on Supporting Group Work*, 179–188. ACM.

Wenger, Etienne, Richard A McDermott, and William Snyder. 2002. *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Harvard Business Press.

Wicks, Paul, Timothy E Vaughan, Michael P Massagli, and James Heywood. 2011. "Accelerated Clinical Discovery Using Self-Reported Patient Data Collected Online and a Patient-Matching Algorithm." *Nature Biotechnology* 29 (5) (May): 411–4. doi:10.1038/nbt.1837.

Zeng, Q, S Kogan, N Ash, and R a Greenes. 2001. "Patient and Clinician Vocabulary: How Different Are They?" In *Studies in Health Technology and Informatics*, 84:399–403.

Zeng, Qing T, Jonathan Crowell, Robert M Plovnick, Eunjung Kim, Long Ngo, and Emily Dibble. 2006. "Assisting Consumer Health Information Retrieval with Query Recommendations." *JAMIA* 13 (1): 80–90. doi:10.1197/jamia.M1820.specific.

Zeng, Qing T., and Tony Tse. 2006. "Exploring and Developing Consumer Health Vocabularies." *JAMIA* 13 ((1)): 24–30. doi:10.1197/jamia.M1761.A.

VITA

Albert grew up in Korea then moved to Virginia to earn his bachelor's and master's degree in Computer Science at Virginia Tech. He moved back to Korea for work before deciding to pursue biomedical informatics in Seattle. While at University of Washington, Albert was involved in a NSF project that aims to match peer-mentor for online health community members with wonderful mentors Dr. Wanda Pratt, Dr. David McDonald, Dr. Andrea Hartzler, and Dr. Jina Huh. In his doctoral thesis titled "Enhancing Health Information-Gathering Experiences in Online Health Communities," Albert described his work to help understand how to best sustain and support online health communities and processing these important sources of patient knowledge and support. He cherishes his time with BHI, Imed, his cohort, and the city of Seattle. Albert is looking forward to his postdoctoral researcher position at the University of Utah in Biomedical Informatics Department.