Online Health Support Communities: A Domain Specific Example of the Messiness of Human Centered Data Science

Abstract
Data Science for health care includes a good deal of structured and unstructured, formal and informal data. Often presented in, at best, forms requiring cleaning and reshaping. The design of systems to provide informal, patient centered social support is the focus of our intended contribution to this workshop. Social support can significantly improve health outcomes for individuals living with disease, and online forums have emerged as an important vehicle for providing social support. Here, we suggest a sociotechnical mechanism that enables these communities to be resilient while providing social support to a large number of people.

Our work will focus on making sense of the data around conversations between members of thirteen disease-specific forums in the WebMD online health community. Overall site activity is strongly correlated with the proportion of replies from densely connected cores of members who maintain strong relationships.

Author Keywords
Health informatics, health information, complex adaptive systems theory.

Josh Introne
Michigan State University
East Lansing, MI
JIntrone@msu.edu

Sean Goggins
University of Missouri
Columbia, MO
Goggins@missouri.edu

Bryan Semaan
Syracuse University
semaaan@syr.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s). CSCW ’16, February 27 - March 02, 2016, San Francisco, CA, USA.
INTRODUCTION
We describe a specific, health care related case of how to conduct data science in the realm of unstructured data employed for social computing research. A significant gap in our understanding of the human side of data science is grounding the discourse in two premises we find important in our work to date: First, data generated online, the source of most social computing research, is messy. It requires a lot of cleaning and reshaping. Often, these processes are lost to history. Second, data science, and human centered data science, are fundamentally domain specific endeavors. With the rest of this paper, we describe our initial experiences and our case.

Social support can play an important role in health outcomes for individuals with chronic diseases, impacting an individual’s mental and physiological well-being, as well as improving clinical outcomes for a variety of diseases [43]. Despite this, there is no defined role for social support in the public healthcare system [50].

Perhaps in response to this gap, a significant amount of social support activity is now transacted online. This has led researchers and service providers alike to investigate the design and function of these systems. In this vein, we investigate two research questions:

- What is the social organization of an online health support community?
- How is the task of providing social support allocated amongst community members in these systems?

Based on our answers to these two questions, we propose an answer to a third research question:

- By what mechanism might an online community sustain itself while providing social support to a large population?

Our investigation focuses upon a set of forums in the WebMD online health community, covering roughly five years of data scraped from web-forums hosted on the site. We begin our analysis with the 55 “featured” forums on the site, and then narrow our focus to the thirteen forums that name a specific disease or mental disorder, and also exhibit the structural property of a tightly connected core group of users.

We begin below by describing a selection of the established findings about social support, and more generally the structure and presumed sociotechnical mechanisms underlying online communities. This is followed by a detailed introduction to the WebMD discussion boards. Following this introduction, we present our analysis in four parts. First, we examine the social networks constructed from inferred relationships between participants, observing a distinct core / periphery / extra-periphery structure. Second, using our classification to distinguish between user groups, we examine who posts, who replies, and who talks with whom. Third, we apply content analysis to illustrate persistent differences in the kinds of support contained in responses from the core to other kinds of users. Finally, we synthesize these findings to propose a mechanism by which the communities we examine sustain themselves while delivering support to a large population of people with support needs.
BACKGROUND
Social support is a multi-dimensional construct covering the various kinds of support individuals provide one another. Various taxonomies of support have been proposed, but three dimensions appear frequently in literature on online support: community/companionship support (membership in a group with shared interests), emotional support (expressions of caring and comfort), and informational support (provision of knowledge or advice) [2,17,46]. These dimensions have been found to be correlated, yet distinct, and their expression depends both upon context and the nature of the relationships involved [17,49].

Giving and receiving social support has been demonstrated to have a range of profound impacts on the physical and mental well-being of individuals. Social support has been associated with improved clinical outcomes for a variety of conditions (e.g. [7,40,41]). Various reasons have been offered to explain this effect. Researchers have shown, for instance, that social support reduces the production of stress hormones, leading to improved immune response and reduction in stress-related disease [43]. Social support can assist with various forms of self-management, such as adherence to diet and medication regimens [20]. Informational support clearly has instrumental value in helping individuals manage their conditions [12], but its receipt and delivery can also be part of a coping process [44].

While the mechanisms and function of social support are current and active areas of inquiry, its value has long been recognized, leading to explicit calls for policies and clinical guidelines that mobilize social support [23,29]. Whereas no such institutional scaffolding has been forthcoming, online health forums may be viewed as an organic response to this need [43].

At a high level, an online health forum is a place where people with medical conditions can go to find social support that complements the services provided by professional health care providers [11,21,25,32]. However, this description obscures the vast diversity found within online support communities.

For example, Wang et al. [46] found that members in a breast cancer support community clustered into seven different groups according to the kinds of support sought and provided and, consistent with Wang et al. [48], found that these groupings were also predictive of member tenure and engagement. Bambi [2] found that members in a breast cancer support forum segregate into “takers” who receive support and “providers” who deliver it, and furthermore, that the types of support sought and provided vary between these groups. She also found evidence for a core-periphery social network organization. Such structures have been found to be present in a variety of social media [13,14], and it has been suggested that having a committed core of people is essential for sustaining online communities [37].

Few analyses specific to online support have sought to articulate the ways in which these different roles interact to perform the “work” of providing large-scale social support. Such mechanisms have, however, been articulated in other online communities. For instance, Crowston et al. [13] described how open-source software communities are organized according to a
core-periphery structure, with different roles for different users. Similar findings have been made about Wikipedia and other online communities [6,9,34].

How such intra-community stratification can give rise to sustainable communities is a nuanced topic. "Critical mass" may be necessary to sustain online communities [26], but the decay of Wikipedia [42] suggests there is more at work. Ren et. al [36] suggest a framework in which shared identity in an online community (e.g. breast-cancer survivors) might result in some members forming strong bonds, but this could ultimately have a negative effect as these members dominate conversations or reject newcomers.

In the following, we seek to extend such work to online support. Our approach is to examine the internal structure of several online support communities hosted by WebMD, followed by the kinds of "work" (in terms of the types of support) done by different classes of users as they interact. We will then use these findings to propose a mechanism that allows these communities to function as a relatively stable sociotechnical engine for large-scale online support.

**WEBMD**

WebMD is a popular online health resource (as of January 2015, Alexa\(^1\) ranked WebMD.com as the 120\(^{th}\) most popular site in the US, and second most popular health site, behind NIH.gov [51]). In addition to providing news, resources to assist with self-diagnosis and a nationwide directory of medical professionals, WebMD also hosts an array of topic-specific message forums. These forums are conventionally used as question & answer forums, in which top-level posts are questions and replies are answers, but a large amount of conversational interaction occurs within this context. 55 of these forums are currently “featured,” which means that WebMD experts participate in discussions, and that spam is deleted.

**Implications for Human Centered Data Science**

Social computing, HCI and data science intersect with each other in new and interesting ways. To date, purely computational discourse around statistics and modeling have been the focus of research and the center of what is widely regarded as 'data science'. In our case, we raise a host of data structuring, cleaning, processing and interpretation issues around a complex data science domain. We argue that our case directly addresses how the discourse for human centered data science might be framed, by surfacing the messiness of these practices for social computing scholars. The domain of health care helps, we believe, to ground discourse around this problem and will make a unique contribution ot the workshop.

**References**


---

\(^1\) Alexa is a commercial tool used to rank various Internet sites. It can be considered a rough indicator of site popularity in the absence of a standardized index.


1.