The Impact of Virtually Crowdsourced Social Support on Individual Health: Analyzing Big Datasets for Underlying Causalities

Emergent Research Forum papers

Langtao Chen
J. Mack Robinson College of Business, Georgia State University, Atlanta, GA
lchen18@gsu.edu

Detmar W. Straub
J. Mack Robinson College of Business, Georgia State University, Atlanta, GA
dstraub@gsu.edu

Abstract

With the rise of online health communities, patients or consumers are using these communities to exchange support through enhanced social relations and interpersonal transactions. An emerging and interesting area of research is to comprehensively understand the interaction dynamics within online health communities. The current study examined the impact of virtually crowdsourced social support on individual health via analyses of big health data. Based on previous research, we propose a conceptual framework of social relations in the context of online health communities and test it through a quantitative field study. Specifically, text mining techniques are utilized to automate the content analysis of big health data. Contributions of this research will not only extend current understanding of social influence in online health communities, but also shed light on the general approach of coping with big datasets in research as well as the design and management of online health communities.

Keywords

Online Health Communities, Social Support, Big Data, Health Crowd, Automatic Content Analysis, Text Mining, Support Vector Machine (SVM), Latent Dirichlet Allocation (LDA), Unified Medical Language System (UMLS)

Introduction

Online health communities are social networks where people with common health interests can share experiences, request questions, seek or provide emotional support (Eysenbach et al. 2004). As an inseparable part of the personalized preventative medicine (Swan 2012), online health communities are changing the way patients treat and/or manage their health. Two major purposes of participants joining online health communities are to seek health information regarding self-management options and to receive emotional support by seeing that their peers care (Hajli et al. 2014). Advanced services such as posing questions to physicians, quantified self-tracking of health conditions, and clinical trials access can also be provided to consumers (Swan 2009). When individuals share their personal health records (PHR) with peers, they are “crowdsourcing” the collective wisdom of a huge community (Eysenbach 2008). This can significantly lower the cost of health care and alleviate burdens on the health care system.

Within the current era of “Big Data,” data generated from Web 2.0, social media, mobile devices, and ubiquitous sensors have been experiencing an exponential growth in terms of volume, velocity, and variety (Russom 2011). The rise of health social networks such as PatientsLikeMe, DailyStrength, and MedHelp provides unique opportunities for research focusing on healthcare decision support and patient empowerment (Miller 2012). With the abundant big data being generated by online health communities, scholars are able to obtain insights into highly detailed, contextualized, and rich contexts. However, there is a lack of research in IS field that empirically addresses this phenomenon and its underlying theoretical relationships via analyses of big health data, hence the RQ:

RQ: What is the impact of virtually crowdsourced social support on individual health?
Theoretical Background

Our theoretical model is presented below as Figure 1. We first discuss the key constructs in the model and then the underlying relationships between the variables that lead to our hypotheses.

Social Support

Studying online health communities involves modeling the social relations among individuals (Eysenbach 2008). Individuals join the online community to exchange social support through creating social relationships and engaging in interpersonal transactions (Hajli et al. 2014; Lakey and Cohen 2000). Although there are different classification schemes of social support, the most widely accepted typology in the literature on online health communities was developed by Cutrona and Suhr (1992). It includes: (1) informational support (providing suggestion or advice on coping with the stress), (2) emotional support (communicating love, care, or empathy), (3) esteem support (communicating respect and confidence in abilities), (4) tangible support (providing or offering to provide goods or services), and (5) network support (communicating belonging to a group or persons with similar interests and concerns). Informational support and emotional support have been found to be the two most frequent types of social support exchanged online (Braithwaite et al. 1999; Gooden and Winefield 2007; Huang and Chengalur-Smith 2014; Mo and Coulson 2008). In this research, we focus on informational support and emotional support exchanged within online health communities. Although the impact of social support on health outcomes is decidedly positive, the underlying mechanisms of such influence has not yet been fully addressed (Swan 2009).

Health Outcomes

To assess potential improvements in the quality of healthcare, various quality and patient safety (QPS) metrics have been devised by the healthcare industry. Since outcomes are the ultimate or acid test for effective healthcare (Lazar et al. 2013), health outcomes emerging from crowd social relations are of vital importance for good research on online health communities. As the definition of health and healthcare is extended to wellness maintenance and condition prevention rather than the single target of curing disease (Swan 2012), there are various kinds of supplementary outcome measures reported in extant literature (Eysenbach et al. 2004). Typically these outcome indicators are morbidity, recovery or restoration of function, readmission, length of stay (LOS), and quality of life (Donabedian 2005; Lazar et al. 2013). A comprehensive evaluation of the quality of health intervention should also include intermediate outcomes such as changes in the individual’s attitudes, knowledge, skills, confidence, and behaviors related to self-management of health (Fowles et al. 2009; Yoo and Bock 2014).

Health Crowd

With the large scale information sharing taken place between online health community peers, individuals and organizations can “crowdsource” the collective wisdom of a vast number of community members (Eysenbach 2008). This can significantly reduce cost of healthcare, improve quality of health and healthcare, and empower patients or consumers with more knowledge and confidence in self-management of health and stress. The availability of huge numbers of patients and consumers belonging to online health communities creates tremendous opportunities to efficiently conduct clinical trials (Darrow 2014), health service review (Adams 2011), and patient experience mining (Kaiser and Bodendorf 2012). The wisdom of health crowd is well founded on the big data being generated by self-managed healthcare sites as well as the voluminous online health community participants (Boulos et al. 2011). However, the interaction dynamics of health crowd in the online health community has not been fully explored.

Homophily

Homophily is the concept that people with similar attributes tend to associate with each other (McPherson et al. 2001). In an offline setting, homophily is found to affect the ultimate persuasiveness of communicators (McCroskey et al. 1975). In the context of online health communities, homophily is believed to enhance source credibility (Wang et al. 2008) as well as members’ perceived empathy from
other members (Nambisan 2011). Homophily can be measured by certain behavioral, sociodemographic, and intrapersonal characteristics such as beliefs, values, education, social status, etc. (McPherson et al. 2001; Rogers and Bhowmik 1970). In the online health community setting, members belonging to the same group share similar conditions and symptoms; thus we use group membership to measure homophily between the thread initiator and the responder.

**Research Model and Hypotheses**

**Research Model**

The research model for this study, as shown in Figure 1, incorporates the four major theoretical concepts just introduced including social support, health outcomes, crowd consensus, and homophily. Specifically, informational support and emotional support are used to predict health outcomes. In addition, homophily is treated as a moderator of the effects of social support on health outcomes while crowd consensus is posited to moderate the effect of informational support on health outcomes.

![Figure 1. Research Model](image)

**Research Hypotheses**

**Relationship between Social Support and Health Outcomes**

To share information about health conditions and treatments and to seek and provide social support are two important aspects of the online health community discourse. Being better informed about health self-management, patients or consumers sharing personal health data within online communities can clearly benefit from the process (Frost and Massagli 2008). There are various perspectives that explain this health-promoting function of social support. From the perspective offered by analogical behavioral processes, social support facilitates healthy behaviors such as exercising, eating right, quitting smoking, and actively engaging in medical regimens (Uchino 2006). Thus, it is hypothesized that support received in online health communities will influence health outcomes.

**H1a**: Informational support received will positively relate to health outcomes.

**H1b**: Emotional support received will positively influence health outcomes.
**Moderating Effect of Crowd Consensus**

Crowd consensus is defined as the extent to which opinions of the crowd converge. Crowd consensus has been found to be related to social influence by positively contributing to the perceived credibility of subjective experiential information shared in online health communities (Fan et al. 2013). Investigating the effect of crowd consensus on online health communities is a strong indicator of the quality of the social support environment (Bliese and Britt 2001). According to social influence theory (Kelman 1961; Kelman 2006), a high level of crowd consensus makes it easier for the social support recipient to assimilate and internalize information received. Hence, we expect that:

**H2**: Crowd consensus will positively moderate the effect of informational support on health outcomes.

**Moderating Effect of Homophily**

In online health communities, homophily refers to the degree to which two online health community members are similar in health conditions, symptoms, and treatments. With this similar background, the social support provider can more fully understand the situation of others, thus providing helpful advice and social support. From the perspective of recipient, the more homophilous online social support stimulus received, the more there should be information assimilation and greater confidence in self-management of disease and health conditions (Nambisan 2011). Thus, we expect that homophily between the social support provider and the recipient will positively moderate the effect of social support on health outcomes.

**H3a**: Homophily will positively moderate the effect of informational support on health outcomes.

**H3b**: Homophily will positively moderate the effect of emotional support on health outcomes.

**Control Variables**

To fully account for the heterogeneity among members, four control variables are tested including: (1) gender (Huang et al. 2014; Vaux 1985); (2) age (Vaux 1985); (3) tenure in the community (Huang and Chengalur-Smith 2014); and (4) degree of participation (Dholakia et al. 2004; Venkatesan et al. 2014).

**Research Method**

Given the explanatory nature of this study, we conducted a quantitative field study on an online health community to empirically test the proposed model.

**Data Collection**

Data were collected from DailyStrength, a large U.S.-based social networking website. The site contains online health communities that specifically focus on different health conditions. To get representative samples, we selected 9 forums covering various kinds of health conditions including: (1) general conditions (chronic pain, and obesity), (2) behavioral conditions (depression, anxiety, alcoholism, physical & emotional abuse, and insomnia), and (3) specific diseases (diabetes type 2, and HIV). An Internet crawler program was used to extract user-generated contents from these forums.

We collected all forum messages posted from August 2006 to November 2014. In total we obtained 238,617 online discussion threads containing 2,305,288 posts generated by 32,405 members. A thread is a group of messages discussing a question or topic initiated by a member, while a post or response is a message by another member replying to the initial message.

**Content Analysis**

Content analysis refers to “a research technique that makes replicable and valid inference from texts (or other meaningful matter) to the contexts of their use (Krippendorff 2004, p. 18 ).” Content analysis provides an unobtrusive way for researchers to gather information. Most previous research on online health communities employs a manual content analysis approach, whereby researchers read through the online messages and manually assign categories to them. Although some recent literature (e.g., Huang...
and Chengalur-Smith 2014; Huang et al. 2014; Wang et al. 2012) utilizes text mining algorithms to automate part of the content analysis work, automatic content analysis in past scholarship has still been severely limited in terms of both scope and depth.

To guide the content analysis of the online health messages, we used the Social Support Behavior Code (SSBC) developed by Cutrona and Suhr (1992) to code the social support for each message. This typology of social support is thought to be ideal for content analysis of online messages as it does not require the access to full range of nonverbal cues for the identification of social support (Braithwaite et al. 1999). Explanation with examples of social support provided by Mo and Coulson (2008) were also referenced. With the observation that many messages indicate more than one type of social support, we followed the same rule used by Loane and D’Alessandro (2013) to allow multiple social support types to be assigned to a single post.

To validate the metrics, a set of posts were manually coded for the types of social supported provided as well as the positive and negative attitudes expressed. Then this set was used as a training pool to train the automatic text classifiers which are based on support vector machine (SVM) model, a widely used text classification and opinion mining technique. Then the training classifiers were used to automatically code the rest of the online posts. The SVM-based automatic qualitative content analysis has been shown to provide results comparable to those concluded from traditional manual content analysis (Huang et al. 2010; Wang et al. 2012).

Following the method used by Huang and Chengalur-Smith (2014), we employed the count of the unified medical language system (UMLS) terms (Bodenreider 2004) presented in informational support messages to assess individual’s healthcare related knowledge. To assess the extent of crowd consensus, latent Dirichlet allocation (LDA) (Blei et al. 2003) extracted topics from online discussion messages. A similar latent semantic analysis (LSA) procedure has been used by Cao et al. (2011) to extract topic characteristics from online user-generated contents. LDA is an appropriate topic modeling tool for this research since its generative probabilistic topic modeling mechanism can deal with the limitation of LSA where words and documents are represented as points in Euclidean space (Steyvers and Griffiths 2007).

The unit of analysis for this study is thus at the individual level. Text classification of the social support, attitudes, the degree of healthcare knowledge, and the topic structure expressed at the message level are aggregated to individual level to calculate the indicators for all focal constructs in the proposed research model. The estimated sample size will most likely exceed 22,000 online health community members.

**Measurement Development**

Table 1 presents the operationalized definition of constructs being explored as well as their measurement items and analytical method used to extract them.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Item</th>
<th>Definition</th>
<th>Analytical Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informational Support</td>
<td>Quantity of Informational Support</td>
<td>The number of informational support replies received by the member.</td>
<td>SVM text classification; descriptive statistics</td>
</tr>
<tr>
<td>Emotional Support</td>
<td>Strength of Informational Support</td>
<td>Average number of word count per informational support reply received by the member.</td>
<td></td>
</tr>
<tr>
<td>Homophily</td>
<td>Same Group Membership</td>
<td>Average number of same groups that the member and her/his responders belong to.</td>
<td>Descriptive statistics</td>
</tr>
</tbody>
</table>

**Table 1:** Operationalized Definition of Constructs and Measurement Items.
**Table 1. Construct and Measurement**

<table>
<thead>
<tr>
<th>Health Outcomes</th>
<th>Opinion Consensus of the Crowd</th>
<th>Topic Consensus of the Crowd</th>
<th>Personal Health Status</th>
<th>Increase in Positive Attitude</th>
<th>Decrease in Negative Attitude</th>
<th>Increase in Health Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average opinion similarity among replies to threads initiated by the member.</td>
<td>Average number of topic similarity among replies to threads initiated by the member.</td>
<td>The health status self-reported by the member on her/his online profile; a possible status includes excellent, good, ok, bad, and horrible.</td>
<td>Percentage increase of the number of posts with positive attitudes posted by the member across the two periods of her/his community involvement.</td>
<td>Percentage decrease of the number of posts with negative attitudes posted by the member across the two periods of her/his community involvement.</td>
<td>Increase of average number of UMLS terms used in the member’s informational support posts across the two periods of her/his community involvement.</td>
</tr>
<tr>
<td></td>
<td>SVM text classification</td>
<td>LDA topic modeling</td>
<td>Descriptive statistics</td>
<td>SVM text classification</td>
<td>SVM text classification</td>
<td>UMLS term extraction</td>
</tr>
</tbody>
</table>

**Data Analysis**

Structural equation modeling (SEM) and hierarchical modeling will be used to test the hypotheses in the proposed research model.

**Conclusion and Contribution**

Drawing from the tenets of multiple theoretical bases including social support theory, social influence theory, and the theory of homophily, our proposed research model intends to build a comprehensive picture of the interaction dynamics within online health communities.

First, it will extend current understanding of social interaction in online health communities via analyses of big datasets of user-generated contents.

Second, the present study employs and validates text mining techniques for automatic content analysis of qualitative research data. The method used in this study can provide insights into how to deal with big textual data for future research on similar settings such as online communities and social media.

For practitioners, this research will provide valuable insights on the design and management of online health communities.

**REFERENCES**


