Measures of User Interactions, Conversations, and Attacks in a Crowdsourced Platform Offering Emotional Support

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Computer Engineering

By

SAMIR YELNE
B.E., Nagpur University, Nagpur, India, 2014

2016
Wright State University

WRIGHT STATE UNIVERSITY
GRADUATE SCHOOL

DECEMBER 13, 2016

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Samir Yelne ENTITLED Measures of User Interactions, Conversations, and Attacks in a Crowdsourced Platform Offering Emotional Support BE ACCEPTED IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science in Computer Engineering.

Derek Doran, Ph. D.
Thesis Director

Mateen Rizki, Ph.D.
Chair, Department of Computer Science & Engineering

Committee on
Final Examination

Prof. Derek Doran, Ph. D.

Prof. Junjie Zhang, Ph. D

Prof. Tanvi Banerjee, Ph. D

Robert E. W. Fyffe, Ph.D.
Vice President of Research and Dean of the Graduate School

Abstract

Samir, Yelne. M.S.C.E., Department of Computer Science and Engineering, Wright State University, 2016. Measures of User Interactions, Conversations, and Attacks in a Crowdsourced Platform Offering Emotional Support .

Online social systems have emerged as a popular medium for people in society to communicate with each other. Among the most important reasons why people communicate is to share emotional problems, but most online social systems are uncomfortable or unsafe spaces for this purpose. This has led to the rise of online emotional support systems, where users needing to speak to someone can anonymously connect to a crowd of trained listeners for a one-on-one conversation. To better understand who, how and when users utilize these systems, and to evaluate their safety, this thesis offers a comprehensive examination of the characteristics of users and their interactions from a massive, leading emotional support platform. From a big data set of millions of conversations across hundreds of thousands of users, the study employs statistical measurement techniques and predictive analytics to shed light about the ways these platforms are utilized, and the extent to which users behave in un-wanting ways. The analysis leads to recommendations on promoting positive system utilization and an understanding of the effectiveness of protections in place to thwart emotional attacks. This work is likely the first to measure the activities and interactions in an online social system for emotional support.

Contents

List of Figures

List of Tables

Acknowledgment

I would like to thank my thesis director Professor Derek Doran for his expert guidance and encouragement throughout my research for this work. I would also like to thank my colleagues Professor Maria-Carla Calzarossa, Luisa Massari from University of Pavia, Italy and Nripesh Trivedi from the Indian Institute of Technology Varanasi who provided thoughtful discussions and contributions to the presented research.

1 Introduction and Motivation

Internet and online based social media platforms, such as Facebook, Instagram, Twitter and some instant messaging services such as Snapchat or Kik are rising as the dominant way people in society communicate with each other. In addition to this, online emotional support system is an emerging kind of online platform whereby those seeking some kind of emotional help (e.g. a poor medical prognosis, loss of a loved one, or depression) consult a volunteer from a crowd of others that, ideally, offer advice and support [1]. An emotional support system depends on this crowd to be well intended, to be trained in active listening techniques, and to be selfless in their support of others. The rapid increase in popularity of emotional support systems indicate that they are effective tools where users achieve positive outcomes [2]. Their popularity is further bolstered by the fact that traditional social systems are public (e.g. Twitter) or semi-public (e.g. Facebook), making them unsuitable to seek emotional support for major issues. This is because users of public and semi-public systems must navigate the tension between sharing honestly about their struggles and asking for help against the positive and inviting impression they strive to develop for their social network, followers, and outsiders (e.g. a potential employer) who may search and discover their profile [3].

Many kinds of emotional support systems have been implemented, recent examples include 7 Cups of Tea, BlahTherapy, and CrisisChat. 7 Cups of Tea (7cot) is a canonical, long standing example of an online emotional support service. 7cot has a vast community of active listeners who are ready to help and listen to those who are in need. There are two types of registered users on the website, namely listeners and members. The unregistered users are called Guests. Listeners are individuals who go through active listening training on the site before becoming available to chat one-on-one with members who need support. 7cot has seen a remarkable growth in the context of number of registered active listeners and members on the website since its inception. The growing demand of such platforms, therefore make it suitable for our study of understanding how users utilize such platforms, their design choices that encourage high user engagement on such platforms.

With society's ever increasing dependence on online social systems as a means to communicate with others, there has also been increase in some of the societal problems associated with them. Online harassment, or cyberbullying, is one of the greatest problems born out of online social systems. We define cyberbullying as the practice of sending an offensive, crude, rude, or demeaning message to another user online with the intention of attacking them personally. A Pew research survey from October 2014 reports that 73% of adult internet users have observed a user being harassed in some way online and 40% have experienced it themselves which is an alarming percentage. Past studies confirm that victims of online harassment face terrible psychological effects like depression, low self-esteem and even suicidal tendencies. For example, a 15 year old girl once committed suicide after facing a barrage of offense and slanderous messages on Facebook [4]. Cyberbullying is thus a grave problem with potentially devastating consequences to its victims. Therefore it is important to explore whether cyberbullying exists on platforms providing emotional support and study the measures adopted to deal with cyberbullying are sufficient enough to deal with this menace.

This thesis explores in detail how users on this platform tend to connect with each other, how they behave, utilize these systems, what design choices and user behaviors promote high user engagement on these systems and sheds light on the grave problems of cyberbullying on this platform[1] The findings provide useful insights on improving existing platforms providing emotional support, creating new ones that are effective, on understanding how the Internet may be used as a crowdsourced clinical psychology tool to help people from immediate emotional distress and other emotional problems, discuss the problems of cyberbullying and the ways to cope with it.

[1]Portions of this thesis were previously written and published by the author in [2, 5].

2 Related Research

Studies in the past have tried to understand the users and their behaviors on online emotional support systems. Maloney-Krichmar et al. investigated the dynamics of group interactions among an online self-help group for knee injuries [6]. Barak et al. established a positive relationship between the amount of activity of adolescents in an online support group and the emotional relief they felt, underscoring the importance of building online systems that facilitate user interactions [7]. Ploderer et al. delved into the discussion topics on a Facebook group of people trying to quit smoking, and found that most supportive responses come from those who just began trying to quit, rather than long-term quitters [8]. Yuen et al. highlighted how remote assessment, treatment and consultation which are provided via internet through self help websites, and videoconferencing have great potential to increase access to high quality psychological services. They also discussed clinical, ethical, logistical challenges involving security, competence, usability and technical difficulties on such platforms [9]. Wang et al. analyzed the relationship of emotional and informational support a user is exposed to on online support groups to their commitment in online health support groups [10]. Zhang et al. studied a facebook diabetes group and concluded that participating in such online groups have a lot of advantages for the users to share important information regarding diabetes irrespective of where they came from, their differences in languages and diversity [11]. Valerie et al. emphasized that online social networks play an important role in supporting different types of decision making, as they provide their participants various forms of support, ranging from the instrumental to emotional and informational [12]. Saha et al.

demonstrated how online social networks are easy and accessible communication platforms particularly in the context of users of autism to help, share and connect with other users having autism or with their families, caregivers etc. which helps them considerably by extracting useful information and at the same time also provide them social support [13]. Past work have also studied characteristics of online social networks, for example Mislove et al. studied characteristics of many online social network graphs such as flickr, youtube, LiveJournal, orkut at a large scale to understand the properties of these networks so as to improve the design of such systems and designing new applications which could promote high user engagement [14]. Han et al. compared structural properties of Weibo and Twitter networks to understand differences and nuances of how the users use these online social systems differently [15].

A lot of work have also been done on studying cyberbullying on online platforms. For instance, An October 2014 Pew research survey offers the best evidence that cyberbullying is a major phenomenon that impacts Internet and social media users. Academic studies have also demonstrated the negative factors associated with cyberbullies that attack U.S. teenagers, and unearthed the fact that bullying is intrinsic to users rather than to a particular platform. In other words, no matter the online social system, cyberbullying should be expected to occur. Previous studies also demonstrate the ill effects of cyberbullying to its victims which could be as grave as suicide [4]. Past works have also tried to understand how users utilize online social networks and social media to manage emotional and personal problems. For example, Newman, et al. showed that people are cautious while sharing health related information online on social networks like Facebook. They also demonstrated that users are hesitant to share certain types of information, especially personal or information that may be used as fodder for cyberbullies, to protect themselves and to manage their online impression [3].

Thus past studies have covered various aspects of online social networks such as understanding structural properties of networks to better understand the use of such systems, thereby suggesting recommendations for improving such systems, demonstrated advantages of using online social communities to share information about wide range of problems such as emotional or medical problems. On the other hand there also has been lot of work regarding the menace of cyberbullying on online systems, the adverse effects of cyberbullying on the victims and the challenges associated with it.

All the above mentioned aspects regarding online social networks have been explored individually in the previous studies. Also very little work has been done studying how the platforms providing online emotional support are being utilized by the users and whether such platforms are advantageous and popular. This thesis cohesively covers all the above mentioned aspects such as studying structural properties and characteristics of the online social networks, understanding user engagement, user behavior and finally understanding the problems and challenges of cyberbullying from the context of online social network providing emotional support.

3 7 Cups of Tea: An Overview

7 cups of tea (7cot) is an online emotional support service that offers crowdsourced emotional support. It was launched in December 2013. It has a vast community of active listeners who are ready to help and listen to those who are in need of emotional support. There are two types of registered users on the website namely listeners and members. The listeners are the users who volunteer to help other users (members or guests) on the website. They are trained in active listening to support people (members). There are currently 42 interactive training courses for a listener which the website provides. Examples of these courses include trainings such as active listening, self-harm, cultural diversity, bullying, work related stress, sleeping well, and a variety of symptom specific courses. Members on the other hand are the users who may have wide range of emotional problems or who may just want to talk to someone (listener). Becoming a member is free and has many advantages like sending a private message to a listener to set up a conversation, scheduling regular listening sessions in the future, connecting to a listener without any time limit etc. Users may also take on multiple types; for example a listener that passes the required training class may become a hybrid who is also a member and can switch his/her listener and member accounts as per the need. The unregistered users are called guests who do not wish to go through a such a registration process as a member goes through and opt to connect to a listener immediately. A guest thus don't enjoy as much privileges as a registered member. There are many other resources available on the website for a registered member if he or she does not wish to have a conversation with a listener. For example the website provide self-help guides which consist of useful information and

videos related to diverse emotional problems such as stress, grief, anxiety, managing emotions etc. and other problems such as alcohol/drug abuse, managing finances, college life etc. The members can access this valuable information very easily, if they prefer to work alone to find solutions to their problems. There is also a feature called mindfulness exercise where the member can listen to audio clips provided by the service on the previously mentioned topics.

We note that 7cot maintains a unique identifier for each guest, based on browser signature and a cookie, so that it can keep track of the activities of the same guest over multiple sessions. Guests and members have the option of identifying themselves as a teenager or an adult. This classification makes it easy for them to connect to listeners who have expertise as per different age groups. Users communicate with others in three different channels: group chats, conversations, or forums. Group chats are free exchanges that multiple members and listeners may participate in. There are group support rooms which are classified as per different emotional problem. For example there is a group support room exclusively for anxiety related issues. Similarly there are many other group support rooms available based on various problems. Conversations are private exchanges of messages between members and listeners. A conversation is a single, permanent connection between a member and listener, lasting for an indefinite amount of time. Conversations maintain state even after a member or listener logs off of the site and may be returned to at any time. A conversation is personal if the user selects a specific listener to speak with. The conversation is general if, instead of picking a listener, the user asks the service to connect with any listener presently available. Users can also participate on different forums pertaining to different topics and can post their opinions and views about the topic (e.g. forum on anger management, anxiety etc). The service provides lot of resources in the form of videos or audio clips or specific information related to specific problem in addition to the large pool of trained listeners ready to listen and help. The information is also updated on quite a regular basis so the members can take full advantage of it.

In recent years, 7cot has seen a remarkable growth in the context of number of registered active listeners and members on the website since its inception, It has attracted more than 150,000 listeners who have helped over 450,000 members in over 3 million one-on-one conversations (private asynchronous or real-time message exchanges). This suggest the fact that online emotional support services offering crowdsourced clinical psychology is both effective as well as popular [2]. The tremendous popularity of 7cot also demonstrates a demand for safe online spaces providing emotional support. Gaming or progress mechanisms are integrated into the site to represent user reputation and experience. For eg. listeners gradually accrue "cheers" over time, and after attaining certain amounts their listener level is upgraded to a more prestigious category. Listeners also achieve badges displayed on their profile for accomplishing tasks such as helping members facing a specific type of need (e.g. loss of a loved one). The level for listeners start from Listener advancing to Peer Supporter then Mentor, Mentor Leader and finally they attain the level of Ambassador. The level they attain is thus based on their level of involvement. Similarly, members accrue "growth points" for performing simple activities such as posting on the forum, or sending messages during a conversation. Accruing enough growth points will upgrade their member level, a rank that reflects a commitment to the site and progress toward improved mental health.

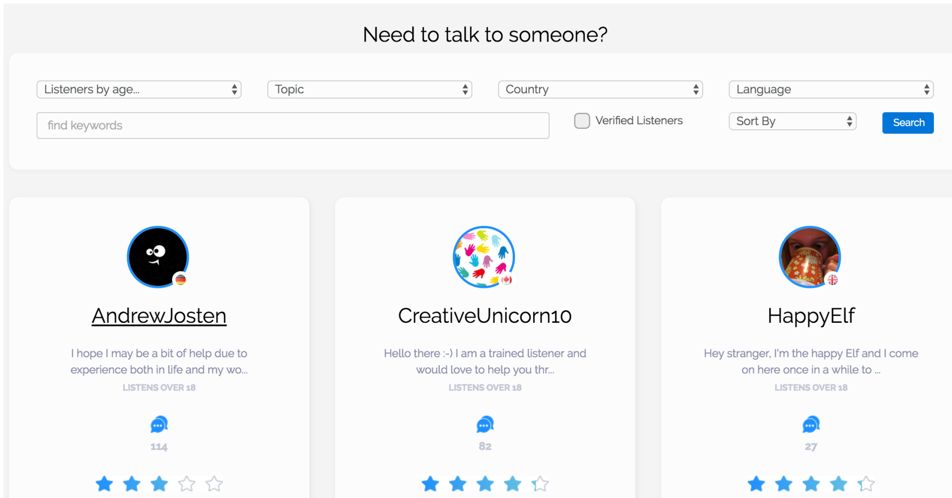


Figure 3.1: Browsing for Listeners on 7cot

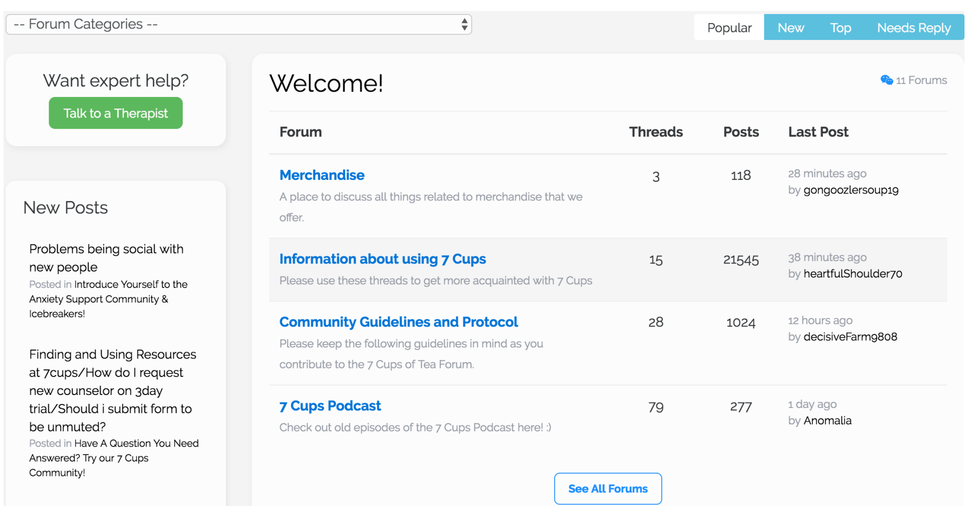


Figure 3.2: Forums on 7cot

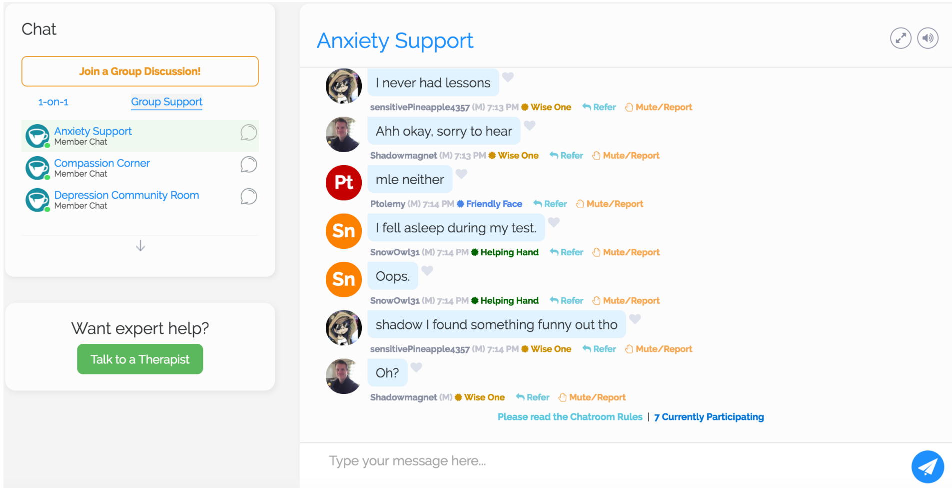


Figure 3.3: Group Chats on 7cot

Figure 3.1 shows the interface for the seven cups of tea website. Users can browse through various listeners by viewing their complete profile, bio, photos, ratings and reviews. They can search for a listener by using various filters like age, topic, country and their language. Figure 3.2 shows how users can participate in forums based on different topics. In a forum, people of different ages, points of view, backgrounds, and experience can share information on a certain topic which would conclusively offer a broad response to that topic at hand. Therefore forums play an important role to provide diverse perspectives and additional information on specific problems. The users can visit on the forums anytime they want and participate in it. Forums therefore makes an excellent resource specially for those users who may be reserved or too shy to open up about their problems during a conversation with a listener and are quite introverted. Figure 3.3 shows group chats on the website where many users can have a talk regarding a particular topic simultaneously For example the figure shows users participating on anxiety support group chat. Group chats could also lead to positive user experience. It offers interactivity as many users participate in it at the same time, which could lead to friendships and community building.

We explore the research questions using a database of user interactions and behaviors shared by 7cot. This database, which cannot be shared publicly due to a non-disclosure agreement, capture the attributes, interactions, and activities of all users performed since its inception on December 5th, 2013 through August 18th, 2015. It includes metadata about every user except for those attributes related to the users true identity and contact information. Attributes of each conversation record were limited to participant identifiers, the date the conversation commenced, the number of messages exchanged by each party, if the conversation was terminated by the member or listener, and the timestamp of the last message sent. We use this data to investigate our research questions next.

4 Users Interaction Structure

4.1 Interaction Structure

We next study the patterns of member engagements with listeners on 7cot. The patterns are found through analysis of a network where members and listeners are connected if they held at least one conversation with each other. We also study networks that connect members (listeners) to each other if they had a conversation with at least one common listener (member). Structural analyses of the networks inform how members are choosing to engage with listeners on 7cot, if some subsets of listeners are more popular than others, and if a pattern of members selectively choosing listeners can be seen.

We represent all 7cot interactions as a bipartite network from members to listeners. We consider all conversations that contained at least one message sent by either a member or listener (note that guests are excluded from this analysis and will be the subject of future work). Figure 4.1 illustrates the structure of this bipartite network. Here, listeners are colored green while member nodes are sized larger and colored hotter by their degree. The coloring finds a presence of some members that connect to vast numbers of other listeners, while the majority of members connect to a similar number. Moreover, we note that green nodes in the center of the network represent listeners who are embedded between a large number of other members, while listeners at the periphery of the network only connect to a very small number of others. The ratio of embedded to periphery nodes in the structure may indicate

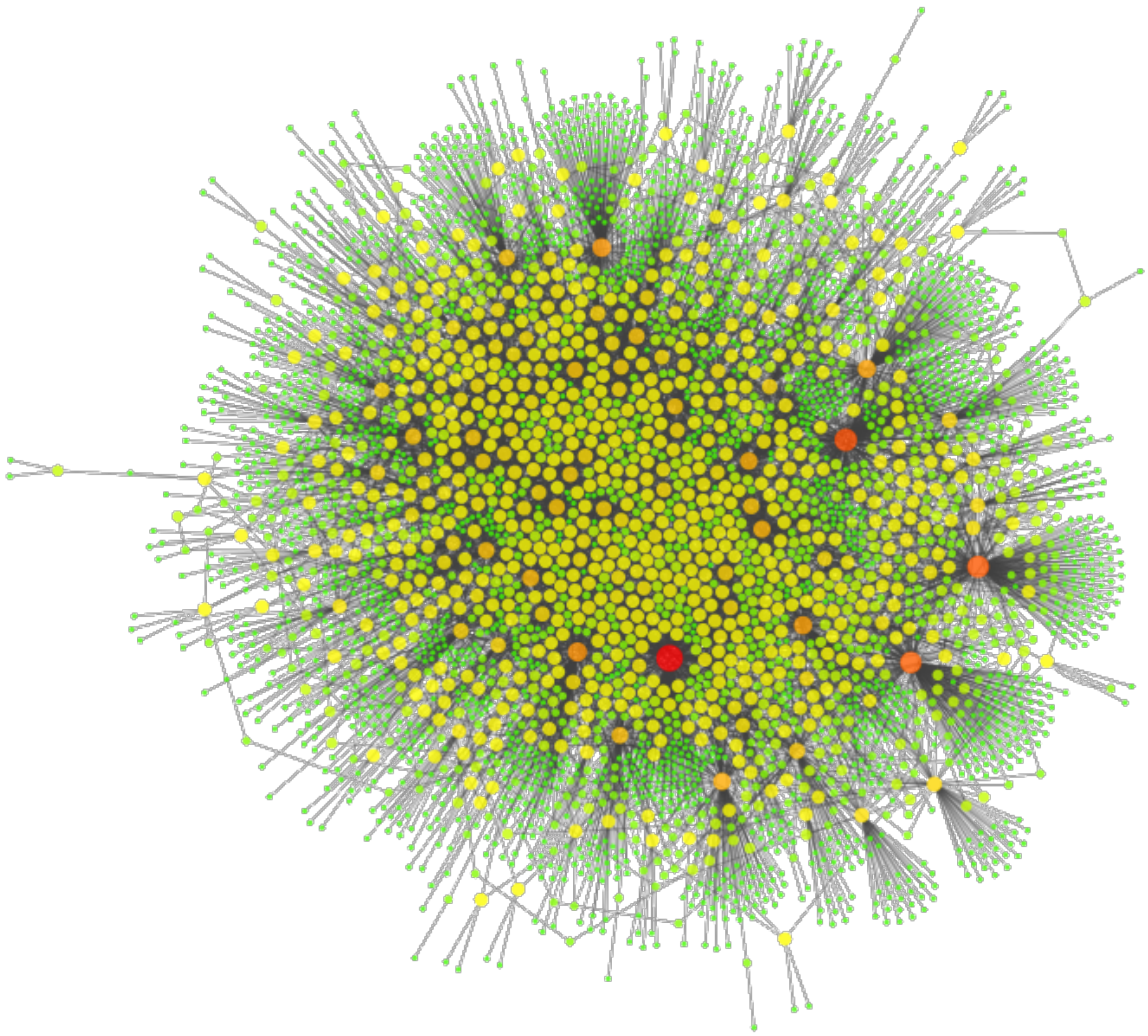
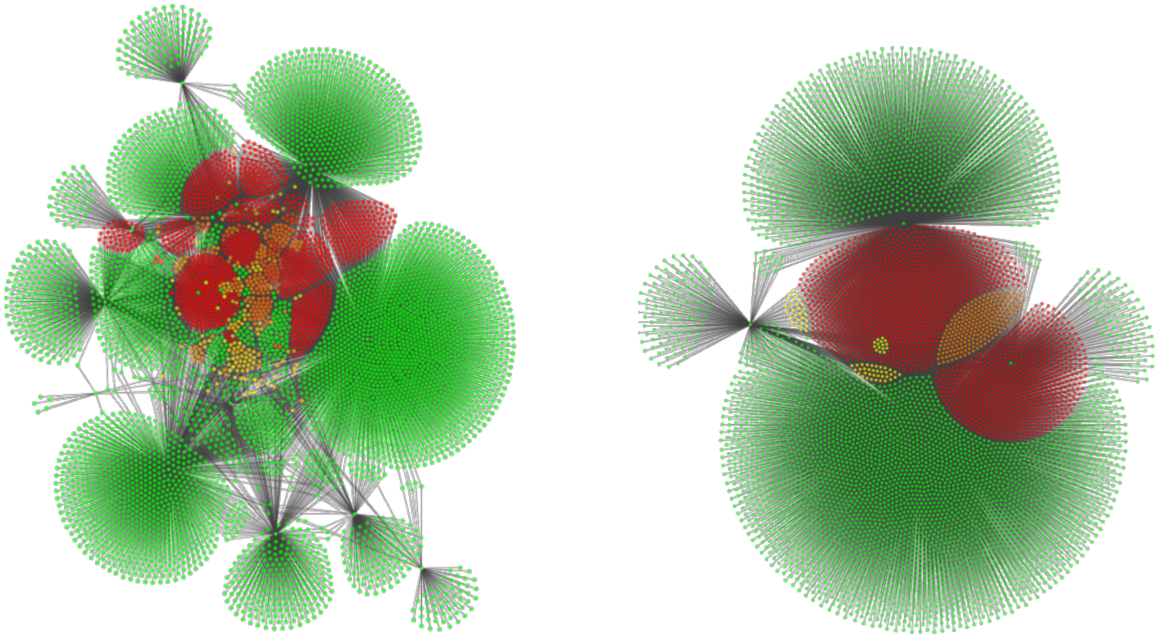


Figure 4.1: 7cot bipartite network; listeners are green and members are colored with hotter colors by their degree.

	Bipartite Network	Member Proj.	Listener Proj.
$\|V\|$	117,372	86,877	30,495
$\|E\|$	465,437	12,657,611	10,359,604
$\langle k \rangle$	5.39	291.39	679.43
$\mathcal{C}$	N/A	0.734	0.636
$\mathcal{A}$	N/A	-0.10	-0.06
$\bar{d}$	3.46	2.56	2.30
ρ	N/A	0.003	0.022
Components	447	447	447
GCC Size	116,411 (99.2%)	86,364 (99.4%)	30,047 (98.5%)

Table 4.1: Bipartite and projection network features

that most listeners go underutilized or simply choose to connect to a small number of others. Table 4.1 lists the structural features of this bipartite network. The network has an average degree $\langle k \rangle$ of 5.39, i.e. members tend to connect to between five or six distinct listeners during their time on the service. This reaffirms the idea that members seek help from a number of others, perhaps to obtain different viewpoints or thoughts about their emotional problem. We also computed the number of connected components in the network. Only 477 disconnected components exist, the largest of which (GCC) includes virtually every user (99.2%) on the platform. In other words, there are virtually no members or listeners on 7cot who choose to exclusively search for and communicate only with each other. The single large GCC lets us compute the average path length in the network as $\bar{d} = \log(|V|/z_1)/\log(z_2/z_1) + 1$, an expression valid for networks that are nearly fully connected [16], where z_1 and z_2 are the average number of others a user can reach within one and two hops respectively. The small average path length $\bar{d} = 3.46$ may be indicative of the existence of a large 'core' of members and listeners serving as hubs that connect members and listeners to others across the bipartite structure. Listeners in the 'core' may thus connect to large and diverse sets of members, i.e., are the listeners that connect to members who request to speak with any available listener.



(a) Member network sample (b) Listener network sample

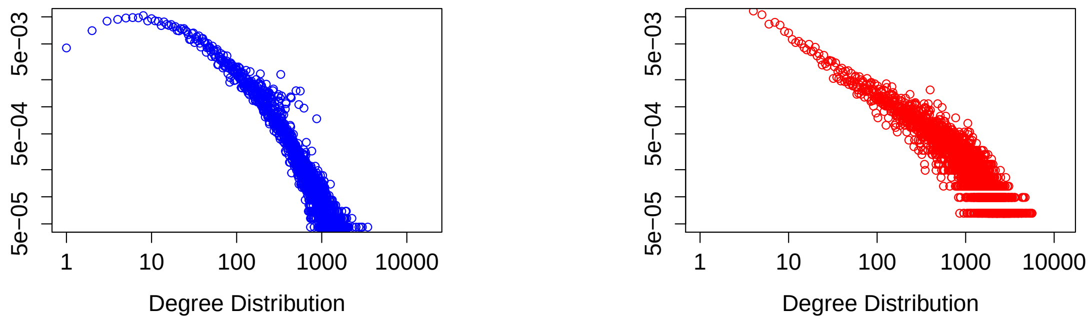
Figure 4.2: Edge sampled projection networks with nodes colored by clustering coefficient

We omit measuring the clustering coefficient $\mathcal{C}$, degree assortativity $\mathcal{A}$, and density ρ of bipartite network because their definitions are closely related to measurements taken over the network's *one-mode projections* [17]. One-mode projections capture the structure of interaction co-occurrences among the g listeners and n members of 7cot. Given a matrix $\mathbf{B} \in \mathbb{R}^{g \times n}$ where $\mathbf{B}_{ij} = 1$ if listener i has a conversation with member j, we define $\mathbf{P}^{(m)} = \mathbf{B}^T\mathbf{B} \in \mathbb{R}^{n \times n}$ and $\mathbf{P}^{(l)} = \mathbf{B}\mathbf{B}^T \in \mathbb{R}^{g \times g}$ as the adjacency matrices of the member and listener projection networks, respectively. We then have $\mathbf{P}^{(m)}_{ij} = c$ ($\mathbf{P}^{(l)}_{ij} = c$) if members (listeners) i and j hold a conversation with c common listeners (members). Structural patterns within the projection networks are discussed next.

4.1.1 Connectivity patterns

Table 4.1 gives the mean degree, global clustering coefficient, degree assortativity, average path length, density, and GCC size of the member and listener projection networks. These statistics may be compared with a visualization of a random sampling [18] of 10,000 edges of the projection networks in Figure 4.2. Nodes are colored hotter in the figure if they have a higher local clustering coefficient $\mathcal{C}_l$ (green nodes have $\mathcal{C}_l = 0$ and red nodes have $\mathcal{C}_l = 1$) and are drawn under a force directed layout so that nodes separated by small distances are positioned closer together. Although sophisticated sampling algorithms are needed to create samples that maintain many structural features of the sampled network [19], edge sampling still conveys the shape of the global network within the interconnected core of the sample (nodes participating in excessive numbers of open triangles are likely an artifact of edge sampling). The high mean degree, large GCC size, and small average path lengths of both projections further support the hypothesis that members and listeners do not limit themselves to interact with a small subset of listeners (members). They both exhibit weak negative degree assortativity, suggesting a small inclination for members (listeners) who share just a few common listeners (members) with others share them with those who have large numbers of listeners (members) in common with others. However, the lower degree, larger clustering coefficient, and larger path lengths of the member network imply a weak penchant for members to form clusters by the common listeners they connect to. Such clusters can be seen in Figure 4.2a as cliques in the core of the member network. These clusters may be traces of member groups that connect to similar 'types' of listeners.

We find the degree distributions of the projection networks, presented in log-log scale in Figure 4.3, to take dissimilar shapes. The listener degree distribution exhibits a near straight line pattern indicative of a power-law distribution, but the pattern is less pronounced in the member degree distribution. We quantify this difference by running a maximum likelihood based test of the null hypothesis H_0*: the empirical data has a power-tailed distribution* (the test also yields best fitting power-law exponent α under the null) [20]. The test leaves little room to reject H_0 for the listener degree



(a) Member projection

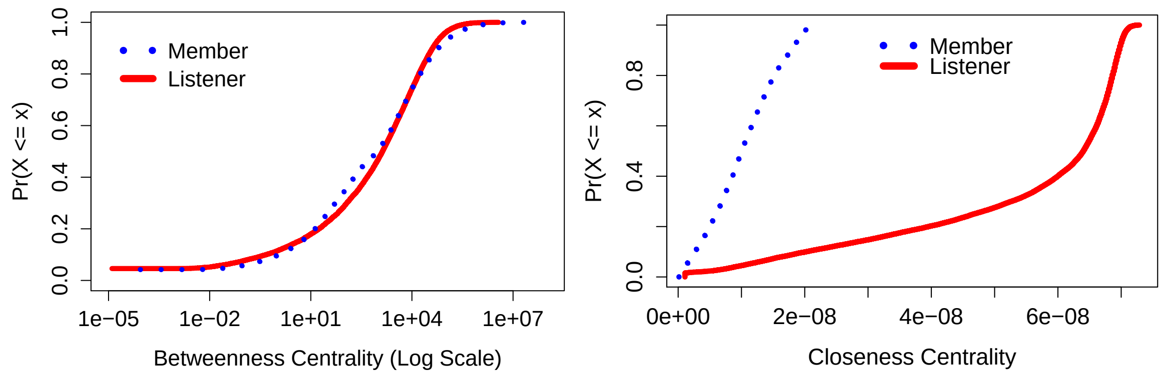
(b) Listener projection

Figure 4.3: Projection network degree distributions

distribution ($p = 0.985; \alpha = 2.51$). However, there is more doubt for the member degree distribution ($p = 0.362; \alpha = 2.34$). That the listener degree distribution has a power-tail suggests significant variation in the number of common members listeners share with each other, and that the probability of sharing orders of magnitude more members than expected is not negligible. A similar statement could be made about members, however they may exhibit less variation since we are less confident if a power-tailed trend exists. The difference of the distributions shape may be explained by members who only need to connect to a limited number of listeners in order to have many problems resolved, or by members who choose to connect deeply with a small number of listeners. Such behaviors place a ‘soft limit’ on the largest number of listeners members may connect to, weakening the support for a power-tail to emerge [21]. On the other hand, so long as a listener is available for newly added members to connect to, there may be no limit on the number of new members a listener may connect to over time.

4.1.2 Centrality analysis

We also study connectivity-based notions of network centrality in the projection networks. We first consider the betweenness centrality of a user u, defined as $b(u) = \sum_{u \neq i \neq j} \sigma_{ij}(u)/\sigma_{ij}$ where σ_{ij} is the number of shortest paths from users i to j and $\sigma_{ij}(u)$ is the number of such paths that include u. This measure reflects the notion that a user is 'central' if she is often part of the shortest path among two others in the network. Figure 4.4a plots the cumulative distribution (CDF) of the centrality scores across the two networks on semi-log scale. Its rapid ascent and long left tail indicate that almost all users are part of a number of shortest paths in the network. The networks are therefore structurally robust to the loss of users. We also consider the closeness centrality of a user u, defined as $c(u) = (\sum_j d(u,j))^{-1}$ where $d(u,j)$ is the distance from user u to j. Figure 4.4b gives the CDF of closeness centrality on the two networks (note that the x-axis is not in log scale). That the CDF for the listener distribution is stretched farther than the member distribution is only because there are fewer nodes in the network. Unlike betweenness centrality, the closeness centrality CDF of the two networks takes on different shapes. The CDF of the member network has only a slight curvature at its left and right tail, with a nearly linear body. This suggests that the centrality scores exhibit a small peak around the mean of the distribution but are otherwise uniformly distributed. The centrality scores of listeners are uniformly distributed up to approximately the 40^{th} percentile, at which point they become heavily skewed. A majority of listeners, therefore, are at a much shorter distance from those below this 40^{th} percentile. This pattern may be indicative of a core-periphery structure [22] in the listener projection network that does not exist in the member one, where those in the core (periphery) have high (low) closeness centrality. The probability of a listener falling in the core may be correlated with the diversity of the members she connects to: connecting to many



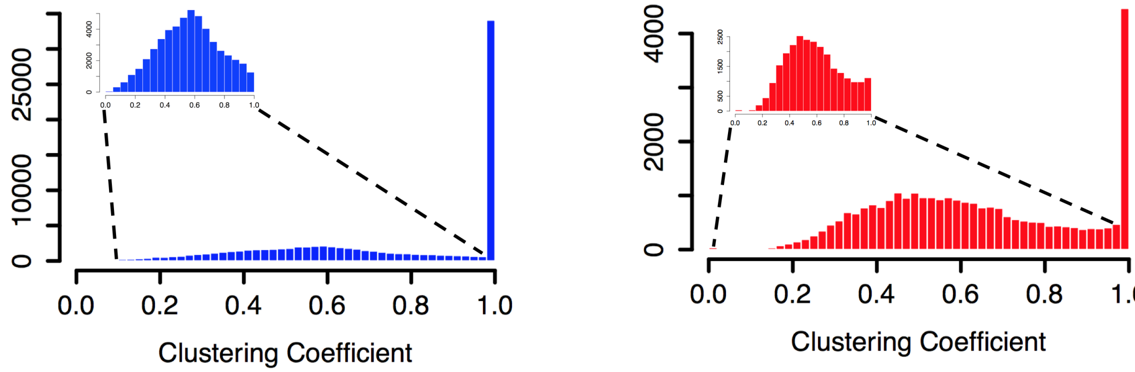
(a) (b)

Figure 4.4: Centrality distributions for members and listeners

different members increases the probability of sharing a connection with a listener already in the core.

4.1.3 Network transitivity

Finally, we use the local clustering coefficient distributions of the projection networks to study the tendency of transitive relationships among members and listeners. A transitive relationship is one where if user A is a member (listener) connected to user B and B is connected to C, then A is connected to C. Table 4.1 lists the global clustering coefficient, defined as the average of the number of closed triangles in a user's neighborhood divided by the number of possible links that could exist within it [23], as $\mathcal{C} = 0.734$ and 0.636 for the member and listener projections respectively. The large coefficients signify that transitive relationships dominate the projection networks. However, histograms of the local clustering coefficients $\mathcal{C}_l$ in the member and listener network in Figure 4.5 show that the large values are driven by the 38.9% of members and 13.2% of listeners whose $\mathcal{C}_l = 1$. The high values of $\mathcal{C}$ are therefore driven by a small proportion of users with fully connected neighbors. When we consider users whose $\mathcal{C}_l < 1$, clustering coefficients appear to be



(a) Member projection

(b) Listener projection

Figure 4.5: Projection network cluster coefficient distributions

normally distributed. Normally distributed $\mathcal{C}_l$ distributions is a typical phenomenon in co-occurrence networks spanning many systems, including scientific paper authorship [24], [25], e-commerce co-purchases [26], and "related page" relationships on search engines [27], but the surge of members where $\mathcal{C}_l = 1$ is unique to 7cot interactions. This suggests that users with $\mathcal{C}_l = 1$ may not emerge from some natural or universal process innate to all co-occurrence networks. This is evidence that both members and listeners perform deliberate actions that drive them into fully connected neighborhoods in the projection networks. For example, members may be selectively connecting to the same pool of listeners that may have similar ratings, experiences, or bio's suggesting an expertise that members in their neighborhood do.

4.2 Summary of Findings

In summary, this chapter identified a number of important qualities about the users of 7cot and the conversations they have. Key takeaways from this study are:

- The giant connected component of the bipartite network include almost all the users of the platform, which tells us that only few users choose to exclusively search for and speak with each other.

- The degree distribution of listener projection network follows a power law but same may not be said for the degree distribution of member projection network. This suggest that member may tend to develop deep and strong relationships with some of the listeners rather than having an "exploratory" behavior where they are trying to connect with as many listeners as possible.

- The clustering coefficients of the member and listener projection appears to be normally distributed as is seen in many co-occurrence networks. A small percentage of members and listeners exhibit perfect clustering coefficients which is unique to this platform.

5 Understanding User Engagement

Next, we perform an engagement analysis of members on 7cot. Engagement analysis offers insights about the user and platform features that encourages members to return, listeners to stay active, and for members to have multiple, fruitful conversations. Such insights are practically important to help a platform retain new members and grow its community of listeners. They also identify qualities that encourage people to seek follow-up emotional support.

5.1 Factors driving engagement

We first relate the features and behaviors of members and its relationship to a measure of site engagement. Since sharing with listeners is the purpose of the service, we quantify engagement as the message rate of a member, that is, the average number of messages sent per day in conversations. We consider features and behaviors that, based on discussions with psychologists and designers at 7cot, may be related to engagement: (i) number of coins, growth points, and compassion hearts, which are gaming and progress measures related to a members reputation and experience; (ii) signup and last login date; (iii) reported distress level when they register; (iv) number of group chat messages; (v) number of page views from the 7cot Web and iOS applications; (vi) number of logins; (vii) number of conversation requests sent; (viii) number of self help page views; (ix) number of forum posts, views, and up-votes. Table 5.1 gives the Pearson correlation coefficient between the features and a members message rates. The coefficients make clear that the gamification features of the

Coins	0.247	Growth Points	0.977
Compassion Hearts	0.243	SignupDate	-0.009
Last Login Date	0.133	Distress Level	0.004
Group Chat Msgs.	0.120	Page Views (Web)	-0.002
Page Views (iOS)	-0.001	Login Count	-0.001
Conv. Requests	0.001	Self Help Views	0.005
Forum Posts	-0.001	Forum Views	-0.001
Forum Up Votes	0.201		

Table 5.1: Pearson correlation between message rate and user and behavior features

platform (accumulated coins, hearts, and growth points) are strongly related to the engagement of a member. However, conversation messages sent by members directly increase growth points, giving this correlation little meaning. Member attributes and behaviors unrelated to communication (signup and last login date, distress level, page views, and help article views) exhibit virtually no correlation, suggesting that users dealing with any type and degree of emotional distress, at any time, exhibit similar levels of engagement on the site.

Many features exhibit little correlation with user engagement, but subsets of features may feature interactions that are. For example, users who exhibit a high distress level and submit many conversation requests may have a high level of engagement even though the features are individually not correlated. Instead of exhaustively exploring all multiway interactions, we consider a random forest model that predicts user engagement by a regression over all features. A random forest is an ensemble of decision trees, each of which is trained over different bootstraps of the data. During training, each tree is limited to the use of distinct small subsets of the features to make spitting decisions. The bootstraps, limited choice of features for tree splitting, and averaging of results across the tree ensemble ensures the forest does not overfit the data even for large N [28]. We compute the importance of each feature to the random forest regression model by the mean square error (MSE) of the random forest predictions against the actual engagement of every user.

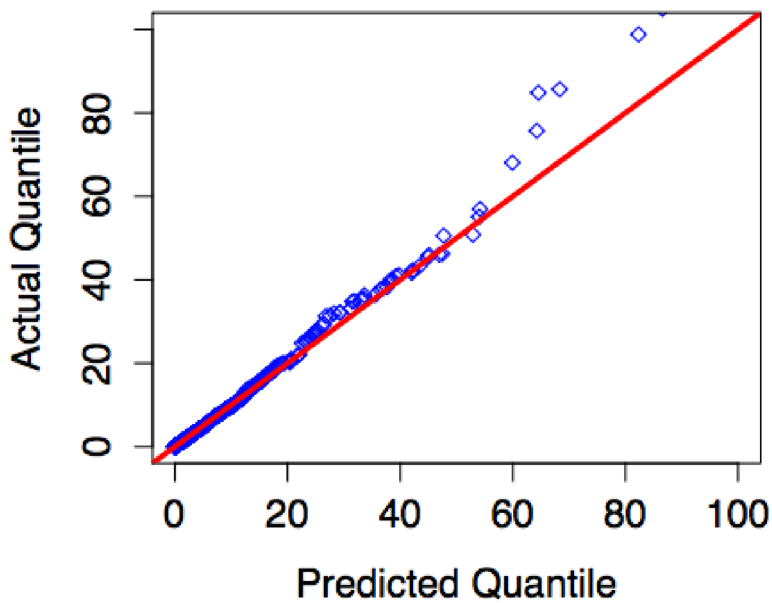


Figure 5.1: Quantile Plot

We trained a random forest regression model using 75% of the user data for a forest with N = 1000 trees and randomly choose 1/3 of the features for every tree splitting decision. The figure 5.1 shows the quantile plot and figure 5.2 shows the prediction scatter plots of the predicted and actual message rates for the 25% of users not used to train the random forest. These figures demonstrate that the decision tree models engagement very well (R2 = 89%), as the quantile plot shows a linear relationship between the distribution of the predicted and actual engagement rates up to the 60th quantile. The predicted vs. actual engagement rates in the Quantile plot only show normally distributed errors for users with low engagement.

Since the random forest reasonably models the relationship between member and behavioral features, we use it for feature importance analysis. Figure 5.3 shows the percent increase in MSE of a random forest trained with data where each factor was individually perturbed across the training data. As anticipated by Table 5.1, the total number of growth points of a member is the most important factor for predicting user engagement due to its direct correspondence with her message rate.

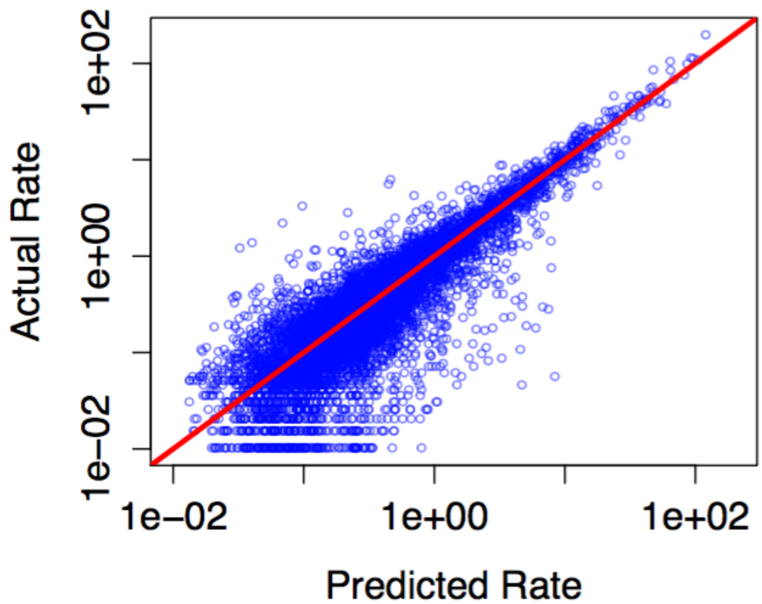


Figure 5.2: Predictions

Members signup and last login dates are the next most important features, each of which increases MSE by over 20% when they are perturbed. The signup date of a user is weakly anti-correlated with engagement according to Table 5.1, thus recent logins have a weak relationship to engagement. The number of messages sent in group chat is the next non-gaming related feature that is important for user engagement. This suggests that participating in group chats encourage users to become more engaged in their one-on-one conversations. It may be the case that users find group settings to be easier or less intimidating to participate in, and builds their confidence to have lengthy sessions with a listener. Finally, we note that the number of forum up votes introduces noise in the model, since perturbing this factor decreases the MSE of the random forest. One explanation may be that members who gain recognition for their forum posts may be disinclined to participate in conversations since they achieve recognition and perhaps satisfaction by only participating on the services forum.

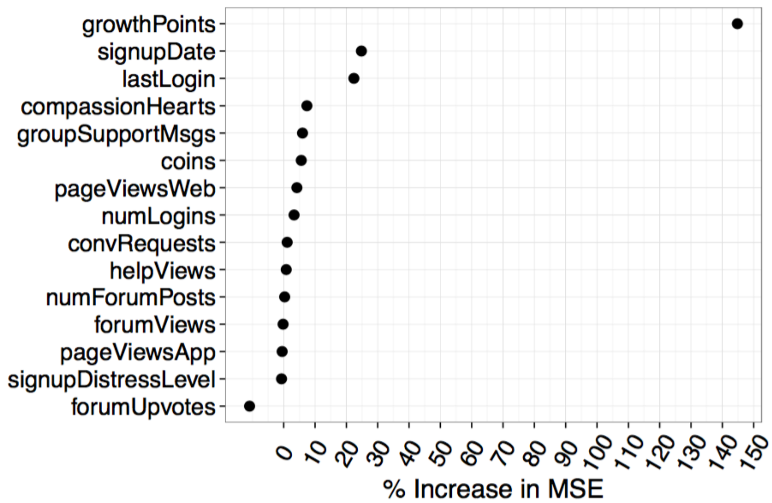


Figure 5.3: Feature Importance in Random Forest Model

5.2 New user engagement prediction

New members to a service are often active for a brief period of time, then quickly become inactive and never return or else they could become lurkers on the website who don't contribute any positively. Early identification of new users likely to become inactive or potential lurkers helps a platform identify those who could be encouraged to continue seeking help, or become listeners to bolster its community. Feature importance analysis of such a classifier performing such predictions may also reveal the user behaviors and attributes that promote people to return and seek follow-up emotional support.

We consider a random forest classifier that identifies if a member who, based on actions during her first two weeks on 7cot, will become an active user. Since there is lack of a standard definition for an active user of a Web service, we consulted with 7cot administrators to define an active user as one who: (i) has been registered for at least six weeks; and (ii) has performed at least two actions on the service over the past month. We also define a new user as one who has registered within the last two weeks. We identified all members who registered between May 7th (the first date user action data was recorded) and November 28th, 2014 (the end of our data set) and mark them as active or inactive. We then collected the following actions they performed during their first two weeks on the site: (i) number of conversation requests and messages sent; (ii) number of forum posts made and viewed; (iii) number of logins performed; (iv) number of help page views; and (v) number of site pages accessed via 7cots Website and iOS app.

21% of the total members became active and 79% became inactive or potential lurkers during this time period. We created a training set by randomly sampling 66% of the registered members for a random forest classifier to predict if they are active. Trees are trained in a similar fashion to regression. Each tree yields its own prediction of if a member will be active or inactive given her actions during the first two weeks and a majority vote decides the class to be predicted. Due to the imbalance in the number of inactive and active members in the training data, we randomly oversample the minority class so that an equal number of inactive and active cases are provided for training, which is a common approach [28]. The trained random forest was tested over the 33% of users not considered in the training set. The classifier achieves a very promising accuracy of 92.5% and the ROC curve in Figure 5.4 demonstrates only a moderate false positive rate (ROC curves approaching the (0,1) corner of the plot are perfect classifiers; the y = x line represents a classifier that performs random guessing).

As before, we assess the importance of the factors used for predicting active users. Since the concept of MSE is incompatible with the notion of a binary classification decision, we instead consider the Gini index of decision tree nodes in the forest [28]. The Gini index measures the average gain of purity by splits of a given variable. A Gini index close to zero suggests that the splitting rule at the parent divides the data into separate classes, which is a property of strong decision tree classifiers. We

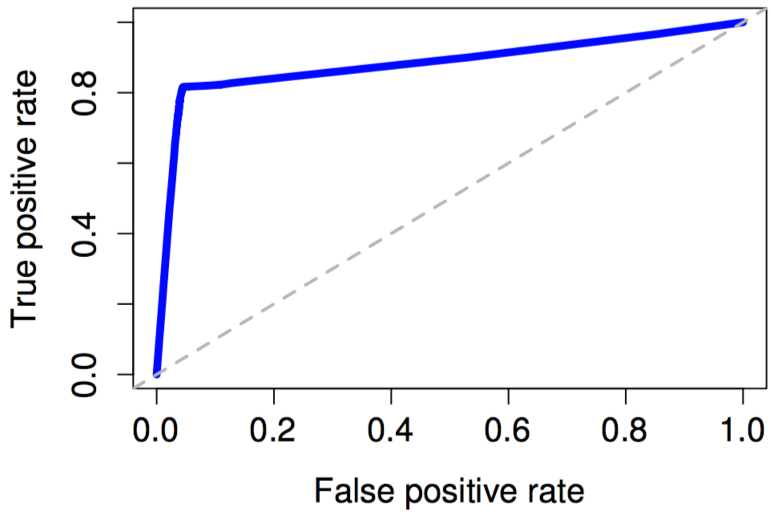


Figure 5.4: ROC Curve

thus rank the importance of a factor by the average decrease of the Gini coefficient across all splits in all trees of the forest that involve it in figure 5.5. It reveals that the number of messages sent in conversations and conversation requests submitted within the first two weeks are the actions that best predict whether a user will become active. We further examine the interaction between these two features by plotting the percentage of new users who became active and submitted greater than x messages in their first two weeks in Figure . Each trend corresponds to subsets of members that also submitted less than the specified number of conversation requests. It shows how for small numbers of conversation requests, the total number of messages sent in one-on-one conversations strongly influences members to become active. But once approximately five conversations are created, the number of messages sent in a conversation loses its importance. This may be because new users that connect with greater numbers of listeners feel more obligated to return to these connections again in the future. On the other hand, when a user connects to only a few listeners, a stronger bond between them (i.e. more messages shared) is necessary to drive the member to return to the site.

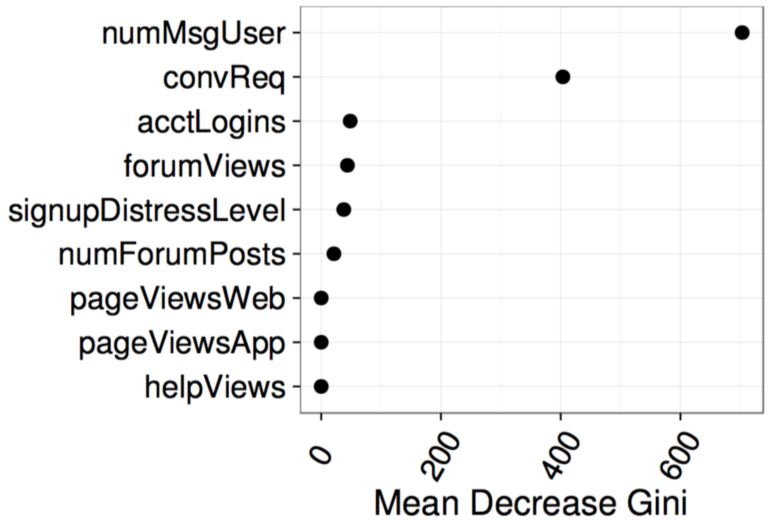


Figure 5.5: Gini Coefficient

Figure 5.6 also shows that the number of account logins performed, the users distress level, and activity related to the online forums within a members first two weeks are not major predictors of her becoming active. The frequency with which a member accesses the platform is thus unrelated to whether she will become an active member; what matters is not the number of times a member visits, but the quality or productivity of those visits as measured by the number of messages send and conversations requested. Furthermore, since members are equally likely to become active no matter their distress level, people suffering from both basic and complex problems may be equally willing to become active in online emotional support platforms. Finally, public spaces to post messages, such as forums, do not encourage new members to become active ones. This may be because forums serve as a less personal, more public medium of communication.

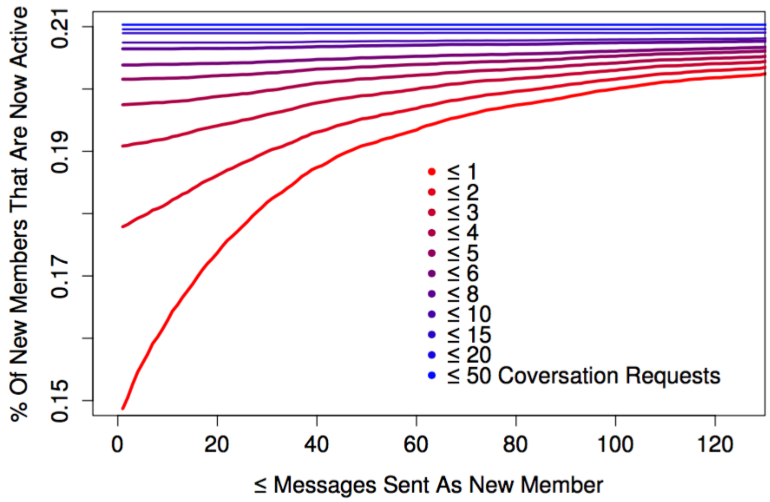


Figure 5.6: Percent of Active Members retained

5.3 Summary of Findings

In summary, this chapter identified a number of important insights about user engagement on this platform. Key takeaways from this study are :

- Users tend to exhibit similar level of engagement on the website irrespective of their emotional distress level or type.
- The gamification mechanisms integrated in the platform (for eg. coins, growth points etc.) are very important for the user to stay engaged on the website. For instance, the listeners may find it very interesting to constantly try and level up their rank thereby gaining recognition and attention by the members on the website. This gives them additional motivation to get highly involved on the website apart from volunteering to help members facing distress.
- Number of messages sent in conversations and conversation requests sent within the first two weeks best predict whether the user will become active in the long term.

6 Cyberbullying Analysis

This section presents our data-driven analysis of cyberbullying on 7cot. While 7cot and other emotional support systems are successful, their members and the listeners on such sites are especially vulnerable targets for cyberbullies. Members, who are already visiting to address an emotional need due to a distress would be taken to even more serious levels of stress and depression by a listener who cyberbullies. When listeners are exposed to cyberbullying, it is not only the health of the user that suffers, but also the health of the entire platform. This is because listeners who are harassed by nefarious members may suddenly become demotivated to participate on the site and opt to no longer volunteer their time or service. This loss of listeners, who are in most ways the lifeblood of a strong emotional support system, may have long term, possibly even fatal effects on the viability of the platform. It is therefore, important to begin to unearth information about the nature of cyberbullying on emotional support platforms, and in particular, to learn about the extent to which harassment occurs, the effects of this harassment on users and on the platform, and whether basic methods to curtail harassment are effective. The analysis is anchored around two important research questions :

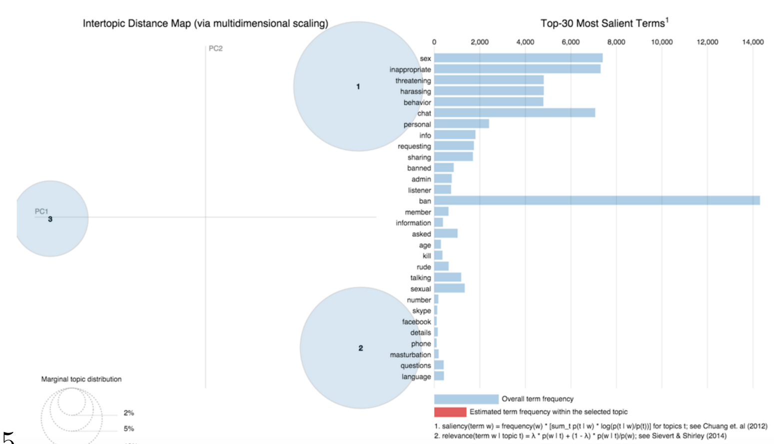
1. Is cyberbullying prevalent on an emotional support service? If so, what broad types of cyberbullying are occurring?

2. Does the act of blocking a bully from conversing with another change their negative behaviors?

While there is support in the research literature and popular press that cyberbullying is present, the levels to which they are prevalent in a social system has not been explored from a data driven perspective. Finally, checking if and how interventions widely adopted in social systems, such as electing to block someone, has an effect on the behavior of cyberbullies help us evaluate if new safeguards are even necessary. While previous studies on cyberbullying focuses on the effects of the practice on individual victims or even on the offenders of cyberbullying, which is an extremely worthy endeavor [29]. Our study takes the unique perspective of exploring the effect of cyberbullying on the overall health and viability of the online social system itself. We do so by evaluating how listeners may be bullied and discouraged from participating in the service by members.

6.1 How Much and What Kinds of Cyberbullying Exist ?

We first examine the degree and ways in which harassment occurs on 7cot. Each 7cot listener has the opportunity to block a member from communicating with them in a one-on-one conversation on the site. We postulate that block events only occur if a member bullies a listener during such a conversation. This is because members and listeners can only communicate one-on-one through conversations and because



55

Figure 6.1: Intertopic map of LDAVIS clusters

Users	Count	Percentage
Num. Members	452,605	-
Num. Listeners	169,372	-
Conversations	3.2 M	-
Cyberbully Members	19,281	4.26%
Listeners exposed to a cyberbully	37,262	22%

Table 6.1: Summary of Users and Cyberbullies

blocks can only be applied in the conversation interface. When a listener elects to block a member, he/she is allowed to enter a note as to why the block is applied. 7cot administrators regularly review these notes to identify members that should be banned from the site for constantly bullying others.

Table 6.1 provides a summary of the volume of members and listeners who are or were exposed to a cyberbully. For this table, we define a cyberbully as an individual who was either blocked by at least one listener in a one-on-one conversation. We identify a total of 19,281 members who were either blocked or banned during this time period, representing 4.26% of all members on the site. While this percentage appears to be small, these 4.26% of members actually held a one-on-one conversation with 37,262, or 22%, of all listeners on 7cot. Thus, even though a small number of members may have performed an action that led to a block or a ban but they run the risk of exposing a large proportion of listeners to emotionally damaging behaviors.

Notes left by listeners when they elect to block a member provides some insight into how and why a listener may have been harassed by the cyber- bully. We employ an extension of Latent Dirichilet Allocation (LDA) [30] called LDAVIS [31]. LDA is a learning algorithm that partitions documents in a corpus into clusters, with clustered documents containing a similar distribution of words. Evaluation of the words in the cluster are thus suggestive of a latent topic or theme in the documents. LDAVIS extends LDA by choosing to filter away words that appear across a number of topic clusters based on a measure of the relevance of a term to a topic. The relevance of a term to a particular topic is calculated by adding two quantities. The first quantity is the probability of a term under a particular topic and the second quantity, called lift, is defined as the ratio of a terms probability under a topic to its marginal probability across the entire corpus. This quantity is used to decrease the the rankings of globally frequent terms and gives higher rankings to terms that occur within the particular topic. Some other notable features of this tool are - We can see how words could be used in different context by hovering over a specific word which shows the different topics where the particular word has been used.

Figure 6.1 presents an intertopic distance map of the clusters found by LDAVIS. In this map, a topic is represented as a circle whose centers are positioned by a distance measure between topics. This distance is defined by a multidimensional scaling the projects the term space onto two dimensions [32]. The visual readily identifies three clusters, each of which has moderate size. The clusters are very far from each other, and interestingly, cover distinct quadrants. This is a strong indication that the terms of each cluster are very distinct, i.e., that there are three prevailing themes of the kinds bullying performed by members. Topic 1 refers to sexual harassment whereas, topic 2 and topic 3 refers to aggressive behavior and trying to acquire personal information respectively on the cluster map. To evaluate the topics, we present the most relevant words in the following figures. Table 6.2 includes words like sex, porn, horny, naked, and dirty, which is suggestive of online sexual harassment. Sexual harassment is also the most dominant form of bullying on the site, as 45.1% of words across the entire set of blocking notes contain words seen in this document cluster. The distribution of relevant words seen in Topic 2, given in table 6.2, include abusive, insulting, rude, angry, and swearing, suggesting that members also bully my sending aggressive and negative messages. 39.5% of words in the set of notes are seen in this cluster. Words in the final topic cluster as shown in the table 6.2 are personal, information, skype,

Cyberbullying Theme	**Top Words**
Sexual Harassment	Sex, Porn, Horny, Dirty, Naked
Rude/Aggressive Behavior	Abusive, Insulting, Swearing, Angry, Rude
Personal Contacts	Personal, Info, Location, Skype, Phone

Table 6.2: Top words for each theme

Cyberbullying Theme	**Notes**
Sexual Harassment	Asking for sex.
Rude/Aggressive Behavior	Very rude, negative attitude.
Personal Contacts	Asking about where I lived and my last name.

Table 6.3: Examples of block notes

phone and location, suggesting that the blocked members also try to acquire personal and off-site contact information from listeners, which is a very dangerous behavior. 15.4% of all the words in the set of notes are seen in this cluster.

The topic modeling analysis clearly classifies cyberbully behaviors into three broad categories which are sexual harassment, rude/aggressive behavior and trying to acquire personal information. This results are in harmony to Finn *et al.* who found that a majority of online harassment was reported to be from unwanted pornography and by threats and insults [33]. While these common kinds of attacks emerge on 7cot, the data shows that acquiring personal information is another major cyberbullying theme. We are therefore not much surprised as the same themes of cyberbullying are observed on online platform providing emotional support which are observed on any other online social platform. Table 6.3 shows some examples of each of these cyberbullying theme which was provided as notes by the listeners when blocking cyberbullies.

6.2 Is Blocking Bullies Enough ?

Finally, we study the impact of blocking a bully on their behavior. Figure 6.2 shows a box plot of the average number of conversations bullies hold in the day

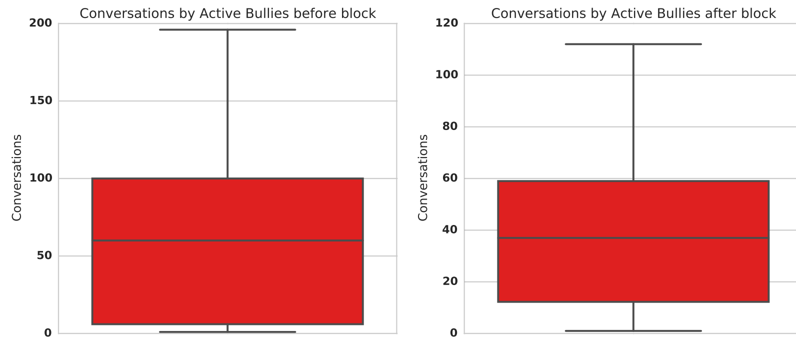


Figure 6.2: Conversations of Bullies before and after block(s)

prior to and day following a block event. Figure 6.3 also shows a violin plot of the same. The violin plot better visualizes the distribution of the data, while the boxplot shows its median and quartiles. The mean median and standard deviation of conversations per day before a block event was 59.60, 60 and 49.25 respectively. On the other hand, the mean, median and standard deviation of the conversations per day after the block(s) was 38.87, 27.48 and 37 respectively. This shows that prior to a block event, the average number of conversations per day were more and the number of conversations per day were more varied as compared to after the block(s) were issued to the bullies which suggest that prior to the block(s) event the bullies may be initiating more conversations on an average per day, with more new listeners and trying to attack a particular listener which they feel may be vulnerable. But, this exploratory behavior becomes restricted as the bullies are discouraged to future conversations by being blocked. This also tell us that standard measures used to deal with cyberbullying like blocking is really effective on this platform.

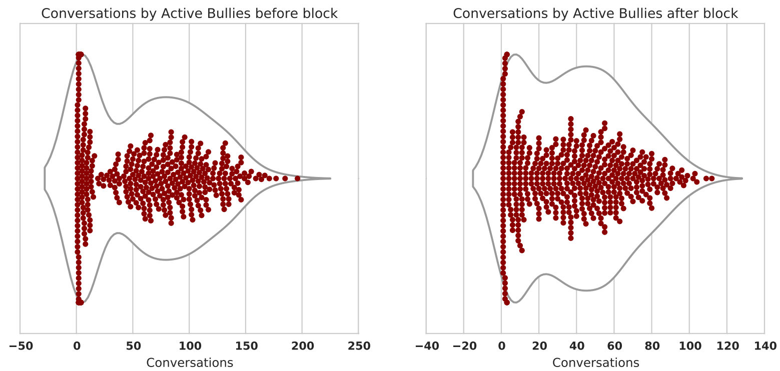


Figure 6.3: Conversations of Bullies before and after block(s)

6.3 Summary of Findings

The key takeaways from this chapter are as follows :

- Cyberbullying is the worst kind of evil born out of online social networks. No matter the online social system, cyberbullying should be expected to occur and 7cot is no exception.

- Language analysis of the notes which are provided when a cyberbully is blocked revealed three distinct themes a cyberbully uses to harass the users, which are sexual harassment, rude behavior and trying to acquire personal information. Past research also revealed such themes for cyberbullying a user, so the themes of cyberbullying can be generalized into a few types. However, getting blocked for trying to acquire personal information looks unique on 7cot.

- The standard measures adopted to thwart cyberbullying on the platform such as blocking or banning are effective.

7 Summary of Findings and Future Work

Online social networks will only gain popularity in the future for connecting with friends, families and dear ones. This thesis presented a data driven approach to shed light on how users of leading emotional support service choose to connect with each other, how they behave, utilize the platform, what design choices and user behavior is responsible for high user engagement, discovered the frequency, kinds, and characteristics of cyberbullying on this system and investigated the effect of blocking a cyberbully, which is the typical method of thwarting their behavior on the social service. We summarize our findings as follows:

1. As the goal of the platform is to provide emotional support and empathy to those who need it, the process of registering as a member on the website could actually lead to more positive health outcomes as compared to not registering and using the platform merely as a guest.

2. The gamification mechanisms integrated in the website are greatly responsible for user engagement.

3. The structural characteristics of network of this platform leads to some unique and useful insights. The giant connected component of the bipartite network include almost all the users of the platform, which tells us that only few users choose to exclusively search for and speak with each other. The degree distribution of listener projection network follows a power law distribution but same may not be said for the degree distribution of member projection network. This suggest that member

may tend to develop deep and strong relationships with some of the listeners rather than having an "exploratory" behavior where they are trying to connect with as many listeners as possible. The clustering coefficients of the member and listener projection appears to be normally distributed as is seen in many co-occurrence networks. A small percentage of members and listeners exhibit perfect clustering coefficients which is unique to this platform. [1]

4. The menace of cyberbullying exist on this platform as well. Types of cyberbullying on this platform can be thematically described as sexual harassment, rude behavior, and as trying to acquire personal information.

5. The worst bullies that are blocked multiple times are also extremely active on the platform and are seasoned users which suggest that such worst bullies initiate a comfortable repertoire with a victim before attacking them personally. It therefore makes it difficult to conclude whether heavy tail is a natural phenomenon in an emotional support system, or if it is because of behavior of such worst bullies and some exceptional normal users.

6. The act of blocking has a strong effect on bullies, with a larger average number of conversations and larger standard deviation of conversations per day prior to a block event as compared to when the block(s) had been issued against them. This suggests that block could act as a deterrent to cyberbulling for most of the bullies.

These findings have many implications about how people use crowdsourced emotional support systems. As mentioned before, the member projection network doesn't quite follow power law distribution which suggest that their behavior may not be exploratory in nature where the objective is to connect with many listeners possible but on the contrary it suggests that they may have thoughtful conversations with

[1] Portions of this thesis were previously written and published by the author in [2].

only few listeners and prefer to connect with them which suggest that members tend to choose to have conversations with their listeners very carefully, maybe only after reviewing their profile, ratings, reviews and overall experience. This may be due to the fact that they are not comfortable to open up about their problems to any listener and therefore prefers to connect with very experienced listener only. So selection of a listener for having a conversation looks highly competitive. This may also suggest that the members on the website are very serious regarding their emotional problems and to find solutions for it. This looks a positive behavior that the members are extra cautious while connecting with a listener and so in a way are more likely of not facing cyberbullying by a nefarious listener as compared to an unregistered guest who may opt to connect to whichever listener available for conversation. On the other hand, the degree distribution of listener projection network follows a power law distribution suggesting that some listeners are willing to help and support to as many members as possible. Also the fact, that listeners are able to support for number of different emotional problems again reaffirms their willingness to help multiple users facing multiple emotional problems. Engagement analysis also demonstrates the importance of gaming mechanisms for tracking user progress, which may serve as additional motivation for the users to help each other. Cyberbullying analysis also explored the problems of cyberbullying on this platform and investigated the effectiveness of standard measures adopted to thwart cyberbullying. The findings show some commonalities between users using other online social networks and users using 7cot such as normally distributed clustering coefficients of the users, while at the same time it also revealed some unique insights which only the users of 7cot tend to follow. For eg. small percent of users having perfect clustering coefficients.

Future work will include studying how cyberbullying grows or spreads, for example if a person have been cyberbullied, how likely is it that he/she will cyberbully somebody else, or in other words does being a victim of cyberbullying encourages the victim to cyberbully someone else ? Does cyberbullying spread like a contagious disease or an epidemic on an online platform and if so how can it be contained ? We would study the networks of the bullies and that of the victims who got cyberbullied and will analyze whether the structural properties of these networks show significant distinctions which could lead to some new insights about the behavior of the cyberbullies as compared to normal users. We will also explore other machine learning models to improve engagement analysis.

REFERENCES

[1] Y. M. Binik, J. Cantor, E. Ochs, and M. Meana, "From the couch to the keyboard: Psychotherapy in cyberspace," *Culture of the Internet*, pp. 71–100, 1997.

[2] D. Doran, S. Yelne, L. Massari, M.-C. Calzarossa, L. Jackson, and G. Moriarty, "Stay awhile and listen: User interactions in a crowdsourced platform offering emotional support," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ACM, 2015, pp. 667–674.

[3] M. W. Newman, D. Lauterbach, S. A. Munson, P. Resnick, and M. E. Morris, "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*. ACM, 2011, pp. 341–350.

[4] H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra, "Detection of cyberbullying incidents on the instagram social network," *arXiv preprint arXiv:1503.03909*, 2015.

[5] M. C. Calzarossa, L. Massari, D. Doran, S. Yelne, N. Trivedi, and G. Moriarty, "Measuring the users and conversations of a vibrant online emotional support system," in *Computers and Communication (ISCC), 2016 IEEE Symposium on*. IEEE, 2016, pp. 1193–1199.

[6] D. Maloney-Krichmar and J. Preece, "A multilevel analysis of sociability, usability, and community dynamics in an online health community," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 12, no. 2, pp. 201–232, 2005.

[7] A. Barak, "Emotional support and suicide prevention through the internet: A field project report," *Computers in Human Behavior*, vol. 23, no. 2, pp. 971–984, 2007.

[8] B. Ploderer, W. Smith, S. Howard, J. Pearce, and R. Borland, "Patterns of support in an online community for smoking cessation," in *Proceedings of the 6th International Conference on Communities and Technologies.* ACM, 2013, pp. 26–35.

[9] E. K. Yuen, E. M. Goetter, J. D. Herbert, and E. M. Forman, "Challenges and opportunities in internet-mediated telemental health." *Professional Psychology: Research and Practice*, vol. 43, no. 1, p. 1, 2012.

[10] Y.-C. Wang, R. Kraut, and J. M. Levine, "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.* ACM, 2012, pp. 833–842.

[11] Y. Zhang, D. He, and Y. Sang, "Facebook as a platform for health information and communication: a case study of a diabetes group," *Journal of medical systems*, vol. 37, no. 3, pp. 1–12, 2013.

[12] V. Sadovykh and D. Sundaram, "Decision making and support in healthcare online social networks," in *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).* IEEE, 2015, pp. 1048–1052.

[13] A. Saha and N. Agarwal, “Demonstrating social support from autism bloggers community on twitter,” in *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2015, pp. 1053–1056.

[14] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, “Measurement and analysis of online social networks,” in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 2007, pp. 29–42.

[15] W. Han, X. Zhu, Z. Zhu, W. Chen, W. Zheng, and J. Lu, “Weibo, and a tale of two worlds,” in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ACM, 2015, pp. 121–128.

[16] M. Newman, *Networks: an introduction*. Oxford university press, 2010.

[17] J.-L. Guillaume and M. Latapy, “Bipartite graphs as models of complex networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 371, no. 2, pp. 795–813, 2006.

[18] N. K. Ahmed, J. Neville, and R. Kompella, “Network sampling: From static to streaming graphs,” *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 8, no. 2, p. 7, 2014.

[19] D. Doran, “Triad-based role discovery for large social systems,” in *International Conference on Social Informatics*. Springer, 2014, pp. 130–143.

[20] A. Clauset, C. R. Shalizi, and M. E. Newman, “Power-law distributions in empirical data,” *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.

[21] L. Lipsky, *Queueing Theory: A linear algebraic approach*. Springer Science & Business Media, 2008.

[22] M. P. Rombach, M. A. Porter, J. H. Fowler, and P. J. Mucha, "Core-periphery structure in networks," *SIAM Journal on Applied mathematics*, vol. 74, no. 1, pp. 167–190, 2014.

[23] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-worldnetworks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[24] J. Yang and J. Leskovec, "Defining and evaluating network communities based on ground-truth," *Knowledge and Information Systems*, vol. 42, no. 1, pp. 181–213, 2015.

[25] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: Densification and shrinking diameters," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 1, no. 1, p. 2, 2007.

[26] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," *ACM Transactions on the Web (TWEB)*, vol. 1, no. 1, p. 5, 2007.

[27] X. Niu, X. Sun, H. Wang, S. Rong, G. Qi, and Y. Yu, "Zhishi. me-weaving chinese linking open data," in *International Semantic Web Conference.* Springer, 2011, pp. 205–220.

[28] T. Hastie, R. Tibshirani, and J. Friedman, "Unsupervised learning," in *The elements of statistical learning.* Springer, 2009, pp. 485–585.

[29] S. Hinduja and J. W. Patchin, "Bullying, cyberbullying, and suicide," *Archives of suicide research*, vol. 14, no. 3, pp. 206–221, 2010.

[30] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[31] C. Sievert and K. E. Shirley, "Ldavis: A method for visualizing and interpreting topics," in *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, 2014, pp. 63–70.

[32] J. Chuang, D. Ramage, C. Manning, and J. Heer, "Interpretation and trust: Designing model-driven visualizations for text analysis," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012, pp. 443–452.

[33] J. Finn, "A survey of online harassment at a university campus," *Journal of Interpersonal violence*, vol. 19, no. 4, pp. 468–483, 2004.