Computers in Human Behavior 86 (2018) 227–234

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

# Quantifying the Effect of Informational Support on Membership Retention in Online Communities through Large-Scale Data Analytics

Wanli Xing [a,*], Sean Goggins [b], Josh Introne [c]

[a] Department of Educational Psychology & Leadership, Texas Tech University, Lubbock, TX 79409, USA
[b] Department of Computer Science, University of Missouri, Columbia, MO 65211, USA
[c] Department of Media & Information, Michigan State University, East Lansing, MI 48824, USA

ARTICLE INFO

*Article history:*
Available online 23 April 2018

*Keywords:*
Informational support
Online communities
Text mining
Social role
Membership retention
Survival analysis

ABSTRACT

Participating in online health communities for informational support can benefit patients in various ways. For the online communities to be sustainable and effective for their participants, membership retention and commitment are important. This study explores how informational support requesting and providing by users holding different social roles (core user and periphery user) are related with participants' retention in the community. We first crawled six years of data in the WebMD fibromyalgia forum with around 200,000 posts and 10,000 users. Then a supervised machine learning model is trained and validated to automatically identify the requesting and providing informational support posts exchanged between the members in the community. Lastly, survival analysis was employed to quantify how the informational support requesting and providing by different social roles predicts the member's continued participation in the online community. The results reveal the different influencing mechanism of requesting and providing support from different social roles on the patients' decision to stay in the community. The findings can aid in the design of better support mechanisms to enhance member commitment in online health communities.



## 1. Introduction

Large numbers of American adults engage with online health communities to seek health-related informational support (Qiu et al., 2011). Doing so can not only help them to deal with stress, but can also teach them to manage their health conditions in new ways (Coulson & Greenwood, 2012). “Informational support” refers to information provided via online forums in which participants can ask about health problems and receive information and advice about treatments, coping with symptoms, side effects, and financial and other burdens (Chuang, 2013; McCormack, 2010). Online health communities have become important vehicles for informational support, and they assure users that they are cared for (Chuang, 2013). However, patients must continue to participate in online communities to receive all of their benefits.

Extensive research has been conducted into the informational support provided via online health communities. This research has primarily examined the information exchanged via these communities and the effects of such exchanges. For example, several researchers have identified various types of informational support provided via posts to online health communities and have tested how these types of support relate to individual health outcomes and a sense of community (Zhang, Liu, Deng, & Chen, 2017; Kirk & Milnes, 2016; Greene, Choudhry, Kilabuk, & Shrank, 2011). Such studies generally use methods like surveys, interviews, content analyses, and ethnographic cases, and they generally involve between a few and several hundred subjects.

While such studies are common, little is known about how the informational support exchanged among the members of online health communities affects their commitment to these communities. Understanding user commitment is important to both individual participants and whole communities. The benefits of becoming involved in a given online health community likely depend to a large extent on the information exchanged among the members of that community (Wang, Kraut, & Levine, 2012). If a member does not remain in a community, they are less likely to receive the benefits offered by that community. Moreover, members are primary sources of resources. Because communities

* Corresponding author.
*E-mail addresses:* wanli.xing@ttu.edu (W. Xing), gogginss@missouri.edu (S. Goggins).

https://doi.org/10.1016/j.chb.2018.04.042
0747-5632/Published by Elsevier Ltd.

aggregate their knowledge, a larger community is likely to know more about a given problem than a smaller one does (Butler, 2001). A sustainable level of participation cannot be taken for granted, however. Many studies have shown that large numbers of members drop out before they can contribute to or benefit from their communities (Resnick, Janney, Buis, & Richardson, 2010; Wang et al., 2012; Yang, Kraut, & Levine, 2017). For this reason, an online health community must understand the factors that affect its members' participation if it is to remain sustainable and effective.

Since members join online health communities to cope with their medical conditions, the amount and quality of informational support that they request and receive should greatly impact their commitment to their communities (Wang et al., 2012). Many sustainable online health communities exhibit a core–periphery structure (Cobb, Graham, & Abrams, 2010), in which a small but active group of core users provides most of the support requested by a much larger group of periphery users. We hypothesized that interacting with users holding different social roles have different influences on their commitment to the community. Certain members have more experience and expertise than others do in meeting participants' informational needs (Atanasova, Kamin, & Petrič, 2017), and this is likely to influence users' decisions to continue participating in their communities. Given the fact that an online health community can easily have thousands of members and tens of thousands of forum posts, understanding members' commitment to online health communities requires automated data processing methods beyond those traditionally used in the social sciences. For this reason, this study uses data mining to process posts on a large scale.

The research question for this study is, "How do requests for and provisions of informational support by members with different social roles influence members' continued participation in online health communities?" To answer this question, we crawled six years' worth of posts to the WebMD fibromyalgia forum, accumulating approximately 200,000 posts from 100,000 users. A large-scale data analysis was then used to process these posts. This analysis had two parts. First, a supervised machine-learning model was trained to automatically identify and distinguish posts requesting informational support and posts providing informational support. Second, a survival analysis was used to quantify how requests for and provisions of informational support by members with different social roles impact members' continued participation in the online community. The findings of this study could aid in the development of support mechanisms designed to increase members' commitment to online health communities.

## 2. Background

Members of online health communities typically do not know each other personally. Instead, they communicate online to acquire information and advice about a variety of health-related issues, including shared health conditions, treatments, and side effects. Studies show that the informational support provided via online health communities can benefit the members of these communities in a number of ways. McCormack (2010), for example, found that exchanging information and support can improve a patient's ability to deal with stress. Rodgers and Chen (2005) observed that patients who interacted frequently with other members of an online breast-cancer community had better mood profiles than those who did not. Braithwaite, Waldron, and Finn (1999) found that the informational support provided by online communities for chronic diseases is especially useful: it benefits not only the patients, but also their health providers and families. However, patients must continue to participate in online communities to receive all of their benefits.

### 2.1. Theoretical foundations of online-community retention

According to the resource-based model of online communities (Butler, 2001), people use their resources—e.g. their time, energy, and knowledge—to benefit the community. Äkkinen (2005) argues that if the perceived benefits of participating in a community exceed the resources sacrificed, then the community creates value for its members. Butler (2001) links perceived benefits and resources sacrificed to the number of members in the community. If the perceived benefits from a community exceed the resources sacrificed, then the number of members in the community will increase. If the perceived benefits from a community do not exceed the resources sacrificed, however, members will leave the community. Butler's study generated a comprehensive list of the benefits offered by online social communities. One of the most important of these benefits is informational support. If an online community provides to a member informational support whose benefits exceed the sacrifices made by the member, the member will likely continue to engage in the community. In contrast, if the informational support provided does not exceed the sacrifices made by the member, the member may gradually drop out of the community. Gu and Jarvenpaa (2003) echoed Butler's claim from the perspective of economic theory, stating that people will only contribute if the benefits outweigh the costs. In collective settings, such as online communities, the tangible and intangible returns valued by the members (e.g. informational support) incentivize their continued participation.

Social theories also offer arguments linking the informational support provided via online communities to user retention. Social exchange theory claims that people contribute and engage because of the benefits they expect to receive in return; this is called "future reciprocity" (Bearman, 1997). In other words, individuals engage in social interactions because they expect that doing so will result in social rewards, including status and the support of others. This theory argues that the expectation of reciprocity motivates individuals to contribute and engage beyond the equilibrium predicted by economic models of utility (Blau, 1964). Members of online communities may hold different roles (e.g. "core" and "periphery") based on their relative contributions and activity levels. Interacting with different users in requesting and providing informational support may produce different reciprocity expectations, which can, in turn, influence members' decisions to stay or to leave.

### 2.2. Empirical research into informational support and retention

Many studies on informational support have examined the contents of posts to forums. To identify the characteristics of informational support, for example, Zhang, He, and Sang (2013) qualitatively analyzed the content of 1352 posts sampled from various online health communities. Hwang et al. (2010) examined the informational support provided via an online weight-loss community. They conducted semi-structured interviews with 13 patients and characterized the support these patients received by qualitatively analyzing the transcripts of their interviews. Many other studies have investigated informational support from a relationship perspective, examining how informational support relates to other problems and/or entities (e.g. coping with diseases and a sense of community). For instance, Welbourne, Blanchard, and Boughton (2009) administered surveys to 122 members of an online infertility community. They found that informational support improves a user's sense of community and serves as a buffer between their physical health symptoms and their stress. Ginossar (2008) statistically analyzed 1424 posts to two online cancer communities to identify differences in how men and women use informational support. Høybye, Johansen, and Tjørnhøj-Thomsen

(2005) used an ethnographic case study to investigate how informational support provided via online health communities affects social isolation.

Few studies have systematically examined how requesting and receiving informational support via online health communities affects members' continued participation. Only Wang et al. (2012) has examined how receiving informational support influences members' commitment to online health communities. By analyzing a large number of posts to an online breast-cancer forum, this study found that users' time commitments decreased with the informational support they received. To fully explore the influence of informational support on user commitment, however, studies must examine more than the quantity of informational support received—they must also consider the members they actually interact with. Some members have more experience and expertise than others in providing informational support, and these members better satisfy participants' needs (Atanasova et al., 2017). For this reason, those who interact with such members may be more satisfied with the information they receive and, as a result, more committed to the community.

Many online forums have core–periphery structures, in which different users occupy different social roles in providing and requesting informational support. These forums include open-source software communities, Wikipedia communities, and online health communities (Introne, Seeman, & Goggins, 2016). The core users are a highly inter-connected group who actively provide the informational support requested by a much larger group of periphery users (Solomon & Wash, 2014). Crowston and Shamshurin (2016) examined the core–periphery structure of an open-source software community and found that the core and the periphery members differed significantly in their volumes of contributions and in their use of inclusive pronouns. Stewart, Astat, and Abidi (2012) examined the core–periphery structure of an online clinical-discussion forum. Their network analysis revealed that the core users were responsible for most of the communication. The author (hold for review) found that 13 online health communities exhibited core–periphery structures. Members play different roles in online health communities, and members with different roles may differentially affect the continued participation of other members.

Because it examined an online health community with a large number of members and a large number of posts, this study used an automatic method of data analysis to quantify users' commitment to the community. Studies have begun to research the informational support provided via online health communities by using data mining methods to analyze posts on a large scale. For example, Mukherjee, Weikum, and Danescu-Niculescu-Mizil (2014) used text-classification techniques to automatically analyze 2.8 million posts from 15,000 users of an online community for drug side effects. By comparing with officially listed side effects, their text-classification model was found to have an accuracy of 85%. Zhang, Liu, Li, and Deng (2013) used text clustering to automatically detect specified conversation topics among 3 million threads. The Rand index indicated that they obtained a very high reliability. Given that data mining allows data to be efficiently analyzed on a large scale, this study used data mining to process a large quantity of online posts. Survival analysis was then used to quantify how exchanges of informational support affected the member retention of an online health community.

## 3. Methodology

### 3.1. Research context and data

This study used the WebMD fibromyalgia forum, which contains information on health and healthcare related to this chronic disease. 172,633 posts between 2008 and 2014 were gathered from 9364 participants. Of these posts, not all were informational: many of them were non-information-related posts, including daily social interactions like “How are you?” and “Thank you for the help!” Since the WebMD fibromyalgia forum is unstructured, the reply structure between the different users had to be known before we could distinguish the core and the periphery users. We built on established work—Lin and Wang (2009) and Wang, Wang, Zhai, and Han (2011)—to develop a rule-based machine learning algorithm that could identify the reply structure of the fibromyalgia forum. Our algorithm is 90% as accurate as human inference. For details about this process, refer to our previous work (Withhold for Review). Then, to distinguish the core and the periphery users, an ideal prototypical core–periphery network was constructed. A low-rank approximation of the network matrix was then used to classify the nodes in the actual network to best match the ideal situation. Finally, a modified Pearson's correlation for weighted networks was applied to evaluate the fit of the core–periphery classification. For more details about the procedure used to distinguish the core and the periphery users, one can refer to our previous work (Withhold for Review). Generally, the core users are in the relative center of the interactions and have a smaller user sample size; periphery users exhibit less interaction with community members in the forum but have a larger user sample size. In this fibromyalgia online community, 41 members were identified as core users and 9320 were identified as periphery users.

### 3.2. Text classification

To examine the informational support exchanged via online health communities, all the “request informational support” posts and “provide informational support” posts must first be identified. However, given the substantial number of posts, it was impractical to manually identify the informational posts. An automated identification method was required. Text classification was an effective way to solve this problem. Text classification uses a supervised machine-learning method to automatically assign posts to different classes or types. In this study, since we needed to check for both “request informational support” and “provide informational support” attributes, two text classifiers were built to identify provide informational support/none and request informational support/none.

To construct the text-classification models, a training dataset was constructed (this is Step 1 in Fig. 1). A random sample of 5000 posts were selected from the WebMD health forums. Then, Amazon Mechanical Turk (MTurk) workers were employed to rate each post as “provides informational support,” “requests informational support,” “both,” or “none.” MTurk is an online market for crowd-working. It lists jobs that workers can select based on their interest (Buhrmester, Kwang, & Gosling, 2011). “Jobs” refers to the paid units of the Human Intelligence Tasks—in our study, this task was to rate a post as “provides informational support,” “requests informational support,” “both,” or “none.” Each post was paid $0.1. Snow, O'Connor, Jurafsky, and Ng (2008) found that the combined ratings of a relatively small number of novice coders were very similar to the ratings generated by content experts. Previous studies of online health communities have shown that ratings can be highly reliable, with intraclass correlation coefficients reaching approximately 0.8 (Introne et al., 2016). The Turk workers were given strict training tasks that required them to correctly classify 23 out of 25 posts before they could begin work on the final task. Each post was rated by five workers. See the specific instructions below:

*In the following, you will be asked to code posts taken from an online support message forum for the kind of informational support*

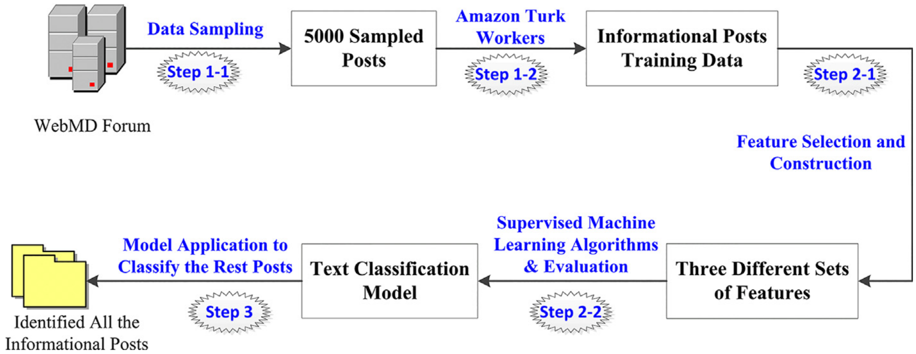



**Fig. 1.** Text classification workflow to characterize informational posts.

*content they contain. Some posts will provide and request informational support, and some will not contain any support content. You must answer each question, and you must get at least 23/25 of the following correct to be granted this qualification.*

*The types of support are as follows:*

- *Request Informational Support—When seeking informational support, the writer is trying to get advice, referrals, or knowledge related to their disease/condition.*
- *Provide Informational Support—Informational support messages provide advice, referrals or knowledge related to their disease/condition.*

*If a post does not contain any of the above, please select "None".*

In order to determine whether a given post was coded "0" (not an informational post) or "1" (an informational post), the codes from five workers were aggregated. These codes represented whether a post provided informational support, requested informational support, or offered no informational support (these posts were coded "none.") Specifically, the aggregation labeled each post with the code applied by the majority of the five workers. For example, if three out of five workers labeled Post A "provides informational support" and two other workers labeled Post A "none," then Post A was labeled "provides informational support." To assess the reliability of the workers' ratings, intra-class correlation (ICC) values were calculated. ICC is a classic measure for evaluating the consistency of a quantitative measure when objects are not rated by the same judges (Wang et al., 2012)—this was the case with the Amazon Turkers. The ICC values were very high for both "provide informational support" (ICC = 0.93) and "request informational support" (ICC = 0.90). Both achieved excellent performance ratings (0.75, 1.0) according to Cicchetti (1994). A sample post coding is provided in Table 1. It shows that the training data set for the first step of the text classification model was completed.

Next, to improve the performance of the text-classification model, three feature sets (Table 2) and a variety of algorithms were tested (this is Step 2 in Fig. 1). The first feature set was an n-gram feature, a popular linguistic feature tool used to represent text (Ghiassi, Skinner, & Zimbra, 2013). In this study, each post was transformed into an n-gram with stop words. The second feature set was the post-level features in the online forums. The following were transformed into binary features: whether or not the post was a top post, whether or not the post was a reply to the top post, and whether or not the author was an author of a top post. Finally, the depth of the post in the particular thread was transformed into a numeric value and treated as a post-level feature. The third feature set was for punctuation-related features. This study specifically targeted posts that identified, requested, and provided informational support. Intuition suggests that these posts might tend to have more question-mark-related punctuation. For this reason, all punctuations containing question marks (such as "?", "?!", and "!???") were retrieved from all posts. These punctuation features were then transformed into binary features so that each post would show whether that post contained that specific punctuation. All of these features were extracted from the training posts and then inputted into different text-classification algorithms (Feldman & Sanger, 2007)—e.g. Naïve Bayes, Logistic Regression, Support Vector Machines, and Decision Tree. All of the text-classification models were tested under 10-fold cross validation (Refaeilzadeh, Tang, & Liu, 2009) to improve the reliability of the constructed model.

After the text-classification model was built and its performance was validated, the third step was to apply the classification model to the remaining posts in the 23 WebMD forums (this is Step 3 in Fig. 1). As a result, each post in every WebMD forum was labeled "provide information," "request information," "both," or "none."

### 3.3. Survival analysis

While controlling for factors unrelated to informational support, this study used survival analysis to quantify the effects of requests for and provisions of informational support by members with different social roles on a person's commitment to an online health community. Survival modeling can estimate the truncated nature of

**Table 1**
Sample post coding.

| Post | Request Infor. | Provide Infor. | None |
|---|---|---|---|
| *Hi, I'm new, I have fibromyalgia, but lately I have been sufefering of muscles cramping is that part of fibromyalgia, I can't sleep or exercise, even walking hearts. I would like to know if someone have an idea what can I do to help me with problem. Thank you in advance.* | 1 | 0 | 0 |
| *If I may suggest … try drinking green tea w/o sugar, eating oatmeal with a small amount of sugar (not pre-flavored), apples, and take a fiber supplement. Green tea speeds up your metabolism and the rest of it helps your body feel full longer! Eat very small meals throughout the day and this will also keep up yor metabolism.* | 0 | 1 | 0 |
| *Bless you …. ….I needed that laugh right now !!!!!!!!* | 0 | 0 | 1 |

**Table 2**
Feature set.

| Features | Description |
|---|---|
| N-gram | One gram includes stop words |
| Post features | Top post, reply to top post, author of top post, depth of the post |
| Punctuation | Specifically focused on question mark usage in the post |

time-series data in a less biased manner than can standard regression techniques (Yang, Wen, Howley, Kraut, & Rose, 2015). The Hazard Ratio was used to determine the influence of an independent variable on the probability of a user's dropping out (Klein & Moeschberger, 2005). A parametric regression of survival analysis was assumed for the Weibull distribution of survival times. The timestamp of a user's first post was used to determine the starting point of their participation, and the timestamp of the user's last post was used to determine the end point of their participation.

#### 3.3.1. Dependent variable

"Dropout": A member was considered a dropout from the community if they failed to post within three months of their last post. Users who posted only once were not considered. Three months is an appropriate interval and has been used in many other studies (Wang et al., 2012). Therefore, a user could have dropped out of the community and rejoined it several times. Of the total users, only 791 users reentered the forum showing the three month is a sufficient threshold to determine whether the users leave the community or not. We considered these users as having survived, but their survival time was calculated as the sum of the time intervals smaller than three months. Since the users whose last posts were within three months of the end of the data collection period could still have been participating, these users were right-censored in the analysis.

#### 3.3.2. Control variables

"Average Post" (AvePost): This variable was the average number of posts a user contributed to the forums in three months. It was considered a baseline in measuring the user's engagement.

"Thread Starter" (Tstarter): Users who started conversations may have been different from those who participated in them. This variable was quantified as the average number of threads a user started over three months.

#### 3.3.3. Independent variables

"Request Informational Support" (two variables): These independent variables measured the average number of informational support requests logged by other members every three months. It was calculated by dividing the total number of posts requested directly by the community's users by the number of three-month periods a user stayed in the community. Since two types of users can request informational support, two variables were built for "requesting informational support": "core user requesting informational support" (CRIS) and "periphery user requesting informational support" (PRIS).

"Provide Informational Support" (two variables): These independent variables measured the average number of informational support posts provided by other members every three months. Therefore, it was actually a measure of the informational support received by the user. It was calculated by dividing the total number of posts provided directly by the community's users by the number of three-month periods the user stayed in the community. Similarly, two variables were built for "providing informational support": "core user providing informational support" (CPIS) and "periphery user providing informational support (PPIS)."

## 4. Results

### 4.1. Text classification

The training data was inputted into various features and machine-learning algorithms to optimize the performance of the text-classification in identifying informational posts. The performance results for specific text-classification models are presented in Table 3. This table shows two text-classification models constructed using training data for "provide informational support" (the best performance was 83.5%) and "request informational support" (the best performance was 91.5%), respectively. The general accuracy range for the text classification models in the literature is between 60% and 85% (Khan, Baharudin, Lee, & Khan, 2010). Therefore, "provide informational support" and "request informational support" classification models are considered excellent text-classification models.

We then applied these text-classification models to the rest of the data from the 23 forums to automatically identify all of the informational posts. Each post was labeled "provide informational support," "request informational support," "both provide informational support and request informational support," or "none." Of the total posts in the community, 38,014 were identified as informational posts, 14,389 were posts requesting informational support, 23,001 were posts providing informational support, and 624 were posts both requesting and providing informational support. Table 4 shows the descriptive statistics for the variables used in the survival analysis, which is reported in the next section.

### 4.2. Survival analysis

While controlling for the effects of "average number of posts" and "average threads started," this study conducted a survival analysis to examine how requests for and provisions of informational support by core and periphery users influenced members' continued participation in the health community. The effect was quantified via hazard ratio to explain the influence of a predicating variable on the retention rate of a user. All variables were standardized with a mean of 0 and a standard deviation of 1. Therefore, the hazard rate predicted the ratio of the probability of whether a user left the community when there was a unit increase in the number of these six independent variables.

In Table 5, Model 0 shows the effects on a members' commitment time of the control variables of "average post number" and "threads started." The hazard ratio for "average post number" is 0.808, indicating that those members who contributed a standard deviation more posts than the mean were 19.2% ($100\% * (1-0.808)$) more likely to survive than those with numbers of posts lower than the average. Similarly, the hazard ratio for "threads started" was 0.799, indicating that those members who started a standard deviation more threads than the mean were 20.1% more likely to stay engaged in the health community.

While controlling for "average number of posts" and "threads started," Model 1 demonstrated the impact on dropout of requesting informational support from users holding different social roles. The results show that requesting informational support

**Table 3**
Classification performance.

| | Naïve Bayes | | Logistic Regression | | SVM | | Decision Tree | |
|---|---|---|---|---|---|---|---|---|
| | P | R | P | R | P | R | P | R |
| N-gram | 59.3% | 62.6% | 82.1% | 86.9% | 80.3% | 84.7% | 76.3% | 82.2% |
| Vectorized Feature + postural + Question | 60.8% | 60.9% | 78.1% | 89.8% | 80.0% | 90.0% | 71.1% | 86.1% |
| N-gram + postural + Question | 70.8% | 62.9% | 83.5% | 91.5% | 81.6% | 89.6% | 76.6% | 89.0% |

P: Provide Informational Support.
R: Request Informational Support.

**Table 4**
Descriptive statistics for the survival analysis.

| | Core | | | | Periphery | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Min | Max | Mean | SD | Min | Max |
| AvePost | 393.48 | 372.87 | 28 | 1738 | 9.58 | 21.59 | 2 | 310 |
| TStarter | 66.25 | 71.97 | 3 | 359 | 3.55 | 7.50 | 0 | 205 |
| CRIS | 3.70 | 16.04 | 0 | 90 | 0.79 | 2.14 | 0 | 29 |
| PRIS | 51.10 | 41.71 | 0 | 212 | 4.65 | 8.15 | 0 | 116 |
| CPIS | 1.82 | 8.15 | 0 | 44 | 0.08 | 0.34 | 0 | 5 |
| PRIS | 16.18 | 24.53 | 0 | 150 | 0.61 | 1.51 | 0 | 18 |

from core users did not influence a member's survival time in the community. However, those who requested informational support from periphery users one standard deviation more times than the average were 29.3% more likely to remain in the community.

Model 2 shows that when “average number of posts,” “threads started,” and “request informational support” are controlled for, the informational support provided by core and periphery users influenced members' survival rates, albeit in opposite directions. Users who received support from core users’ one standard deviation more times than the average were 9.2% more likely to stay in the health community. In contrast, those who received support from periphery users one standard deviation more times than the average were 11.4% more likely to leave the group. In other words, receiving more informational support from core users was associated with staying, while receiving more informational support from periphery users was associated with leaving. Fig. 2 illustrates the results graphically. Fig. 2(a) shows “mean survival” with all of the predicating variables. Fig. 2(b) shows “mean survival” with each predicating variable.

## 5. Discussion and conclusion

Participating in online health communities for informational support can benefit patients in multiple ways (McCormack, 2010). However, people cannot benefit from such communities if they drop out. In this study, we quantified the effect of exchanging informational support on whether users remained in an online health community. To do so, we built a machine-learning model to accurately and automatically detect requests for and provisions of informational support. We then used survival analysis to compute the effects on continued participation of requesting from and providing informational support to core and periphery users. The results revealed that requesting and providing informational support from users with different social roles differentially impacted users’ commitment. This study analyzed posts to the fibromyalgia forum because users of communities for people with chronic diseases tend post more actively.

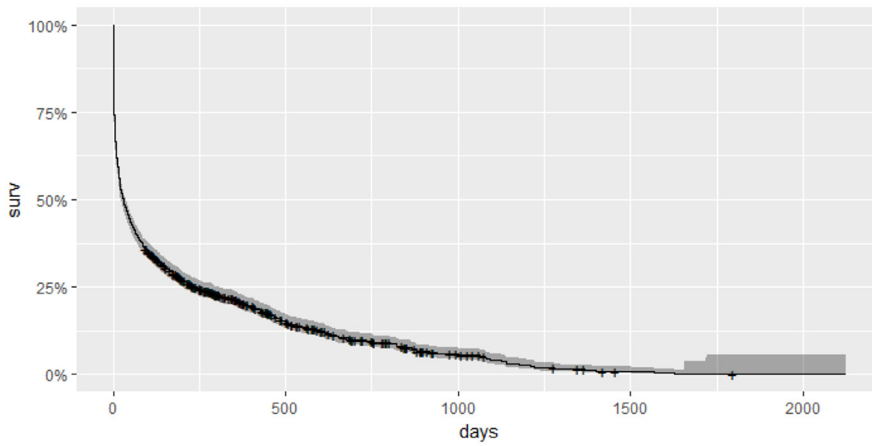
The results suggest that requests for informational support from periphery users are associated with higher community commitment, while requests for informational support from core users are not. Since core users are usually much more experienced and active (Atanasova et al., 2017), questions from core users may be more difficult for other members to answer and may turn such users away. Additionally, since core users mainly provide support, requests for support from core users may not raise the awareness of other users to an extent sufficient to influence their decisions to stay. In contrast, the main purpose of online health communities is to help periphery users, who have much less experience and expertise in dealing with chronic diseases (Solomon & Wash, 2014). For this reason, requests for support from periphery users are easier to respond to. Moreover, core users may spend more time and effort interacting with periphery users. As a result, users are more likely to continue participating in the community.

The provision of informational support manifests a totally different mechanism of influence. While support provided by core users is positively associated with users’ commitment to the online community, support provided by periphery users is negatively associated with user commitment. This phenomenon also likely results from the differences in experience and expertise between core users and periphery users. Support provided by core users is likely to be extensive and understandable, while support from periphery users may be less professional and may even contain errors (Cobb et al., 2010). It is therefore not surprising that support from different users differentially impacts the time that users commit to a community. In contrast to previous studies, which found that exposure to informational support negatively influences member commitment (Wang et al., 2012), this study provides a granular mechanism that better explains the influence of informational support on member retention.

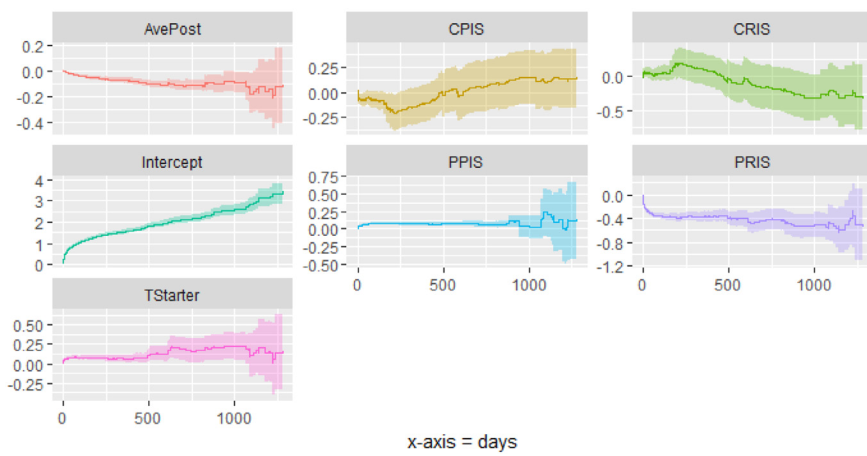
These findings have significant implications for the delivery of informational support via online communities. Although the machine-learning mining model was constructed to investigate

**Table 5**
Survival analysis results.

| | Model 0 | | Model 1 | | Model 2 | |
|---|---|---|---|---|---|---|
| | Hazard Ratio | *p* | Hazard Ratio | *p* | Hazard Ratio | *p* |
| AvePost | 0.808 | 0.000*** | 0.838 | 0.006** | 0.857 | 0.019* |
| TStarter | 0.799 | 0.000*** | 1.089 | 0.202 | 1.071 | 0.257 |
| CRIS | | | 0.977 | 0.438 | 1.054 | 0.209 |
| PRIS | | | 0.707 | 0.000*** | 0.668 | 0.000*** |
| CPIS | | | | | 0.908 | 0.037** |
| PPIS | | | | | 1.114 | 0.005*** |




(a) Overall survival curve




(b) Independent variable effect on survival

**Fig. 2.** Overall and independent effect on members' survival time.

behavior in online health communities, it can be modified to serve as a basis for active interventions. If a given member receives informational support only from periphery users, for instance, the forum moderator could direct additional resources to that user. These resources could include support from core users or from a medical professional, and they could decrease the likelihood of that user dropping. Moreover, to remain sustainable, online health communities must encourage periphery users to request informational support and discourage them from providing it.

These results also have important methodological implications. Research into informational support and online health communities in general usually relies on traditional social science methodologies, including ethnographic analyses, surveys, interviews, content analyses, and statistical testing (Zhang & Sang, 2013; Welbourne et al., 2009). This study also demonstrates that automatic data-mining techniques can identify informational posts with sufficient accuracy. In fact, the technique used in this study was built on the work of previous researchers. Future studies, including studies into online health communities, should consider going beyond traditional methods by employing large-scale data analytics.

An important limitation of this study was that even though we addressed longitudinal effects, the findings were correlational. While the results were consistent with the assumption that informational support influences commitment, they could also be interpreted as expressions of preexisting differences among members of the community. Further experiments using different survey methods (Shin, 2017 & 2018) might determine more precisely the extent to which informational support influences the continued participation of members of health communities. Another limitation of this study is that it only considered users who had posted more than once. As a result, it may have overlooked "lurkers," who participate only by reading. A final limitation of this study is that it examined only one chronic disease forum: an online fibromyalgia forum. For this reason, its findings should be applied with caution to other online health communities.

A number of directions for future research are apparent. First, even though our findings indicate that the impact of informational support on user retention varies with the roles of the individuals requesting and providing the support, why this is the case is not completely clear. Future studies could interview and survey patients to determine what motivates them to stay or to leave. Second, informational support is just one of the many types of support exchanged via online health communities. Future studies could investigate the influences of other types of social support on user retention, including the influence of emotional support and the influence of community support. Third, patients do not participate in online health communities to promote participation in or to reduce dropout from these communities. It would be useful to study the extent to which informational support impacts patients' health conditions.

## Conflicts of interest

The authors declare that they have no conflict of interests.

## References


Äkkinen, M. (2005). *Conceptual foundations of online communities*. Retrieved from: https://aaltodoc.aalto.fi/bitstream/handle/123456789/12293/isbn951791959X.pdf?sequence=1.

Atanasova, S., Kamin, T., & Petrič, G. (2017). Exploring the benefits and challenges of health professionals' participation in online health communities: Emergence of (dis) empowerment processes and outcomes. *International Journal of Medical Informatics, 98*, 13–21.

Bearman, P. (1997). Generalized exchange. *American Journal of Sociology, 102*(5), 1383–1415.

Blau, P. M. (1964). *Exchange and power in social life*. New York: Wiley.

Braithwaite, D. O., Waldron, V. R., & Finn, J. (1999). Communication of social support in computer-mediated groups for people with disabilities. *Health Communication, 11*(2), 123–151.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3–5.

Butler, B. S. (2001). Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research, 12*(4), 346–362.

Chuang, K. Y. (2013). *Understanding user interaction patterns in health social media*. Doctoral dissertation. Drexel University.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284.

Cobb, N. K., Graham, A. L., & Abrams, D. B. (2010). Social network structure of a large online community for smoking cessation. *American Journal of Public Health, 100*(7), 1282–1289.

Coulson, N. S., & Greenwood, N. (2012). Families affected by childhood cancer: An analysis of the provision of social support within online support groups. *Child: Care, Health and Development, 38*(6), 870–877.

Crowston, K., & Shamshurin, I. (2016 May). Core-periphery communication and the success of free/libre open source software projects. In *IFIP international conference on open source systems* (pp. 45–56). Springer International Publishing.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press.

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications, 40*(16), 6266–6282.

Ginossar, T. (2008). Online participation: A content analysis of differences in utilization of two online cancer communities by men and women, patients and family members. *Health Communication, 23*(1), 1–12.

Greene, J. A., Choudhry, N. K., Kilabuk, E., & Shrank, W. H. (2011). Online social networking by patients with diabetes: A qualitative evaluation of communication with facebook. *Journal of General Internal Medicine, 26*(3), 287–292.

Gu, B., & Jarvenpaa, S. (2003). Online discussion boards for technical support: The effect of token recognition on customer contributions. In *ICIS 2003 proceedings* (p. 10).

Høybye, M. T., Johansen, C., & Tjørnhøj-Thomsen, T. (2005). Online interaction. Effects of storytelling in an internet breast cancer support group. *Psycho-oncology, 14*(3), 211–220.

Hwang, K. O., Ottenbacher, A. J., Green, A. P., Cannon-Diehl, M. R., Richardson, O., Bernstam, E. V., et al. (2010). Social support in an Internet weight loss community. *International Journal of Medical Informatics, 79*(1), 5–13.

Introne, J., Seeman, B., & Goggins, S. (2016). A sociotechnical mechanism for online support provision. In *Proceedings of the ACM conference on human factors in computing (CHI)* (San Jose, CA).

Khan, A., Baharudin, B., Lee, L. H., & Khan, K. (2010). A review of machine learning algorithms for text-documents classification. *Journal of Advances in Information Technology, 1*(1), 4–20.

Kirk, S., & Milnes, L. (2016). An exploration of how young people and parents use online support in the context of living with cystic fibrosis. *Health Expectations, 19*(2), 309–321.

Klein, J. P., & Moeschberger, M. L. (2005). *Survival analysis: Techniques for censored and truncated data*. Springer Science & Business Media.

Lin, C., Yang, J. M., Cai, R., Wang, X. J., & Wang, W. (2009 July). Simultaneously modeling semantics and structure of threaded discussions: A sparse coding approach and its applications. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (pp. 131–138). ACM.

McCormack, A. (2010). Individuals with eating disorders and the use of online support groups as a form of social support. *Computers Informatics Nursing, 28*(1), 12–19.

Mukherjee, S., Weikum, G., & Danescu-Niculescu-Mizil, C. (2014 August). People on drugs: Credibility of user statements in health communities. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 65–74). ACM.

Qiu, B., Zhao, K., Mitra, P., Wu, D., Caragea, C., Yen, J., & Portier, K. (2011 October). Get online support, feel better—sentiment analysis and dynamics in an online cancer survivor community. In *Privacy, security, risk and trust (PASSAT) and 2011 IEEE third inernational conference on social computing (SocialCom), 2011 IEEE third international conference on* (pp. 274–281). IEEE.

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In *Encyclopedia of database systems* (pp. 532–538). Springer US.

Resnick, P. J., Janney, A. W., Buis, L. R., & Richardson, C. R. (2010). Adding an online community to an internet-mediated walking program. Part 2: Strategies for encouraging community participation. *Journal of Medical Internet Research, 12*(4).

Rodgers, S., & Chen, Q. (2005). Internet community group participation: Psychosocial benefits for women with breast cancer. *Journal of Computer-mediated Communication, 10*(4), 00–00.

Shin, D. (2017). The role of affordance in the experience of virtual reality learning: Technological and affective affordances in virtual reality. *Telematics and Informatics, 34*(8), 1826–1836.

Shin, D. (2018). Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience? *Computers in Human Behavior, 78*, 64–73. https://doi.org/10.1016/j.chb.2017.09.012.

Snow, R., O'Connor, B., Jurafsky, D., & Ng, A. Y. (2008 October). Cheap and fast—-but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 254–263). Association for Computational Linguistics.

Solomon, J., & Wash, R. (2014 May). Critical mass of What? Exploring community growth in WikiProjects. In *ICWSM*.

Stewart, S. A., & Abidi, S. S. R. (2012). Applying social network analysis to understand the knowledge sharing behaviour of practitioners in a clinical online discussion forum. *Journal of Medical Internet Research, 14*(6), e170.

Wang, Y. C., Kraut, R., & Levine, J. M. (2012, February). 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* (pp. 833–842). ACM.

Wang, H., Wang, C., Zhai, C., & Han, J. (2011 July). Learning online discussion structures by conditional random fields. In *Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval* (pp. 435–444). ACM.

Welbourne, J. L., Blanchard, A. L., & Boughton, M. D. (2009 June). Supportive communication, sense of virtual community and health outcomes in online infertility groups. In *Proceedings of the fourth international conference on Communities and technologies* (pp. 31–40). ACM.

Yang, D., Kraut, R., & Levine, J. M. (2017, May). Commitment of newcomers and oldtimers to online health support communities. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 6363–6375). ACM.

Yang, D., Wen, M., Howley, I., Kraut, R., & Rose, C. (2015 March). Exploring the effect of confusion in discussion forums of massive open online courses. In *Proceedings of the second (2015) ACM conference on learning@ scale* (pp. 121–130). ACM.

Zhang, Y., He, D., & Sang, Y. (2013). Facebook as a platform for health information and communication: A case study of a diabetes group. *Journal of Medical Systems, 37*(3), 1–12.

Zhang, X., Liu, S., Deng, Z., & Chen, X. (2017). Knowledge sharing motivations in online health communities: A comparative study of health professionals and normal users. *Computers in Human Behavior, 75*, 797–810.



Wanli Xing is an Assistant Professor in Instructional Technology at Texas Tech University, USA with background in learning sciences, statistics, computer science and mathematical modeling. His research interests are online community, educational data mining, learning analytics, and CSCL.

Sean Goggins is an Associate Professor in Computers Science at University of Missouri Columbia. His research focuses on understanding new types of organizations in open source software and health care. Both from a data focused point of view.

Josh Introne is an Assistant Professor in the Department of Media and Information. Dr. Introne holds masters and Ph.D degrees from the Computer Science Department at Brandeis University. His research is mainly collective intelligence.