THE EVOLUTION OF USER ROLES IN ONLINE HEALTH COMMUNITIES – A SOCIAL SUPPORT PERSPECTIVE

Xi Wang, Interdisciplinary Graduate Program in Informatics, The University of Iowa, Iowa City, IA, USA, xi-wang-1@uiowa.edu

Zhiya Zuo, Interdisciplinary Graduate Program in Informatics, The University of Iowa, Iowa City, IA, USA, zhiya-zuo@uiowa.edu

Kang Zhao, Tippie College of Business, The University of Iowa, Iowa City, IA, USA, kang-zhao@uiowa.edu (Corresponding author).

Abstract

Online Health Communities (OHCs) have become a major source of social support for people with health problems. Using a case study of an OHC among breast cancer survivors, we revealed the types of social support embedded in each post using text mining techniques. Then we aggregated users' activities in different types of social support and identified different roles that users play in an OHC via unsupervised machine learning techniques. By analyzing how users' roles change over time, we constructed a transition graph to illustrate the evolution of users' roles in an OHC. In addition, we discovered that a user's behavior in receiving social support is correlated with the transition of her role. It was revealed that the types social support received by a user may facilitate or delay her role transitioning. Our research has implications for OHC operators to track users' behaviors in order to manage and sustain an OHC.

Keywords: User Profiling, User Role Evolution, Social Support, Online Health Communities, Machine learning, Text Mining.

1 INTRODUCTION

Nowadays, healthcare becomes a global problem that many people pay attention to. Thanks to the development of the Internet, new technology and information can spread more rapidly and widely than before. According to the report of Fox and Duggan (2013), 35% of U.S. adults have gone online particularly for information related to medical conditions. Besides getting information from healthcare professionals and friends, 24% of adults also sought information or support from peers who have the same health condition. A major venue where people find such peers is Online Health Communities (OHCs), such as "Patientslikeme.com", "breastcancer.org" (for breast cancer survivors) and "quitsmoking.com" (for smokers). Compared with traditional health-related websites that only allow users to retrieve information, online health communities (OHCs) increased members' ability to interact with peers facing similar health problems and as a result better meet their immediate needs for social support. It was estimated that 5% of all Internet users participated in an OHC (Chou et al. 2009).

While people use OHCs for a wide range of needs. Generally speaking, participation in OHCs is helpful for users to get therapeutically information for their concern, enhance the understanding of medical knowledge, heighten the level of their emotional comfort (Lieberman, 2007) and personal empowerment (Barak et.al, 2008) and even strengthen offline social connections. Among all of these, obtaining psychosocial support is one of the key benefits for the users (Kim et al. 2012; Rodgers and Chen 2005). Research has found that such support can help patients adjust to the stress of living with and fighting against their disease (Dunkel-Schetter 1984; Qiu et al. 2011; Zhao et al. 2014) and is a consistent indicator of survival (McClellan et al. 1993).

Early in the 1990s, Ferguson (1996) talked about online self-helpers can receive and give opinions in online self-help forums. Connections between users make the forum a community, and provide an evidence of social support. According to (Shumaker and Brownell 1984), social support refers to the "exchange of resources between at least two individuals perceived by the provider or the recipient to be intended to enhance the well-being of the recipient". Literatures on social support suggest that OHCs mainly feature three types of social support: informational support, emotional support, and companionship (a.k.a., network support) (Bambina 2007; Keating 2013). Informational support is the transmission of information, suggestion or guidance to the community users (Krause 1986). The content of such a post in an OHC is usually related to advice, referrals, education and personal experience with the disease or health problem. Example topics include side effects of a drug, ways to deal with a symptom, experience with a physician, or medical insurance problems. Emotional support, as its name suggests, contains the expression of understanding, encouragement, empathy affection, affirming, validation, sympathy, caring and concern, etc. Such support can help one reduces the levels of stress or anxiety, which is sometimes called appraisal support as well. Companionship or network support consists of chatting, humor, teasing, as well as discussions of offline activities and daily life that are not necessarily related to one's health problems. Examples include sharing jokes, birthday wishes, holiday plans, or online scrabble games. Companionship helps to strengthen group members' social network and sense of communities. Note that instrumental support (which support others with the concrete, such as financial assistance or material goods) may also exist in a support group (House, 1981). However, such support is usually rare in OHCs that span a large geographical area. Therefore, we do not consider instrumental support in this study of OHCs.

As social support is a pillar of OHCs, a natural question to ask would be: when it comes to user participation, are a user's online activities in different types of social support related to her/his participation in an OHC? Previous study showed that receiving more emotional support or contributing more companionship are positively correlated with users' longer involvement in OHCs, while receiving informational support is negatively correlated (Wang et al. 2012, Wang et al. 2014). By analyzing large-scale data from an OHC, our previous work (Wang et al. 2014) identified several roles of users in an OHC and suggested that overall those who only seek social support are more likely to churn from OHCs. However, the conclusion is based on a static snapshot of an OHC. Would a user switch from one role (e.g., support seeker) to another (e.g., support provider) over time? If so, what are the frequent patterns in such role evolution? How can we leverage such evolution for OHC management?

The literature features several studies on users' roles in online communities. For example, Preece and Shneiderman (2009) proposed a framework to describe the user role evolution in computer-mediated social participation. Reader, contributor, collaborator and leader as 4 users roles were defined. The authors also indicated users could move from one stage to another in either nonlinear or linear way. However, there was no empirical data or experiment in that paper to support the framework. Füller et al (2014) identified different type of user roles in a jewelry design contest community based on network metrics, and they suggested successfully managing the diverse roles individuals play in the online community would be helpful to community management. Specifically for OHCs, many qualitative studies found users acting as different roles, and a user's role may change over time. Pfeil and Zaphiris (2007) analyzed patterns of empathy in a website for senior citizens. They found users could be divided as target (who was experiencing the problem) and empathizer (who was providing the help), but the users do not strictly act as one specific role. Loane and D'Alessandro (2013) pointed out that majority of OHC users joined in the community as Information Seekers. As time passes, participants could also contribute as support providers. Such conclusions were based on analyzing the content of posts in OHCs following some coding scheme. Such a manual coding approach significantly limited that amount of data they can analyze.

To the best of our knowledge, our work is the first one to analyze the evolution pattern of OHC users' roles in social support activities using a large-scale dataset. We addressed the following three research questions regarding social support and user role evolution in OHCs:

(1) How many different user role categories are there in an OHC? In addition to support seekers and providers, are there any other types of contributors?
(2) Can we automatically capture the dynamic evolution of users' roles? Or in other words, does a user's role in an OHC shift over time?
(3) If users' roles do evolve, what kinds of transitions are more frequent? What are the factors related to the transition of users' roles?

Our research considered all three major types of social support in an OHC and tried to reveal users' roles based on their activities in different types of social support. The outcome of this research has implications for building and sustaining an active OHC through better thread/post recommendations and community management.

2 SOCIAL SUPPORT DETECTION

In this research, we used a very popular peer-to-peer OHC among breast cancer survivors and their caregivers. With more than 140,000 registered users, the website provides various ways for its members to communicate, including discussion forum, private messaging, friend subscription, listserv, etc. Its online forum has 73 discussion boards. Our dataset consists of all the public posts and user profile information from October 2002 to August 2013. There are 107,549 threads, with more than 2.8 million posts contributed by 49,552 users.

As we mentioned earlier, informational support, emotional support, and companionship are three major types of social supports in OHCs. To assign user role in this OHC, for each post, we need to determine whether it was seeking informational support (SIS), providing informational support (PIS), seeking emotional support (SES), providing emotional support (PES), or simply about companionship (COM). Note that we did not differentiate seeking and provision of companionship, for the nature of companionship is about participation and sharing. By getting involved in activities or discussions about companionship, one is seeking and providing support at the same time. It is also possible that a post could belong to more than one of the categories above. Table 1 lists example posts for each category and a post that belongs to two categories.

Social Support Category	Examples
Companionship (COM)	*Kelly Have a wonderful time in Florida, enjoy the sun and fun. Heather* *I'm loving her new CD. Didn't recognize any of the songs at first, but there are a few now that I find myself singing the rest of the day.* *This game has the poster making a new 2 word phrase starting with the second word of the last post Example: Post : Hand out Next poster: Out cast Next poster: Cast Iron Next poster: Iron Age Now let's begin the game~ Age Old*
Seeking Informational Support (SIS)	*Where do you buy digestive enzymes and what are they called?*
Seeking Emotional Support (SES)	*I feel like everyone else's lives are going forward, they have plans, hopes, aspirations because they feel. I am one of those not yet out of the woods. I was also someone who could never get cancer. I was a good person, exercised, ate well. Good people don't get sick. I have taken the step of antidepressants, they mitigate the damage, but do not block the pain or sadness I feel.*
Providing Informational Support (PIS)	*I had surgery Aug05 for bc recurrance. B4 surgery I had 33 IMRT rads, prior to that had 4A/C & 4 Taxol. I had bc in 2000 & had 37 rads in same general area. Now, my surgery won't heal. Wound doc says there is adema or something on my sternum (shown on recent MRI). My wound has been draining since it broke open in Sept.*
Providing Emotional Support (PES)	*Hope you feel better soon, we are here! Prayers Hugs come from Massachusetts APPLE♥.*
Providing Informational Support (PIS) & Providing Emotional Support (PES)	*I am also the daughter of a 35 yrs BC survivor. Mom is just now going through some more Cancer - alas - they found it in her lung, but it is totally unlikely to be a follow-up of her old BC. I am 45, and was 43 at DX time, my mom was diagnosed at 38... and I am a BRCA2 carrier. Tina, one day at a time. Maybe you'll get good news - it is so hard to wait!!! It is also important to remember that - whatever it is, it is highly treatable, and that YOU WILL SURVIVE too!!! and life goes on after. It will take some time, but it goes on... see my picture? even the hair is back!!! Hugs to all. I am happy you all found your way here, it is a great site for exchanging information, learning and finding support.*

Table 1. Example posts for types of social support.

As it is almost impossible to label all 2.8 million posts manually, we used classification algorithms to decide what kind(s) of social support each post is about. To train the classification algorithm, we leveraged human annotated data. We randomly selected 1,333 (54 initial posts and 1,279 comments) out of our dataset. Five human annotators were asked to read each post following an annotation guideline and decide whether the post is related to one or more categories of social supports. To control the quality of human annotations, we also added to the pool 10 posts that have been annotated by domain experts. For each post, we only accepted results from annotators whose performance on the 10 quality-control posts is among top 3. Results from the other two annotators were discarded. The average value of inter-annotator agreement (Cohen's Kappa coefficient) of top 3 annotators is 0.671, which suggests decent quality of agreement according to Altman, 1991. Then a majority vote was used to determine whether a post is related to a specific category of social support.

Users in OHCs have different writing styles or linguistic preference to express their opinions. To capture these characteristics, we examined each post and extracted various types of features for the classifier: basic features, lexical features, sentiment features, and topic features. A list of features can be found in our previous work (Wang et al 2014). After comparing the performance of different algorithms for the five categories of social support. AdaBoost with Naïve Bayesian as the weak learner was chosen to classify PES[1], PIS and SIS, AdaBoost with DecisionStump (trees) as the weak learner was chosen to classify COM, while logistic regression was the best choice for SES. Overall, our classifiers achieved decent performance with accuracy rates ranging from 0.804 to 0.914 in all five classification-tasks (see Table 2). After applying the best-performing five classifiers on the remaining of the 2.8 million posts, each post was assigned 5 binary labels, indicating whether the post belongs to a specific type of social support. The total numbers of posts in each category are listed in Table 3.

Social support	Results	Naïve Bayes	Logistic Regression	SVM (Poly Kernel)	Random Forest	Decision Tree (J48)	AdaBoost
COM	Accuracy	0.696	0.787	0.783	0.771	0.767	0.804
	ROC Area	0.839	0.817	0.768	0.848	0.75	0.852
PES	Accuracy	0.713	0.830	0.840	0.830	0.81	0.817
	ROC Area	0.823	0.787	0.681	0.825	0.687	0.817
PIS	Accuracy	0.753	0.813	0.823	0.767	0.779	0.801
	ROC Area	0.824	0.83	0.783	0.837	0.717	0.859
SES	Accuracy	0.893	0.901	0.970	0.967	0.963	0.963
	ROC Area	0.749	0.867	0.656	0.851	0.671	0.668
SIS	Accuracy	0.851	0.880	0.943	0.931	0.937	0.914
	ROC Area	0.893	0.803	0.745	0.86	0.766	0.869

Table 2. Performance of 6 classification algorithms of classifying social supports.

Intuitively, there are more posts to provide support than to seek support. This is what most would expect from a popular OHC with a large and active user base. About 37% of the posts provided informational support, making it the largest group among the five. In other words, providing information support is the most popular activity in the OHC. Companionship posts constitute the second largest group, which suggests that members of the OHC did form a strong sense of community and discussed many issues other than cancer. In addition, 197,956 posts were predicted to provide informational and emotional support at the same time, representing the largest group with more than one category of social support. Meanwhile, a few posts were predicted as none of these 5 categories, which did not contain any social support in their content.

Social support category	Number of posts
Companionship (COM)	932,538
Seeking Informational Support (SIS)	284,027
Seeking Emotional Support (SES)	227,188
Providing Informational Support (PIS)	1,034,682
Providing Emotional Support (PES)	497,096

Table 3. Total numbers of posts in each category of social supports.

3 EVOLUTION OF USER ROLE

After estimating the nature of social support in each post, we could then build a profile for each user by aggregating her/his posts by their social support categories. We represented each user's social support involvement with a *social support activity profile*, which is a 1×5 vector. Each element in the vector is the percentage of the user's social supports in a social support category. For example, user Mary has published 10 social support related posts, with 3 companionship posts, 4 posts providing emotional support, 2 posts providing informational support, 1 post seeking emotional support, and no post seeking informational support. Then she will have a vector of <0.3, 0.4, 0.2, 0.1, 0>.

[1] Although the results of accuracy and ROC area of random forest are slightly better than AdaBoost for the PES classifier, the random forest classifier has much worse recall and f-measure. Thus we decided to choose AdaBoost.

With social support distribution vectors of 47,581 users, we applied the classic K-means clustering algorithm to divide users into k groups, so that the users with similar social support distributions would belong to the same cluster. To find the best grouping of users, we tested various K values (from 2 to 20) and clustering results with Davies-Bouldin Index (DBI) (Davies and Bouldin, 1979). DBI is defined as Equation 1, where $D_{intra}(C_i)$is the average distance from all the other users in cluster C_i to the centroid of C_i, and $D_{inter}(C_i, C_j)$ is the distance between centroids of $C_i$and C_j. Euclidean distance was used for this study. Generally speaking, DBI prefers smaller groups, for the value of intra-cluster distance is lower in the smaller group, and penalizes short inter-cluster distances. Therefore, the solution with the lowest DBI provides relative balance of small clusters and long distances between every two clusters.

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{j:i\neq j}\left\{\frac{D_{intra}(C_i)+D_{intra}(C_j)}{D_{inter}(C_i,C_j)}\right\} \qquad \text{(Equation 1)}$$

After testing K values from 2 to 20, K=7 yielded the lowest DBI value and hence the best clustering results. Centroids for each of the 7 clusters are shown in Table 4.

Results in Table 4 suggest that, intentionally or not, OHC users do have different patterns in social support involvement and thus play different roles in the community. Some users' posts focused on one major category of social support. For example, users in cluster 0 published an average of 96.55% of their social support posts are to provide informational support. They obviously act as *Information Providers* in the community. Similarly, cluster 1 is for *Companionship Builders* with 64.92% of the supports in companionship, and cluster 4 consists of *Emotional Support Providers.* The two smallest clusters are for seekers: cluster 3 for *Information Seekers* and cluster 6 for *Emotional Support Seekers*. Meanwhile, users in cluster 2, the largest cluster of the seven, are *All-around Contributors* with relatively balanced profiles in each social support category. Cluster 5 represents a group of *Information Enthusiasts*, who focus mainly on informational support, both seeking and providing. In the previous study (Wang et al 2014), we used two metrics--productivity (i.e., a user's total number of posts) and life span (i.e., the number of days between a user's first and last post)--to investigate how users in diverse groups engaged differently in the OHC. It was found that Companionship Builders and All-around Contributors are the most productive members and stayed within the community for the longest time. By contrast, those who mainly seek support (informational or emotional) published fewer posts than others and involved only a short period.

Social Support	All users	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
COM	0.1126	0.0042	0.6492	0.1271	0.0154	0.0504	0.0408	0.0404
PES	0.1178	0.0074	0.0833	0.1511	0.0053	0.612	0.0315	0.0351
PIS	0.4422	0.9655	0.1277	0.4762	0.0152	0.2394	0.4369	0.0325
SES	0.0743	0.0067	0.0349	0.1245	0.0107	0.0481	0.0494	0.5868
SIS	0.2531	0.0162	0.1049	0.1211	0.9534	0.0501	0.4414	0.3052
# of users	47581	6647	3923	15336	3502	3994	13225	954
% of users		14%	8%	32%	7%	8%	28%	2%

Table 4. Centroids of user profiling clusters

However, results in Table 4 represent a static snapshot of all the users based on their aggregated activities over the 11 years of data in our dataset. In other words, each user was assigned to a cluster based on all her posting behaviors from the first post to the last, no matter how long the user was active in the OHC. To capture the evolution of users' roles over time, we analyzed each user's social support activities on monthly basis. For example, if a user were active for less than a month in this OHC, she would have only one social support activity profile, each of them being a 1×5 vector that represents her activities in the five social support categories in that month. Another user, who continuously contributed to this community for two years, would have 24 such profiles.

To identify which role among the seven roles in Table 4 a user played in a specific month, we adopted a simple Nearest Neighbor classification scheme. Given a user's and a month in which she was active in posting, we assigned to the user the role, whose centroids in Table 4 is closest to the user's social support activity profile in that specific month. Among 384,423 monthly social support activity profiles from 47,581 users, users' monthly roles distributions are shown as Table 5. Note that some monthly social support activity profiles were labeled as "Inactive in Posting Social Support", which indicated all the elements of the 1×5 vector were 0. Such a profile indicates two possibilities: one is that the corresponding user did not post anything during that month; the other one is that the corresponding user posted something other than social support during that month.

User Role	Number of monthly social support activity profiles
Information Seeker (IS)	9,680
Information Provider (EP)	26,711
Emotional Support Seeker (ES)	3,126
Emotional Support Provider (EP)	19,125
Community Builder (CB)	36,759
All-around Contributors (AC)	62,879
Information Enthusiasts (IE)	31,322
Inactive user in Posting Social Support (IA)	194,821
Total	384,423

Table 5. User Roles summarized by monthly activities

Based on each user's role in each month of her activities, we draw a role evolution network to show how users shift from one role to another in Fig 1. This weighted and directed network has 10 nodes. Seven of them correspond to the seven user roles we discovered in Table 4, including Information Seeker, Information Provider, Emotional Support Seeker, Emotional Support Provider, Companionship Builder, All-around Contributor, and Information Enthusiast. There is one node for the status of Inactive in Posting Social Support. We also included two nodes, Registration (REG) and Churn (CHU), to represent the starting and ending points of one's activities in this OHC. In this OHC, we assumed that a user had churned from this OHC if she had no post during the last 12 weeks in our dataset.

A directed link from node A to node B means that at least one user evolved from role/stage A to B after a month at A. The weight of each link was computed as the probability of transiting from one node to another (detail numbers are shown in Table 6). The width of ties in Fig 1 is proportional with such probabilities. For example, the wide tie from IS to CHU showed when a user acted as an Information Seeker in this OHC, there would be a high probability for her to churn from the community next month. By contrast, the thin tie from IS to IE shows that there would be a low chance for an Information Seeker to become an Information Enthusiast. Therefore, the sum of all edges' weights starting from one node was 1. Basically, all users started from Registration, but might end at any other node in this graph as time goes by.

	IP	IS	EP	ES	CB	AC	IE	IA	CHU
REG	0.1639	0.0969	0.0969	0.0269	0.0857	0.2305	0.2989	0.0132	0
IP	0.0132	0.0010	0.0273	0.0035	0.0393	0.1361	0.0576	0.2860	0.3359
IS	0.0345	0.0426	0.0143	0.0066	0.0358	0.0383	0.0708	0.2681	0.4884
EP	0.0375	0.0070	0.2047	0.0045	0.1165	0.1529	0.0276	0.2253	0.2235
ES	0.0398	0.0254	0.0224	0.0150	0.0727	0.0690	0.0636	0.2657	0.4259
CB	0.0299	0.0089	0.0576	0.0059	0.4242	0.1443	0.0252	0.1794	0.1242
AC	0.0724	0.0075	0.0572	0.0045	0.1044	0.4746	0.0549	0.1296	0.0944
IE	0.0642	0.0299	0.0217	0.0080	0.0426	0.1373	0.1185	0.2394	0.3379
IA	0.0392	0.0130	0.0215	0.0042	0.0308	0.0346	0.0332	0.8032	0.0198

Table 6. Probabilities of users' roles transitions. (The first column represents the starting role in a transition, while the first row represents the ending role.)

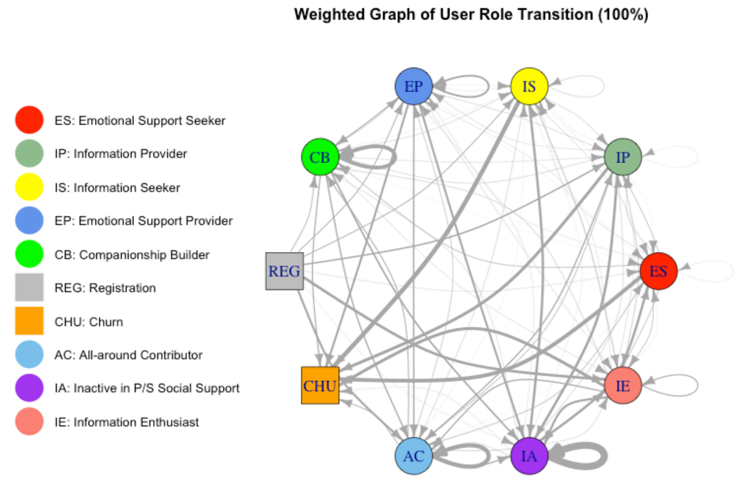
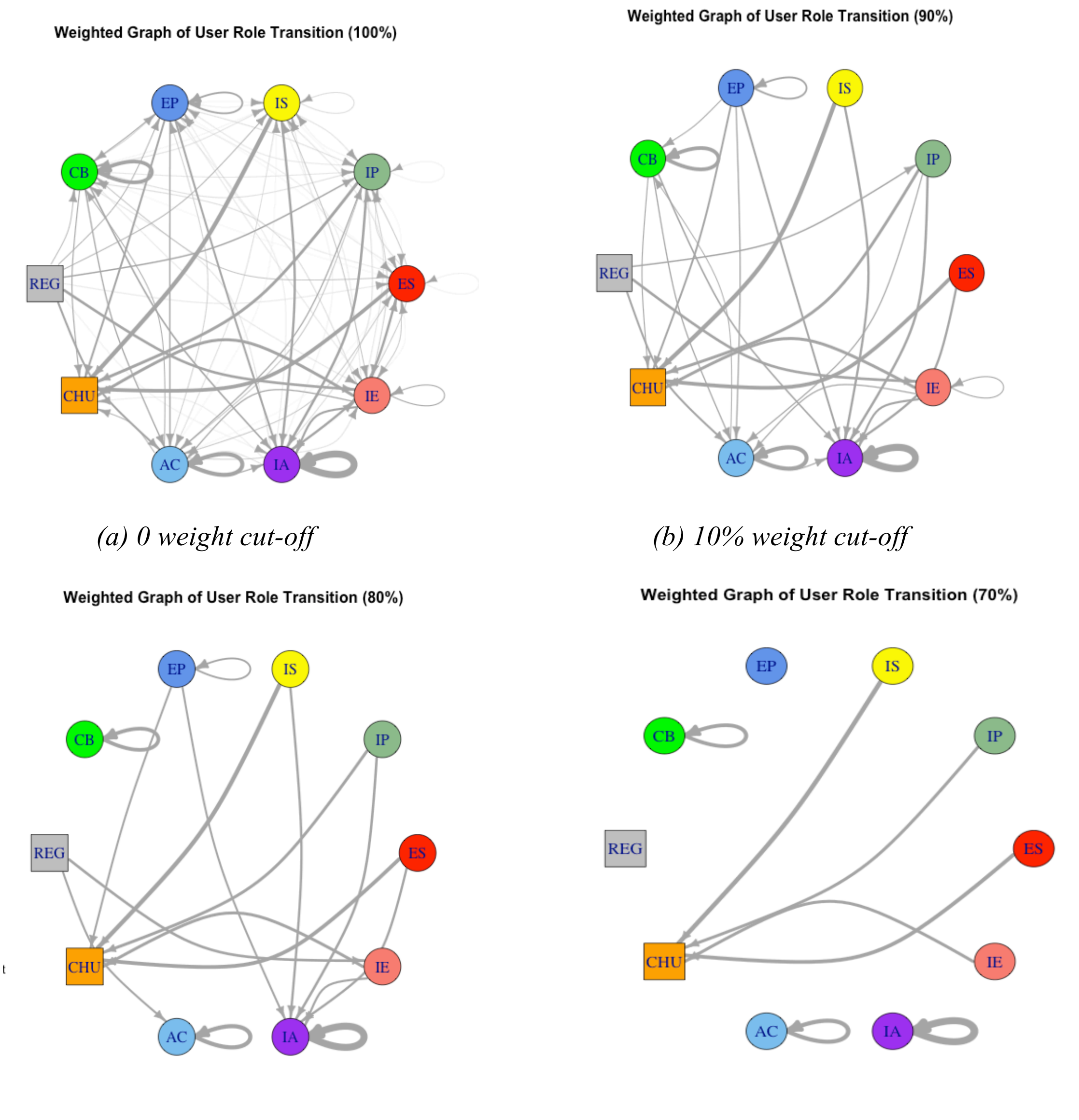


Figure 1. Evolution of user role in this OHC

To find pattern in user role evolution, we generated a series of weighted graphs at increasing cut-off weights from 0 to 30% (shown in Figure 2). Specially, when 0 is the threshold, we show all the links in the evolution network, while the 90% graph only shows the most frequent transitions whose probabilities are over 0.1. Similarly, from Figure 2 (c), we can see all the links with a probability higher than 0.2.

From these evolution networks, we can find some interesting trajectories of user role evolution. First, after registration, users had the highest probability to become Information Enthusiasts and All-around Contributors in this OHC. Different from being Information Enthusiasts, users who focused only on seeking or only on providing social support (EP, IS, IP and ES in the graph) had a higher probability to churn. This is consistent with our previous findings: users who mainly seek support (informational or emotional) published fewer posts than others and involved only a short period. Meanwhile, both support seekers and providers, even Information Enthusiasts, can easily switch to an idle stage--contributing no social support to this community in a month. By contrast, the roles of Companionship Builders and All-around Contributors are relatively stable in the community. Because a previous study of us (Wang et al 2014) suggested that users in these two roles are the most productive members and tend to stay within the community for longer time, the high probability for users to stay in the two roles is good news for the OHC.

Since a well-organized OHC should motivate users' active participations, the more active users posted in the OHC, the more social support would be provided, which can benefit more people. Therefore, how to keep users engaged in is an interesting question to study using this evolution network. In other words, we want to find the signal of users becoming inactive in posting or preparing to leave. In the role evolution network, we can see that some nodes have high probability to transit to CHU (churn). Then what could lead to this transition? What about other high probability transitions to the IA node (Inactive in providing or seeking social support)? We summarized transitions with probabilities higher than 0.2 as an example in Figure 2 (c) and analyzed users social support activities preceding undesirable transitions to CHU and IA.



(a) 0 weight cut-off

(b) 10% weight cut-off

(c) 20% weight cut-off

(d) 30% weight cut-off

Figure 2. Evolution of user role filtered by weight in this OHC

We used logistic regression to reveal factors related to high probability transitions in the evolution network, because shifting from one type of role to another type is a binary choice--the answer could be either “yes” or “no”. Using “transition or not” as the dependent variable, we summarized five independent variables in Table 7. Basically, we want to see if a user’s behaviors in receiving social support in the previous month would modify her role’s shift in this OHC. Therefore, we collected data of receiving information and emotional support both directly and indirectly, namely RISD, RESD, RISI and RESI. In addition, we included a user’s exposure to companionship (RCOM). Detailed definitions of these variables can be found in Table 7. Values of these variables were based on users’ activities within the months preceding their role transitions. Specifically, when a transition occurs at *kth* month, we analyzed data of *(k-1)th* month to see what variables are correlated with the transition.

The results were listed in Table 8. The first column shows the starting node of the transitions and the second row shows the ending node of the transitions. Coefficient estimates showed the relationship between receiving social support and role transitions. For example, when we hold all other independent variables constant, an Emotional Support Provider (EP) directly receiving each additional information support post from the others would lead to 0.031 decrease in the log-odds of her dropping out of the OHC next month.

Indep. Variables	Descriptions
RISD	Direct informational support received--the number of informational support posts a user received after initiating a support-seeking thread.
RESD	Direct emotional support received--the number of emotional support posts a user received after initiating a support-seeking thread.
RISI	Indirect informational support received--the number of informational support posts a user was exposed to in threads that she/he did not initiate but contributed to.
RESI	Indirect emotional support received--the number of emotional support posts a user was exposed to in threads that she/he did not initiate but contributed to.
RCOM	Companionship received--the number of companionship posts a user was exposed to in threads that she/he did not initiate but contributed to.
Note: for RISI, RESI, and RCOM, we assumed that a user read others' replies that were posted within 7 days before the user's replies.	

Table 7 Independent variables for binary logistic regression

	Indep. Variables	Regression Coefficient (B)	
E / S		CHU (User churn)	IA (Inactive in P/S Social Support)
EP	RISD	-0.031***	0.000
	RESD	0.000	-0.009
	RISI	-0.021***	-0.007***
	RESI	-0.044***	-0.011***
	RCOM	-0.010***	-0.006***
IS	RISD	-0.169***	-0.023***
	RESD	0.071***	0.016
	RISI	-0.078***	-0.002
	RESI	0.030	-0.037***
	RCOM	-0.050***	0.012
IP	RISD	-0.066***	-0.013**
	RESD	0.026	0.023*
	RISI	-0.067***	-0.008***
	RESI	0.006	-0.008**
	RCOM	-0.004	-0.001
ES	RISD	-0.128***	-0.007
	RESD	0.012	-0.005
	RISI	-0.029***	-0.005
	RESI	-0.057***	-0.015*
	RCOM	-0.016	0.002
IE	RISD	--	-0.049***
	RESD	--	0.023***
	RISI	--	-0.008***
	RESI	--	-0.007**
	RCOM	--	-0.003

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

Table 8 Results of binary logistic regression on frequent user role transitions. High probability (over 0.2) transitions start from a node/stage in first column and end at CHU or IA. For each transition, a number shows the regression coefficient of one of five independent variables and this transition.

As we mentioned earlier, those in the roles of AC and CB tend to stay in the same roles. Previously, we concluded All-around Contributors (AC) are the largest group in this OHC (32%), and companionship provided by Companionship Builders (CB) could help users participate together. Therefore, keeping those already in AC and CB to stay there would be very useful to build a successful OHC. To find what are behind the stability for the two roles, we applied logistic regression on self-loops for the two roles as well. The results are shown in Table 9. Similarly, we did find the correlation between receiving social support and self-transitioning. For instance, receiving informational support was significantly helpful for a user to keep acting as an All-around Contributor (AC).

AC →AC		CB → CB	
Indep. Variables	Regression Coefficient (B)	Indep. Variables	Regression Coefficient (B)
RISD	0.004***	RISD	0.001
RESD	-0.002	RESD	0.016***
RISI	0.007***	RISI	-0.002***
RESI	0.002***	RESI	0.002***
RCOM	-0.002***	RCOM	0.005***

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

Table 9 Results of binary logistic regression on frequent user role stabilizations

4 DISCUSSION

It is intuitive that, when users became inactive, it is a strong signal that they are leaving. Nevertheless, there is no link starting form inactive users to churn (IA to CHU) in Figure 2. This is because in our dataset, we summarized a user's behavior from her first post to the last post. As a result, a user's last social support profile in the dataset must contain some of her posting information. In other words, a user who churned from the OHC must have a role related to social support. Lurking behaviors are not considered because such data is unavailable for our study.

Based on statistics in Table 8, even if a user started from the same role, differences in receiving social support would lead to different role transitions. Specifically, receiving more informational support directly could effectively decrease churning probability of an Emotional Support Provider, but it cannot guarantee the user continuously providing emotional support to the others. Similarly, the exposure to companionship could keep an Information Seeker staying in this OHC, but helped less to insure her contributing social support successively. By contrast, receiving emotional support indirectly can help both Information Seekers and Information Providers staying productive during the next month. As the smallest group of users in this OHC, Emotional Support Seekers have a high probability to drop out of the community, but reading social support posts in the others' threads or getting informational replies directly from the others could significantly decrease this likelihood. In addition, for Information Enthusiasts, receiving social support from the others essentially decided their levels of activities in the following month. However, a new interesting finding is that some users who were not active in posting social support content would come back to the community after a long time of absence.

Generally speaking, reading others' posts or receiving support directly might effectively decrease the chance of a user to leave or to be inactive in posting social support related posts in the future. However, an interesting observation is that receiving some social support may negatively affect users' engagement in this OHC, and lead them to churn or be inactive in posting. For example, the positive coefficient for RESD in the IS->CHU transition in Table 8 suggested that getting too much emotional support might lead to an Information Seeker's departure. It makes sense because for a user who sought information support, emotional support did not satisfy her needs, which might have contributed to her churn. A similar situation also applies to Information Enthusiasts. Therefore, receiving the social support that one is seeking promptly seems to be the secret to keep users away from churn and inactivity.

Last but not least, what can help All-around Contributors and Companionship Builders to stay in their roles? Intuitively, except reading emotional support directly from the others, all receiving behaviors are correlated with the self-transitions of these two roles. However, the exposure to companionship is not helpful for them to maintain their roles. One possible explanation is that All-around Contributors kept posting all kinds of social support in a balanced way. Reading too much companionship posts might lead them to post more companionship posts and then shift to Companionship Builders. Different from All-around Contributors, the exposure to companionship posts is positively correlated with Companionship Builders' role stability.

Our analysis on the pattern of user role transitions in an OHC could also have implications for the management of OHCs. According to a study by Young (2013), a mature OHC holds the largest size of active members. Maintaining a sustainable active member base can effectively avoid the "death" of the OHC. According to our previous studies, Community Builders and All-around Contributors are those who are much more active than the others. Also, senior members' experience, knowledge, and community network could be precious resource for newcomers. Therefore, encouraging users to stay at or transit to these two roles is helpful to maintain a sustainable OHC.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we used both supervised and unsupervised machine learning techniques to detect the role of users with regard to social support activities in an OHC. To address research questions we proposed in Section 1, we found users in this OHC can be grouped into 8 different types of contributors: Information Seeker, Information Provider, Emotional Support Seeker, Emotional Support Provider, All-around Contributor, Companionship Builder, Information Enthusiast and Inactive users in seeking/providing social support. Based on analyzing their monthly social support behaviors, we found that users' roles do evolve over time. We also included user's registration and churning behavior as two extra nodes to construct the network of user role evolution. Meanwhile, we found some role transitions are with higher probabilities than others, such as from an Information Seeker to Churn. Logistic regression models uncovered that receiving certain types of social support are correlated with high frequent transitions that are of interests for community managers.

This paper only showed preliminary findings on this interesting topic. There are many directions for future work. First, we want to build a dynamic model to describe user role evolutions in this OHC. In this paper, we only considered the probability of transitions without paying attention to the time of the transitions occurred. Thus when a transition will happen may be an interesting question to explore. In addition, we have already showed how the type of social support a user received is related to her role transitions. Then can we recommend some support posts to these users to facilitate or delay a role transition? Thus, designing a good recommender system based on social support is also a possible studying project in the future.

References

Altman, DG (1991). Practical statistics for medical research. London: Chapman and Hall.
Bambina, A. (2007). Online social support: the interplay of social networks and computer-mediated communication. Cambria Press, Youngstown, N.Y.

Barak, A., Boniel-Nissim, M., & Suler, J. (2008). Fostering empowerment in online support groups. Computers in Human Behavior, 24(5), 1867–1883. http://doi.org/10.1016/j.chb.2008.02.004
Chou, W. S., Hunt, Y. M., Beckjord, E. B., Moser, R. P., & Hesse, B. W. (2009). Social Media Use in the United States: Implications for Health Communication. Journal of Medical Internet Research, 11(4). http://doi.org/10.2196/jmir.1249
Dunkel-Schetter, C. (1984). Social Support and Cancer: Findings Based on Patient Interviews and Their Implications. Journal of Social Issues, 40(4), 77–98.
Ferguson, Tom. (1996). Health Online. Reading, MA: Addison-Wesley Company.
Fox, S., & Duggan, M. (n.d.). Health Online 2013. Retrieved from http://www.pewinternet.org/2013/01/15/health-online-2013/
Füller, J., Hutter, K., Hautz, J., & Matzler, K. (2014). User Roles and Contributions in Innovation-Contest Communities. Journal of Management Information Systems, 31(1), 273–308. http://doi.org/10.2753/MIS0742-1222310111
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. http://doi.org/10.1080/10410230701697100
House, J.S. (1981). Work stress and social support. Reading. Addison-Wesley, MA.
Keating, D. M. (2013). Spirituality and Support: A Descriptive Analysis of Online Social Support for Depression. Journal of Religion and Health, 52(3), 1014–1028.
Kim, E., Han, J. Y., Moon, T. J., Shaw, B., Shah, D. V., McTavish, F. M., & Gustafson, D. H. (2012). The process and effect of supportive message expression and reception in online breast cancer support groups. Psycho-Oncology, 21(5), 531–540. http://doi.org/10.1002/pon.1942
Krause, N. (1986). Social Support, Stress, and Well-Being Among Older Adults. Journal of Gerontology, 41(4), 512–519.
Lieberman, M. (2007). The role of insightful disclosure in outcomes for women in peer-directed breast cancer groups: a replication study. Psycho-Oncology, 16(10). Retrieved from http://escholarship.org/uc/item/8rw0v48t
Loane, S. S., & D'Alessandro, S. (2013). Communication That Changes Lives: Social Support Within an Online Health Community for ALS. Communication Quarterly, 61(2), 236–251. http://doi.org/10.1080/01463373.2012.752397
McClellan, W. M., Stanwyck, D. J., & Anson, C. A. (1993). Social support and subsequent mortality among patients with end-stage renal disease. Journal of the American Society of Nephrology, 4(4), 1028–1034.
Pfeil, U., & Zaphiris, P. (2007). Patterns of Empathy in Online Communication. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 919–928). New York, NY, USA: ACM. http://doi.org/10.1145/1240624.1240763
Preece, J., & Shneiderman, B. (2009). The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation. AIS Transactions on Human-Computer Interaction, 1(1), 13–32.
Qiu, B., Zhao, K., Mitra, P., Wu, D., Caragea, C., Yen, J., … Portier, K. (2011). Get Online Support, Feel Better – Sentiment Analysis and Dynamics in an Online Cancer Survivor Community. In 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom) (pp. 274–281). http://doi.org/10.1109/PASSAT/SocialCom.2011.127
Shumaker, S. A., and Brownell, A. 1984. "Toward a theory of social support: Closing conceptual gaps," Journal of Social Issues (40:4), pp. 11–36 (doi: 10.1111/j.1540-4560.1984.tb01105.x).
Wang, X., Zhao, K., & Street, N. (2014). Social Support and User Engagement in Online Health Communities. In X. Zheng, D. Zeng, H. Chen, Y. Zhang, C. Xing, & D. B. Neill (Eds.), Smart Health (pp. 97–110). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-08416-9_10
Wang, Y.-C., Kraut, R., & Levine, J. M. (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 (pp. 833–842). New York, NY, USA: ACM. http://doi.org/10.1145/2145204.2145329
Young, C. (2013). Community Management That Works: How to Build and Sustain a Thriving Online Health Community. Journal of Medical Internet Research, 15(6). http://doi.org/10.2196/jmir.2501
Zhao, K., Yen, J., Greer, G., Qiu, B., Mitra, P., & Portier, K. (2014). Finding influential users of online health communities: a new metric based on sentiment influence. Journal of the American Medical Informatics Association: JAMIA, 21(e2), e212–218. http://doi.org/10.1136/amiajnl-2013-002282