# User-Created Groups in Health Forums: What Makes Them Special?

**V. G. Vinod Vydiswaran[1]*, Yang Liu[1]*, Kai Zheng[2,1], David A. Hanauer[3,1], Qiaozhu Mei[1,4]**
[1]School of Information; [2]School of Public Health, Department of Health Management and Policy;
[3]Department of Pediatrics; [4]Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor
{vgvinodv, owenliu, kzheng, hanauer, qmei}@umich.edu

**Abstract**

Patients and caregivers often seek help and support through online forums for their ailments. In many health forums, users can not only participate in existing, site-defined user groups but also initiate their own groups. To better understand the mechanism and usefulness of such user-created groups, we comparatively studied their characteristics against those of site-defined groups, based on an empirical dataset of over three years of user activities in health forums collected from MedHelp.org. In the study, we first derived a categorization of user-created groups according to the underlying reasons leading to their creation, and then investigated if there are significant differences between user-created and site-defined groups with respect to group memberships, activity levels, social network metrics, and interaction patterns. The results show that users initiate more homophily driven communities and social interactions, and that the members of user-created groups are more vocal and more socially active than those who have only participated in site-defined groups. The findings suggest that allowing users to create their own groups could improve user engagement in health forums, foster rapport among users, and ultimately lead to a cohesive social environment that supports more effective information sharing and community building.

## 1 Introduction

The Internet has increasingly become an important source of health information for patients and caregivers. According to a poll conducted in August 2011 (Harris-Interactive 2011), 39% of the U.S. adults online reported looking for health information "often" on the Internet. In another survey (Fox 2011), 18% of Internet users said they go online to connect with others with similar health concerns, and this fraction increases to 23% among those who have chronic conditions such as high blood pressure, diabetes, and cancer.

Using social media to acquire and disseminate health-related information has become common especially among the younger generation (Chou et al. 2009). Health forums are popular venues frequented by patients and caregivers seeking information and support. For example, in a 2010 poll (Capstrat 2010), 37% of Internet users rated online health forums as somewhat or extremely reliable sources of health information. Research studies have shown that these forums play an instrumental role in facilitating exchanges of health information and/or emotional support (Wang, Kraut, and Levine 2012; Chou et al. 2009) and that their use is associated with higher degrees of patient empowerment (van Uden-Kraan et al. 2009).

In the most prevalent scenario, owners/designers of a health forum provide a predefined list of user communities, often organized around certain medical conditions, ailments, or treatment procedures. Users can join, post questions, and respond to others' questions in these communities. In addition to these site-defined communities, some health forums also allow users to form new communities of their own, on the fly, which creates a whole new paradigm of how users of health forums bond and communicate.

To better understand the motivation and merits of such user-initiated communities (hereafter referred to as "user-created groups"),[1] we comparatively examined their characteristics against those of "site-defined groups." Based on an empirical dataset collected from MedHelp.org, a premier online health forum, we studied the attributes of user-created groups, including reasons leading to their creation, and how user-created groups differ from site-defined groups along multiple dimensions including membership, activity levels, social network metrics, and interaction patterns.

## 2 Data Description

Established in 1994, MedHelp (sometimes referred to as "the Website" in this paper) is one of the earliest and most well-known online forums dedicated to supporting user-driven discussions on health or healthcare related topics. The Website had over 12 million registered users when this work was conducted. It has been the subject of study in several prior research endeavors (Gill and Whisnant 2012; Chuang and Yang 2012a; Hagan 3rd and Kutryb 2009).

In February 2013, we collected a complete set, i.e. all forum posts and all user profiles, from the Website. The dataset consists of over six million messages (both questions and

*Both authors contributed equally.



[1]Note that user communities in online health forums may be called forums, boards, user groups, etc. For simplicity, in this paper, we refer to them uniformly as "groups."

comments in response to questions) posted by over a million unique users in about 1.4 million threads. From the user profiles collected, we extracted the friendship links among the users as well as their group membership if they had explicitly joined certain user groups.

There are five distinct types of user communities in MedHelp: (1) medical support communities, (2) "ask-a-doctor" forums, (3) forums on pets, (4) international forums, and (5) user groups. In this paper, we focus on the medical support communities and the user groups, both of which aim to facilitate interactions among patients and caregivers. The medical support communities are designed and provided by MedHelp (hence "site-defined groups"); whereas the user groups are support communities initiated by end users ("user-created groups").

The dataset analyzed in this study includes all posts in a total of 270 site-defined groups and 747 user-created groups, in addition to the profiles of 1,007,570 users who had posted at least one message in these groups. Among these users, 9,544 were members of at least one user-created group and 502,269 were members of at least one site-defined group. A total of 130,605 friendship links existed among all users, out of which 113,273 (86.7%) are between users in our dataset.

## 3 Categorizing user communities by purpose

The first question we are interested in investigating is why users choose to create so many groups even though the Website has already provided a comprehensive list of well attended groups. In order to understand the potential motives, we conducted a qualitative content analysis of the data to categorize the user-created groups based on their stated purposes. Two authors first individually analyzed a random sample of a hundred user-created groups to derive a primitive list of categories according to the descriptions of the groups. Then, through a consensus development process, the two lists were reconciled and merged to produce a final categorization scheme. The analysis resulted in ten categories, as listed below.

1. **Specific conditions (COND)**: Communities related to particular conditions, ailments, or diseases. This includes addictions or rare diseases that may not have an established cure or understanding. Examples include "Arachnoiditis sufferers," "Granulomatous mastitis," and "Vertigo."
2. **Specific treatment (TRMT)**: Communities related to particular treatments and procedures, including conditions that arise from a specific treatment or procedure. Examples include "Mirena IUD side effects support group," "Methadone community," and "Natural health" (a group about alternative medicines).
3. **Recovery (RCVY)**: Communities related to the process of recovering after a completed treatment regimen, including addiction recovery, smoking cessation, etc. Examples include "Addiction recovery group," "Recovery after vitamin D deficiency," and "Heart surgery recovery."
4. **Family support (FAM)**: Communities related specifically to family members or caregivers of patients or others suffering from specific conditions. Examples include "ADHD parents," "Alzheimer's caregivers," and "Family members of prisoner."
5. **Socializing (SOC)**: Communities created primarily to host social interactions, mostly in a non-medical context, such as chit-chats, specific interests or hobbies, discussions of current events, etc. Examples include "Prayer group," "All about TV shows and movies," and "Dinner table."
6. **Public policy (POL)**: Communities related to governmental agencies, the economy, or public policies, including healthcare and insurance. Examples include "Social security disability or SSDI," "Ideas for economic living," and "FDA recalls, US food and drug administration."
7. **Pregnancy (PREG)**: Communities broadly related to pregnancy, including attempting to conceive, conditions and complications during pregnancy, and post-pregnancy care for the mothers and babies. Examples include "Trying to conceive after 40," "Pregnancy after tubal ligation surgery," and "March 2011 babies."
8. **Goal-oriented (GOAL)**: Communities related to specific health-related goals, including weight loss, healthy diet, etc. This category also includes communities where members could track each other's health or behavior change progress. Examples include "HCG protocol group," "Diet ideas," "Weight gain," and "The 10% club."
9. **Specific demographics (DEM)**: Communities that target towards specific demographic groups. This category is further classified into the following five subcategories:
   (a) *Gender*: Communities targeted towards users of a specific gender. Examples include "Boy problems" and "Christian women with bipolar disorder."
   (b) *Location*: Communities targeted towards users in a specific location. Examples include "California" and "Problems with children and young people services in the UK."
   (c) *Age*: Communities targeted towards users of a specific age group. Examples include "TTC over 40" (TTC is a commonly used abbreviation in health forums for trying to conceive) and "ADHD teens and young adults."
   (d) *Profession*: Communities targeted towards users of a specific profession. Examples include "College students" and "The doctors."
   (e) *Others*: Communities targeted towards users of a particular demographics not included above, such as marital status (e.g. "Mothers and the balance of a stressful life," a group for single working mothers), race (e.g. "Native American / Canadian circle"), etc.
10. **Miscellaneous (MISC)**: Communities that do not fall in any of the above categories, including the ones with an unclear purpose (missing or vaguely described), or those advocating for a particular business organization (e.g. a specific law firm).

Note that the categories described above are not mutually exclusive and therefore a community could be classified under multiple categories. Overall, 79 (10.6%) user-created

| Category | Site-defined $n = 270$ | User-created $n = 747$ |
|---|---|---|
| 1. Specific conditions | 173 (64.1%) | 260 (34.8%) |
| 2. Specific treatment | 15 (5.6%) | 53 (7.1%) |
| 5. Socializing | 16 (5.9%) | 150 (20.1%) |
| 7. Pregnancy | 40 (14.8%) | 122 (16.3%) |
| 9. Demographics | 20 (7.4%) | 81 (10.8%) |
| 10. Miscellaneous | 3 (1.1%) | 77 (10.3%) |
| Other categories (3,4,6,8) | 17 (6.3%) | 93 (12.5%) |

Table 1: Distributions of the categories of site-defined and user-created groups. Note that the columns do not add to 100% since groups may have multiple labels.

groups and 15 (5.6%) site-defined groups were classified into more than one category. Nonetheless, as all pregnancy-related communities were likely gender-specific by nature, we did not include them again in the "Specific demographics / Gender" category. Communities with non-English titles and descriptions were all classified as "Miscellaneous."

**Inter-rater agreement** The inter-rater agreement is high between the two authors who were involved in classifying the user groups. For user-created groups, the two-rater, ten-category Cohen's kappa coefficient is 0.745. Out of the 747 user-created groups, the two raters agreed on the categorization of 600 of them (80.3%). For site-defined groups, the inter-rater agreement is even higher, with the Cohen's $\kappa = 0.854$.

### 3.1 Category distributions

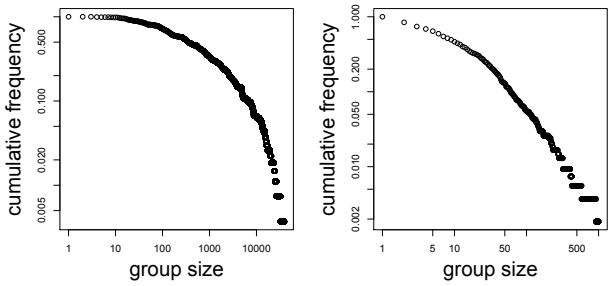
Table 1 summarizes the distributions of the ten categories of site-defined and user-created groups as defined in Sec. 3. Among site-defined groups, the two most frequent categories are those related to "specific conditions" (class 1, ∼64%) and "pregnancy" (class 7, ∼15%). We also note that only 1.1% of the site-defined groups are categorized as "miscellaneous." This also suggests that the classification scheme works well for site-defined groups, although it is derived based on user-created groups.

The category distribution of user-created groups presents some interesting difference from the site-defined groups. As in site-defined groups, the most salient category of user-created groups continues to be "specific conditions" (class 1), but the percentage drops to about 35%.

The second most popular category is related to "socializing" (class 5), constituting 20% of user-created groups. This indicates a significant use case of user-created groups, i.e., to find "friends" and engage in casual conversations about things that may or may not relate to health, rather than purely seeking information about their medical conditions and treatments. Indeed, we notice that typically, such groups are related to hobbies, discussion around current news, religion (prayer groups), or recreational activities (such as a group on creative writing), which would take one's mind away from the pain and suffering.

Pregnancy-related groups constitute ∼16% of user-created groups, forming the third largest category. As we see from Table 1, the proportion is similar to those related to pregnancy in site-defined groups. Our analysis shows that user-created groups related to pregnancy focus more on particular demographics such as age, specific pregnancy-related complications, or questions on parenting and childcare.

About 11% of the user-created groups in our data set are created for specific demographic classes. Many of these groups (∼69%) are also related to a specific condition, pregnancy, or for socializing. For example, there are groups for patients from a specific age-group and location (e.g. a group intended for diabetic teens in Michigan), or for teenagers who suffer from autism but want to socialize and share their experience with each other. These interesting surface statistics motivate us to dig deeper into the actual reasons why the users create groups.




(a) Site-defined groups (b) User-created groups

Figure 1: Cumulative distributions of sizes of site-defined and user-created groups. User-defined groups present a power-law distribution with an exponent of 2.04.

## 4 User-created groups vs. site-defined groups

In this section, we present a detailed comparison of the site-defined groups and the user-created groups. We compare them with respect to membership statistics of the groups, properties of the social network of users, and the forum activity of users.

### 4.1 Analysis based on group membership

Our data set consists of 270 site-defined groups and 747 user-created groups. The median number of members of a user-created group is 5. Figure 1 shows the cumulative distribution of size of the site-defined and user-created groups, measured by the number of members. As the figure shows, the sizes of user-created groups follow a power-law distribution with an exponent of 2.04. Most of these groups have only a small number of members, while very few groups have many members. Indeed, 26.5% of the user-created groups have only one single member (the founder of the group). In comparison, the sizes of site-defined groups do not follow a power-law distribution. The median number of members of a site-defined group is 354.5. Half of the site-defined groups have a moderate number of members between 100 and 2000, and there are relatively few groups with either very high or very low membership.

Out of the 747 user-created groups, 50 (6.7%) groups are marked as private groups. Analyzing the descriptions of all private groups, we find that the main reason for marking a group as private is to restrict membership to specific age groups (such as adults or teenagers), and that 43% of the private user-created groups are created for the purpose of socializing (cf. Sec. 3, class 5).

To further study the membership patterns, we group the users into seven categories, based on their roles and memberships in site-defined and user-created groups. The seven categories, along with the number of users per category, are listed below.

(a) **user-created group members (UCG MEM)**: users who are members of one or more user-created groups ($n = 9,544$);

(b) **site-defined group members (SDG MEM)**: users who are members of one or more site-defined groups ($n = 502,269$);

(c) **strict user-created group members (UCG ONLY)**: users who are members of user-created groups only ($n = 2,738$);

(d) **strict site-defined group members (SDG ONLY)**: users who are members of site-defined groups only ($n = 495,463$);

(e) **both**: users who are members of at least one user-created group and one site-defined group ($n = 6,806$);

(f) **founders**: users who initiated one or more of the user-created groups ($n = 553$); and

(g) **non-members**: users who are not members of either a user-created group or a site-defined group ($n = 502,577$).

Out of the 1,007,570 users in our dataset, 502,577 (49.88%) users did not join any group. 49.2% users joined one or more site-defined groups. Although there are three times as many user-created groups on the Website, only 9,544 (less than 1%) users actually joined at least one of them. On the contrary, it is interesting to note that most of the users who joined user-created groups (71.3%) also joined one or more site-defined groups. There are 553 users who initiated the 747 user-created groups.

## 4.2 Analysis of friendship networks

Next, we analyze the properties of the user friendship network to better understand the social network characteristics of users participating in health forums. In the friendship network, nodes represent users who have posted at least one message in any forum, and two nodes are connected via an edge if the users are identified as friends in their respective profile pages. All network analyses are performed using the igraph library (Csardi and Nepusz 2006).

There are 1,007,570 users in our data set, of which only 30,915 (3.1%) have friendship links. The entire friendship network consists of 30,915 nodes and 113,273 edges. This social network presents a giant component of 30,870 nodes and a disconnected component of 45 nodes. Figure 2 shows the cumulative degree distribution of the global network as a log-log plot. The degree distribution follows a power-law with an exponent of 2.12.

The network does not show an assortative mixing pattern. The assortativity coefficient, which is defined by the Pearson correlation coefficient of the degrees of pairs of linked

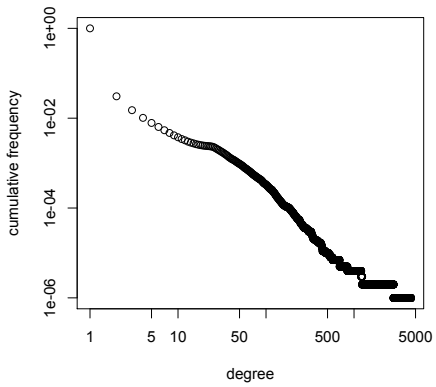



Figure 2: Cumulative degree distribution of the global MedHelp network follows power-law with an exponent of 2.12.

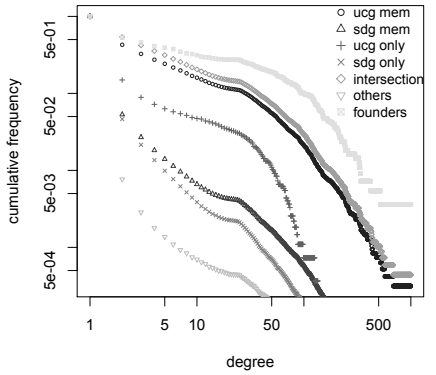



Figure 3: Cumulative degree distributions of users in different groups

nodes, is −0.075. The average shortest path length of the network is 3.81. This suggests that the connected components are still small worlds. We also compute the clustering coefficient of the network. The clustering coefficient, also known as transitivity, is defined as the ratio of the triangles and the connected triples in the network (Watts and Strogatz 1998). The friendship network has a clustering coefficient of 0.031, which is much smaller than that of real world social networks, but much larger than a random network of the same number of nodes and edges. This may suggest the lack of community structure in the global friendship network in MedHelp (Newman and Park 2003).

Next, we study the properties of individual users on the friendship network, grouped by the categories defined in Sec. 4.1. We compute the average degree, maximum degree, average local clustering coefficient, and the mean of two centrality measures: betweenness and closeness. The local clustering coefficient of a node is the ratio of the trian-

| User categories | UCG MEM | SDG MEM | UCG ONLY | SDG ONLY | Both | Founders | Others |
|---|---|---|---|---|---|---|---|
| Observations (n) | $(9,544)$ | $(502,269)$ | $(2,738)$ | $(495,463)$ | $(6,806)$ | $(553)$ | $(502,577)$ |
| Avg. degree | 10.17 | 0.40 | 1.87 | 0.22 | 13.51 | 41.11 | 0.04 |
| Max. degree | 4422 | 4422 | 134 | 1204 | 4422 | 4422 | 325 |
| Avg. local cc | 0.117 | 0.012 | 0.049 | 0.010 | 0.143 | 0.102 | 0.001 |
| Avg. betweenness ($\times 10^3$) | 79.80 | 2.46 | 1.87 | 0.97 | 111.16 | 667.81 | 0.20 |
| Avg. closeness ($\times 10^{-8}$) | 28.21 | 3.52 | 9.20 | 3.07 | 35.86 | 35.33 | 0.51 |

Table 2: Properties of different categories of users on the friendship network. Note: cc stands for clustering coefficient.

| | Site-defined | User-created | Sig. |
|---|---|---|---|
| Number of vertices | 2343.46 | 29.15 | *** |
| Number of edges | 458.54 | 31.36 | *** |
| Density | 0.002 | 0.040 | *** |
| Clustering coefficient | 0.088 | 0.206 | *** |
| Average shortest path | 2.65 | 1.80 | *** |
| Degree assortativity | −0.43 | −0.48 | * |
| Average degree | 0.35 | 0.63 | *** |
| Maximum degree | 53.11 | 4.57 | *** |

Table 3: Network statistics of site-defined and user-created groups. The last column shows significance levels: * for 10% and *** for 1%.

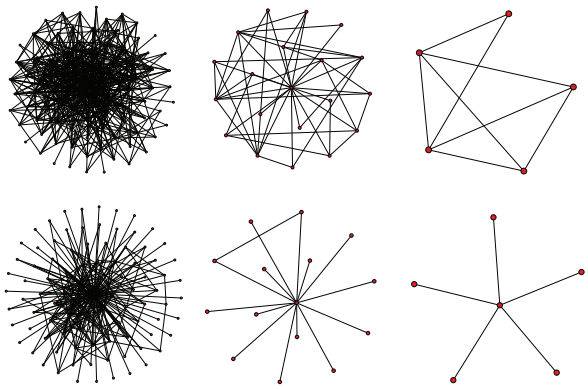


Figure 4: Sample ego networks of users with highest degree in the "pregnancy-related" category (top row) and the "specific condition" category (bottom row).

gles connected to the node and the triples centered on the node. The betweenness centrality of a node is defined by the number of shortest paths going through the node, while the closeness centrality is defined by the inverse of the average length of the shortest paths to all the other nodes in the network (Newman 2003).

The summary statistics are reported in Table 2. We see that the founders of user-created groups are the most popular users on MedHelp, yielding the highest average degree. They also yield the highest centrality measures, which means that they are also the most influential users on MedHelp. Members of user-created groups tend to make more friends and have higher centrality measures than members of site-defined groups. Their friends are also more likely to be friends with each other, suggested by a larger clustering coefficient. Members of both types of groups are even more social and more influential. We see similar results when we consider the properties of the members belonging to the "strict user-created groups" (UCG ONLY) or the "strict site-defined group" (SDG ONLY) categories.

Figure 3 shows the cumulative degree distributions of seven groups of users. From the figure, we observe that the users differ in the number of friends in all seven categories. Some of them have thousands of friends, others just visit the forum without making any friends. Founders, user-created group members, and users who joined both user-created and site-defined groups have broader distributions. We also notice that members of the "strict user-created groups" have a much steeper tail compared to those of "strict site-defined groups", which means the former has a larger proportion of members with relatively high degree.

Next, we construct a sub-network of the friendship network for each site-defined and user-created group, by extracting the members of that group and the links between them. We calculate the network metrics of these group-specific sub-networks. For each metric, we compare the mean of the values among user-created groups against that among site-defined groups. The summary statistics are reported in Table 3. As shown in the table, site-defined groups are much larger in size than user-created groups on average. However, user-created groups have a denser and more clustered friendship network than site-defined groups. This indicates that although the site-defined groups usually attract more users, the user-created groups form better communities since their members are better connected with each other. Unlike real world social networks, both groups showed disassortative mixing, as popular users tend to make friends with users with low degree.

Finally, we compared the network statistics of the ten categories of user-created groups as defined in Sec. 3. The results summarized in Table 4 show that on average, the recovery groups (class 3) have the most members, while members of the pregnancy-related groups (class 7) make more friends than members of other types of groups. Pregnancy-related groups and demography-specific groups (class 9) have the highest density and clustering coefficient in their networks. On the other hand, the networks for groups related to specific conditions (class 1), recovery (class 3), family support (class 4), and specific goals (class 8) have a low clustering coefficient, a shorter average shortest path, and a more negative assortativity. This suggests that the network struc-

| Categories | COND | TRMT | RCVY | FAM | SOC | POL | PREG | GOAL | DEM |
|---|---|---|---|---|---|---|---|---|---|
| Number of vertices | 31.16 | 22.28 | 50.08 | 17.78 | 23.87 | 21.36 | 38.85 | 27.95 | 17.90 |
| Number of edges | 8.16 | 12.88 | 24.08 | 14.83 | 15.60 | 8.36 | 88.35 | 14.90 | 22.12 |
| Density | 0.02 | 0.01 | 0.01 | 0.01 | 0.05 | 0.05 | 0.07 | 0.03 | 0.08 |
| Clustering coefficient | 0.15 | 0.15 | 0.14 | 0.09 | 0.17 | 0.20 | 0.27 | 0.13 | 0.25 |
| Average shortest path | 1.66 | 1.85 | 1.79 | 1.38 | 1.87 | 1.93 | 1.99 | 1.76 | 1.98 |
| Degree assortativity | −0.62 | −0.52 | −0.66 | −0.56 | −0.40 | −0.46 | −0.43 | −0.53 | −0.39 |

Table 4: Network statistics of user-created groups across the nine categories.

| Categories | COND | TRMT | RCVY | FAM | SOC | POL | PREG | GOAL | DEM |
|---|---|---|---|---|---|---|---|---|---|
| Number of questions | 0.66 | 0.85 | 0.74 | 0.73 | 3.17 | 0.78 | 0.92 | 0.74 | 0.67 |
| Number of comments per thread | 0.25 | 0.30 | 0.18 | 0.37 | 0.25 | 0.35 | 0.28 | 0.14 | 0.40 |
| Number of unique users per thread | 0.07 | 0.06 | 0.03 | 0.05 | 0.17 | 0.70 | 0.18 | 0.20 | 0.26 |

Table 5: Forum activity statistics of user-created groups in nine purpose categories, normalized by the group size

tures of these groups may feature a star topological pattern. Figure 4 shows sample ego networks of the most popular users in three user-created groups related to pregnancy (on the top row) and three groups related to specific conditions (on the bottom row). The figure shows that popular members in groups of specific conditions indeed act as a hub and provide their expertise to other group members. On the other hand, popular members in pregnancy groups are connected with each other, which indicates less heterogeneity in their network structures.

### 4.3 Analysis of user activity

We then compare the activity levels of members of user-created and those of site-defined groups, in terms of how the users are involved in forum discussions.

Site-defined groups initiate many more threads than user-created groups. On average, there are $4,375$ threads (questions and corresponding comments) per site-defined group, as compared to only $29$ threads per user-created group. However, on average, a question posted in a site-defined group receives $3.43$ comments, whereas a question posted in a user-created group receives $5.60$ comments. The number of unique users participating in a thread is also different between the two types of groups. On average, $3.03$ users participate in a threaded discussion in a site-defined group, whereas $3.31$ users participate in a thread in a user-created group. Both activity measures are significantly larger for user-created groups compared to site-defined groups ($p$-value $< 0.01$). This shows that discussions in user-created groups are more interactive and involve more users.

Next, we analyze the activity levels of the ten categories of user-created groups, as defined in Sec. 3. Table 5 presents the forum activity levels of groups in each category, normalized by the group sizes. We observe that, on average, members of socializing groups (class 5) initiate the largest number of threads, while a thread in public policy groups (class 6) attract the most unique users.

We then perform a similar analysis at the individual-user level and compare the activities of the seven categories of users, as defined in Sec. 4.2. The results are summarized in Table 6. Consistent with the observations based on network properties, we observe that founders of user-created groups are the most active in forum discussions. They respond to significantly more questions posted by other members. On the other hand, the average number of threads initiated by users who never joined any group is $0.94$. These users typically register on the Website, ask a question, and never come back. As long as a user joins at least one group, the engagement significantly increases. Members of user-created groups tend to be more active (i.e., initiating threads) and more interactive (i.e., replying to threads) than members of site-defined groups. Members of both user-created groups and site-defined groups are even more active and interactive. They initiate seven times more threads and participate in almost twenty times more threads, comparing to users who only join site-defined groups. Figure 5 shows the distribution of number of questions and comments posted by users in the seven categories. If we compare the distributions of a particular category of users in the two plots, we can see that founders, members of user-created groups, and members who joined both user-created and site-defined groups respond to questions more often than asking questions. In contrast, members of the site-defined groups and other members who join neither user-created nor site-defined groups have a large overlap in the two distributions, indicating that most users in these groups tend to ask questions and rarely respond to other people's questions.

## 5 Discussion

We have presented the basic statistics and a categorization of user-created groups on MedHelp, as well as the network properties and activity patterns of the user-created groups and their members. Our analysis presents interesting differences between user-created groups and site-defined groups. In this section, we provide a discussion about the implications of these results and insights about how to improve the design of online health communities.

### 5.1 Why users create new groups?

In general, we find that homophily is one of the driving factors underlying user-created groups. The primary purposes of user-created groups are to socialize and connect with

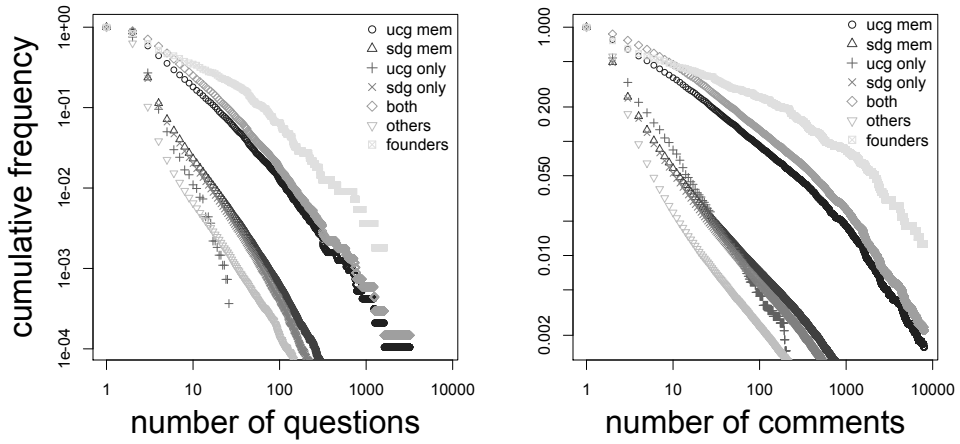



Figure 5: Cumulative distributions of number of questions and comments posted by users in different groups

| User categories | UCG MEM | SDG MEM | UCG ONLY | SDG ONLY | Both | Founders | Others |
|---|---|---|---|---|---|---|---|
| Avg. # of threads initiated per user | 9.26 | 1.85 | 1.32 | 1.70 | 12.45 | 30.03 | 0.94 |
| Avg. # of comments per user | 92.43 | 6.67 | 3.78 | 5.01 | 128.10 | 407.10 | 2.68 |
| Avg. # of threads participated per user | 62.53 | 5.55 | 2.99 | 4.44 | 86.48 | 272.20 | 2.67 |
| Avg. # of threads replied to per user | 54.00 | 3.72 | 1.88 | 2.74 | 74.97 | 244.40 | 1.73 |
| Percentage of repliers | 64.5% | 31.0% | 43.0% | 30.4% | 73.2% | 72.3% | 44.2% |

Table 6: Forum activity statistics of individual users based on group membership

other patients suffering from similar conditions, especially having similar interests or demographic profiles. Based on our analysis, we are able to identify four primary reasons why users initiate new groups given the existing ones. Users tend to create new groups

1. to form communities specific to rare and complicated conditions and new treatments;
2. to communicate with peers with similar demographics;
3. to build or maintain social relationships;
4. without checking if similar groups exist, thus creating duplicated groups.

**Reason 1. Users create communities for rare, complicated, more specific, or more general conditions:** One of the main reasons users create new groups is to connect with other patients suffering from similar conditions. These conditions are rare or complicated so that they are either not covered by the site-defined groups, covered by a site-defined group that is designed for a much more general condition, or partially covered by multiple groups. We find numerous instances where founders, diagnosed with a rare disease, reached out to others to share their symptoms or experience dealing with a particular diagnosis, and in some cases to educate others who might suffer from similar conditions.

To quantify this behavior, we analyze the text descriptions of the groups to look for motivations behind initiating a new group. We search for phrases such as "I decided to start a group . . . " or "I created this group . . . " in the group descriptions. We find 85 groups with such statements, among which the founders of 16 groups (19%) explicitly claimed that the absence of a site-defined group that is specific to their conditions was their motivation to initiate a new group.

For instance, the founder of a mental illness group with 321 members, explained the reason to form a new group as:

> "I noticed there are many groups for each illness. So I thought of an idea to have a group where we can talk about many different mental illnesses, share our stories, support each other, and hopefully make friends."[*sic*]

A specific sub-category of such user-created groups that are prevalent on MedHelp are those related to complications in pregnancy. We find multiple groups related to trying to conceive (especially with age-specific focus), and for women who are in similar stages of pregnancy. Further, there are many groups related to complications or specific conditions related to pregnancy, such as multiple births, multiple pregnancies, or pregnancy after contraception procedures, which are not the focus of any of the site-defined groups.

For instance, the founder of the group on low progesterone with 38 members stated:

> "After seeing millions of postings on MedHelp and elsewhere about women having low progesterone, many having one or more miscarriages as a result, I thought there should really be a group focused on low progesterone."[*sic*]

**Reason 2. Users create groups to find peers with similar demographics:** Looking further at the user-created groups, we observe that even when some conditions have a

corresponding site-defined support group, users still initiate groups about these conditions, but dedicate them to particular demographic groups. For instance, the MedHelp data set includes multiple demographic-specific sub-communities of ADHD patients and caregivers, viz. "ADHD teens," "ADHD seniors," and "parents of ADHD children." Peers in the same demographic group may naturally better understand their situations and better communicate with each other. This is related to the principle of homophily in social science.

Analyzing the group descriptions uncovers explicit reasons from founders supporting this reason. For instance, the founder for the group for young mothers (with 45 members) stated:

> "For moms ages 16-25. I was a mom at the age of 16 and I really wanted a group my own age to go to! There isn't one right now so I thought I would start one!"

**Reason 3: Users create groups to build or maintain social relationships:** Following the principle of homophily, founders tend to form groups to build social relationships with other users similar to them in various aspects, not just in terms of demographics but also including hobbies or interests, religious beliefs, or health goals. For example, the founder of a group on creative writing (with 58 members) stated:

> "This group is for people to write a journal, or thoughts to be viewed by all of us ... By sharing our writings we will grow closer ..."

We also observe that founders initiated groups to stay connected with members in another group that they are part of. For example, a founder created two groups ten months apart from each other – one on "march 2011 babies" (with 64 members, created in July 2010) and the other on "march 2011 moms" (created in May 2011, with 10 members) to continue in touch with members of the former group and invite other new mothers.

**Reason 4. Users create new groups without the knowledge of existing groups with similar objectives:** One of the concerns with allowing users to create groups is that there might be many duplicated groups on the same or very similar topics. This leads to fractured communities, since interested members might get split between two similar groups. To validate if this was a valid concern, we analyze the user-created groups in two ways – (i) by comparing groups based on similar descriptions and (ii) by comparing groups based on overlapping membership.

First, each user-created group is compared against the five most similar groups, based on textual similarity of the titles and descriptions of the groups. An idf-weighted cosine similarity measure is used to compute text similarity.

We find that among the 50 pairs of groups with the highest textual similarity, 32 pairs of groups are duplicates. It is to be noted that since the pairs are sorted by the similarity of their descriptions, it is expected that the pairs at the top are indeed very similar and potentially duplicate groups. This, however, represents a small fraction of overall user-created groups. Some notable examples of duplicate user-created groups are as follows.

- A group on "military wives" (22 members) was created ten days after another founder created a group on "military spouses and family support group" (25 members).
- A group on "pregnancy after tubal reversal" (18 members) was created twenty days after another founder created a group on "pregnancy after tubal reanastomosis (tubal reversal / tr)"[*sic*] (15 members)
- There are four groups on "methadone detoxicating from home" created over a period of three years.

Some duplicate groups are created in error by the same founder, which are ignored in the analysis above.

Another approach to the analysis of duplicated groups is based on membership of the two groups. For each user-created group, we find the most similar site-defined groups, based on idf-weighted cosine similarity measure described above. The pairs that yield a low textual similarity are filtered out. Then, we compute the fraction of members of the user-created group that are also members of a site-defined group.

89 (11.9%) user-created groups are found to have a site-defined counterpart with a high text similarity. Out of these, only 8 (9%) have a membership overlap of at least 50% with the closest site-defined group. Even if we relax the textual similarity measure to consider as many as 238 pairs (31.9%), we find that only 26 (10.9%) of them have a membership overlap of at least 50%. This shows that users do not join duplicate groups. This reinforces our hypothesis that duplicate groups lead to fractured communities. A similar analysis between two user-created groups also produced similar results.

## 5.2 Why are user-created groups great?

Our analysis shows that user-created groups are more social and exhibit high density of friendship networks. Members of user-created groups are more active and interactive than those who participate in site-defined groups. Founders and users who are members of both site-defined and user-created groups are the most popular, social, active, and responsive among all users. On the other hand, users who do not join any group are the least active or social on MedHelp. These findings are consistent with the principles in the social identity theory, which states that group identities (group memberships, in our case) can deter free-riding behavior in public goods provision (inactivity in forums) (Eckel and Grossman 2005), and that salient group identity (membership to user-created groups, in our case) increases the effect (Shih, Pittinsky, and Ambady 1999). We must also note that the memberships to user-created groups are quite small compared to site-defined groups; and our findings are consistent with the observation that smaller groups may lead to lower information overload and hence better participation (Whittaker et al. 2003).

The friendship network structures of the user-created demographic groups, pregnancy groups, socializing groups, and public policy groups present a much larger density and a much larger clustering coefficient, suggesting the effectiveness of these groups to facilitate the connections and communications of their members.

Users in the socializing groups initialize many more threads; users in the demographic groups and the family support groups reply to more threads; and a thread in the public policy groups, demographic groups, and goal-oriented groups involves more users. Users who join a user-created group make ten more friends, initiate ten times more threads, and participate in twenty times more threads.

### 5.3 Who should care?

Based on the analysis presented in Sec. 4 over the MedHelp data set, there are specific implications for designers and researchers who desire to build better online health forums and communities.

1. Allowing users to initiate groups helps build unique communities that are specific to user needs and preferences. These include social groups that might not be directly relevant to a particular disease, but help improve the overall quality of life for patients.
2. Allowing users to make suggestions of new groups on rare diseases or complications would help health forums increase their coverage of medical conditions.
3. Enabling users to initiate groups helps create better communities that are more social and more active. This can be further supported by allowing users to form sub-groups within a larger site-defined group to cater to specific demographics or cohorts of patients suffering specific complications.
4. Enabling search of existing groups and expanding the search to include the title and description of existing groups would help identify existing groups and avoid the creation of duplicated groups.
5. Recommending relevant groups to users can help increase user engagement and reduce fractured communities. The recommendation may be based on demographics, location, or users' hobbies and interests, in addition to specific medical conditions.

These design implications can potentially improve the effectiveness of online health forums in satisfying users' needs for finding peers, seeking health information, and receiving informational and emotional support. They can also increase user engagement in online health communities.

## 6 Related Work

Our work is closely related to studies exploring users' motivation of joining online health communities. The oft-mentioned benefit of communicating health information online is to obtain experience-based information about particular treatments or behavior strategies from patients with similar experiences (Newman et al. 2011; Ressler et al. 2012; Ziebland and Wyke 2012). These studies also suggest that people visit online health communities to seek emotional support, engage with others to make progress on their health goals, and decrease a sense of isolation by connecting with community members. Existing works, such as (Bambina 2007), suggest that participants in online groups can receive emotional support either through direct messages of care and concern from other participants, or indirectly, by comparing or relating to experiences shared by others. Our analysis of user-created groups supports these studies and reveals additional reasons to participate in online health communities.

Researchers have also studied the dynamics of online communities. According to (Arguello et al. 2006), the two key elements of a successful online community are community responsiveness to members' messages and members' commitment to the community. They also suggest that group identity as well as group size may affect the vitality of a community. Other studies have mentioned the importance of member retention and commitment (Wang, Kraut, and Levine 2012). In our analysis of user activity, we found that small, user-created groups induce higher user participation than large, site-defined groups. Other researchers, such as (Maloney-Krichmar and Preece 2005), have found that strong group norms for support and reciprocity and strong subgroups can also actively contribute to the community's stability and vitality.

As one of the earliest of its peers, MedHelp has attracted attention from researchers interested in studying online health communities. In (Yang and Tang 2012), the authors developed a method of identifying influential users in MedHelp using threaded messages in three MedHelp forums. Similarly, in (Chuang and Yang 2012b), the authors applied descriptive content analysis on three months of data from the MedHelp Alcoholism support community. They found that the functionalities that support journal writing and leaving individualized notes to other members provided by MedHelp strengthened relationships between users and increased user engagement. They also identified different types of support exchanged in the alcoholism support forum (Chuang and Yang 2012b). To our knowledge, ours is the first work that explores user-created groups on MedHelp, and quantitatively compares it to site-defined groups.

## 7 Conclusions and Future Work

In this paper, we studied the usefulness of user-created groups over site-defined groups. We derived a categorization of user-created groups according to the underlying reasons leading to their creation, and investigated if there are significant differences between site-defined and user-created groups with respect to group memberships, activity levels, social network metrics, and interaction patterns. Our analyses show that users initiate more homophily-driven communities that are also more social. We also find that the members of user-created groups are more vocal and more socially active than those who participate in site-defined groups. The findings suggest that providing the capability to allow users to create their own groups can improve user engagement in health forums, foster rapport among users, and ultimately create a more effective social environment that supports information sharing and community building. We believe that the analysis presented in this paper helps understand the needs of online patient communities and supports design of better online forums.

This study provides a platform to further investigate the information shared in health forums. One research direction

suggested, but not explored, in this paper is to build a recommendation system for patients based on stronger peer-to-peer matching. Survey-based tools could be deployed to get a deeper understanding of the user intentions to initiate new groups. In the future, we plan to also study the quality and trustworthiness of the content presented in such forums, especially in groups focused on specific medical conditions, and the effect of fractured communities on effective communication. There is also an increasing trend of participation from medical professionals in online forums that is relevant, but beyond the scope of this paper.

## 8 Acknowledgments

This work is partially supported by the National Science Foundation under Award Numbers IIS-0968489, IIS-1054199, CCF-1048168, and BCS-1111019; the National Center for Advancing Translational Sciences under Award Number UL1TR000433; and by MCubed, a University of Michigan internal funding program. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

## References

Arguello, J.; Butler, B. S.; Joyce, E.; Kraut, R.; Ling, K. S.; Rose, C.; and Wang, X. 2006. Talk to me: Foundations for successful individual-group interactions in online communities. In *Proceedings of the SIGCHI conference on Human Factors in computing systems (CHI)*, 959–968. ACM.

Bambina, A. D. 2007. *Online Social Support: The Interplay of Social Networks and Computer-Mediated Communication*. Cambria Press.

Capstrat. 2010. Health care information – where do you go? who do you trust? In *Capstrat Public Policy Poll, conducted Apr 2010*.

Chou, W.-Y. S.; Hunt, Y. M.; Beckjord, E. B.; Moser, R. P.; and Hesse, B. W. 2009. Social media use in the United States: Implications for health communication. *Journal of Medical Internet Research* 11(4):e48.

Chuang, K. Y., and Yang, C. C. 2012a. Interaction patterns of nurturant support exchanged in online health social networking. *Journal of Medical Internet Research* 14(3):e54.

Chuang, K. Y., and Yang, C. C. 2012b. A study of informational support exchanges in MedHelp alcoholism community. In *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer. 9–17.

Csardi, G., and Nepusz, T. 2006. The igraph software package for complex network research. *InterJournal* Complex Systems:1695.

Eckel, C. C., and Grossman, P. J. 2005. Managing diversity by creating team identity. *Journal of Economic Behavior & Organization* 58(3):371–392.

Fox, S. 2011. Peer-to-peer healthcare. In *Pew Research Center's Internet and American Life Project*.

Gill, P. S., and Whisnant, B. 2012. A qualitative assessment of an online support community for ovarian cancer patients. *Patient Related Outcome Measures* 3:51–58.

Hagan 3rd, J. C., and Kutryb, M. J. 2009. Cataract and intraocular implant surgery concerns and comments posted at two internet eye care forums. *Missouri Medicine* 106(1):78–82.

Harris-Interactive. 2011. The growing influence and use of health care information obtained online. In *Harris Polls, conducted Aug 9–15, 2011*.

Maloney-Krichmar, D., and Preece, J. 2005. A multilevel analysis of sociability, usability, and community dynamics in an online health community. *ACM Trans. Comput.-Hum. Interact.* 12(2):201–232.

Newman, M. E. J., and Park, J. 2003. Why social networks are different from other types of networks. *Physical Review E* 68:036122.

Newman, M. W.; Lauterbach, D.; Munson, S. A.; Resnick, P.; and Morris, M. E. 2011. It's not that I don't have problems, I'm just not putting them on Facebook: Challenges and opportunities in using online social networks for health. In *Proceedings of the ACM 2011 conference on Computer Supported Cooperative Work (CSCW)*, 341–350.

Newman, M. E. J. 2003. The structure and function of complex networks. *SIAM review* 45(2):167–256.

Ressler, P. K.; Bradshaw, Y. S.; Gualtieri, L.; and Chui, K. K. 2012. Communicating the experience of chronic pain and illness through blogging. *Journal of Medical Internet Research* 14(5):e143.

Shih, M.; Pittinsky, T. L.; and Ambady, N. 1999. Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological science* 10(1):80–83.

van Uden-Kraan, C.; Drossaert, C.; Taal, E.; Seydel, E.; and van de Laar, M. 2009. Participation in online patient support groups endorses patients' empowerment. *Patient Education and Counselling* 74(1):61–69.

Wang, Y.-C.; Kraut, R.; and Levine, J. M. 2012. To stay or leave? The relationship of emotional and informational support to commitment in online heath support groups. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW)*, 833–842.

Watts, D. J., and Strogatz, S. H. 1998. Collective dynamics of "small-world" networks. *Nature* 393(6684):440–442.

Whittaker, S.; Terveen, L.; Hill, W.; and Cherny, L. 2003. The dynamics of mass interaction. In *From Usenet to CoWebs*. Springer. 79–91.

Yang, C. C., and Tang, X. 2012. Estimating user influence in the MedHelp social network. *IEEE Intelligent Systems* 27(5):44–50.

Ziebland, S., and Wyke, S. 2012. Health and illness in a connected world: How might sharing experiences on the internet affect people's health? *Milbank Q* 90(2):219–49.