Unravel the Drivers of Online Sharing Communities: An Empirical Investigation

Mu Xia
Department of Business Administration
College of Business
University of Illinois at Urbana-Champaign
Champaign, IL 61820
USA
mxia@uiuc.edu

Wenjing Duan
Information Systems & Technology Management
School of Business, George Washington University
Washington, D.C. 20052
wduan@gwu.edu

Yun Huang, Andrew B. Whinston
Center for Research in Electronic Commerce
McCombs School of Business
University of Texas at Austin
Austin, TX 78712
USA
{yun@mail.utexas.edu, abw@uts.cc.utexas.edu}

Abstract

Recently user-oriented online sharing communities have seen explosive growth. Two characteristics of these communities set them apart from traditional online message-based communities such as online forums. First, users have no social ties before joining the community. Second, there is little or no “verbal” communication between users. This research investigates the structure and dynamics of online sharing communities using data collected from an IRC music channel from 2001 to 2006, covering all five years of the post-Napster age. We have collected more than three hundred million individual activities, capturing 0.05% of the global music sharing volume. We find that sharers are an essential part of the community and their activities have a dominant impact on the growth of the community. By contrast, free riders have two opposite impacts on sharer retention. More free riders in *number* make it more likely for a sharer to keep sharing, while more free rider *activities* discourage sharers from contributing. That is, the existence of free riders, despite the congestion caused by their download activities, does to some degree stabilize the community. Most previous literature examines the online community only from the aggregate level. Our study, nevertheless, distinguish the influence and behavior of different members in the community. Instead of paying only attention to the total number of users, our results suggest that understanding the impact of their core members is critical in investigating the dynamics and the sustainability of online sharing communities.

1. Introduction

As the Internet continues to enjoy growth in both its reach and bandwidth, new user-oriented online sharing communities have seen explosive development. Take YouTube.com, a web site for people to share and download videos. Debuted in May 2005, it attracted 9 million unique US visitors in February 2006. In April 2006, 35,000 new videos were posted and 35 million videos were watched daily (Liedtke 2006). Other similar popular user-contributed sharing web sites include Flickr.com, a public picture-sharing site, and KaZaa, where members share music.

An important feature sets these communities apart from traditional online forums where members exchange messages--there is little verbal communication among users. Also notable is the fact that these communities are completely built online. The users have no social ties among themselves. As a result, users communicate in a very detached mode, mainly through observing other users' activities. Such loose ties among users seemingly subject the sustainability and dynamics of the community to a myriad of random factors. Indeed, many communities could not survive long after their launch[1]. Even when a community experiences growth as explosive as YouTube, it is not clear how these communities work. Researchers have very limited understanding for the fundamental question "how do members' behavior, interactions and activities affect the dynamics and sustainability of the online community?"

As more web applications have made it easy to develop such voluntary sharing communities and as a result they mushroomed, the need to understand has never been so pressing. For entrepreneurs, knowing the answer to the above question will help them to better plan and manage a community after its launch and maximize its value for all members. For investors, valuing a

[1] For example, the community site theglobe.com, at its height, boasted 9.3 million registered users and its stock price surged by 600% at its initial public offering (IPO) history (Bobala 2001). Yet within three years the site closed its doors.

burgeoning community's long-term investment prospect requires understanding of this fundamental question about the viability of the community. For researchers, there is so little work studying the new phenomenon that it is not clear if existing theories will apply and how.

In this paper, we aim to take the first step to tackle the question by analyzing a large data set of 300 million individual activities in Internet Relay Chat (hereafter IRC) sharing channels from March 2001 to May 2006. This five-year log dataset allows us to observe individual user activities and study the research issue at both the aggregate and individual user levels. At the aggregate level, we examine the dynamics of different user types and how their activities influence the community variation. To complement and deepen the insights generated from aggregate level analysis, we further examine how sharer's individual behavior affected by their own activities and the community as a whole.

We have several findings. First, sharers are the core part of the community and their behavior has a dominant impact on the growth of the community. By contrast, free riders have two opposite impacts on sharer retention. Their *size* has positive impact—the more free riders in number, the more likely sharers will keep sharing, whereas their *download activities* have a negative impact—the more free riders download, the more likely sharers will stop sharing. Although seemingly surprising, this result reveals the true payoff of a typical sharer—he is happy to see more people enjoy his contribution as long as his own download activities are not affected. We are also able to understand the role of sharer size on sharer retention and find that, when taking into account of a sharer's own download, the total sharer size actually has a negative impact on sharer retention. We suspect this is due to the disproportional download thus network congestion sharers caused.

This paper contributes to the literature in four ways. First, our unique dataset's time span and level of detail surpass existing IS literature studying online communities. The richness of the dataset affords us the freedom and flexibility to examine every measurement reasonable from different angles. Second, while online communities have been attracting researchers' interest, we are the first to identify the effect of the community structure, including member type and activities on the community. Third, our results help deepen the understanding of non-verbal sharing communities, an emerging and important phenomenon fueled by the increasing ease of creating such communities and ever-increasing involvement and contribution of individual users online. Fourth, the insight can assist entrepreneurs and investors to better manage and evaluate online communities.

The rest of the paper proceeds as follows. We review relevant literature in Section 2. In Section 3 we propose the theoretical framework and present our hypotheses. Section 4 describes the data set, introduces the model and presents the results of data analysis. In Section 5 we discuss the findings and their managerial implications, followed by some future research directions.

2. Literature Review

As mentioned in the introduction, an increasing number of online sharing communities are characterized by non-verbal communication and little social context. The only possible impact the community as a whole has on individual members is through the individual member's observation of the system characteristics and aggregate behavior of all other users. While no literature directly tackles the research question raised above in this particular setting, there are some papers studying related issues in environments that have some common characteristics with sharing communities.

One such community is the online message-based community, which the Information Systems literature has been investigating for some time. Among the first systematic studies, Butler (2001) examines the role of membership size and communication activity on online social structures. He proposes a resource-based model that treats membership size and communication activity as resources and benefit providers of the community, as well as costs incurred to members. He applies the model to data from a number of email lists and finds that as membership size grows, the community experiences a faster “churn” rate, i.e. the percentage of membership loss also increases. Similarly, the impact of communication activity on members is also a double-edged sword. While more communication provides more value to members in general, its benefit is not valued uniformly. He cautions theorists and developers of online social structures to be aware of the opposing forces and the endogenous nature of membership size and communication activity, as well as their interplay, and adjust their expectations of the growth of a community accordingly.

Other related work on online social structures studies individual messages to investigate information overloading and its effect on user communication in online forums (Jones et al. 2004), to predict the likelihood of users receiving replies to their posts, and the probability of users staying in a community (Arguello et al. 2006). Unfortunately, when the target community’s communication is non-verbal based, one cannot resort to analyzing messages. Therefore, the findings of these papers cannot carry over. Another approach is to survey participants. Nonnecke et al. (2006) investigate the reasons why people lurk, i.e. participate in an online message-based community but never post. They administer a survey to both posters and lurkers in MSN communities and find that lurkers have many diverse reasons for not posting. Topping the list is “just reading/browsing is enough”. This is exogenous to the community and can be due to various personal characteristics of the user. For example, the subject matter of the community may be just a side hobby for the lurker that does not deserve the time and attention required for active posting. The second most cited reason is “still learning about the group”. This suggests that some of the lurking behavior is temporary and adaptive. Thus it is an endogenous factor that can be influenced by the community. Both of these findings may also apply to our setting to explain the dynamics of sharing communities, since they are not directly based on messages.

To the best of our knowledge, Asvanund et al. (2004) are the first to empirically study online networks with non-message-based communication. They specifically focus on network externalities of peer-to-peer (P2P) music-sharing communities. As with Butler (2001), they recognize the fact that as the network grows, so do the benefits (more resource availability) and costs (network congestion due to user free riding). By sending queries to six P2P networks, they collect query congestion, song availability and download delay data to measure the exact value of positive and negative network externalities. Although in a different setting, their result is similar to Butler's (2001) in that the marginal value brought by an additional user declines in larger networks, while the marginal costs imposed by the new user increases with size. Again, the results suggest that there is an upper bound to optimal size of a centralized P2P network. In addition, their results also show that, consistent with the public economics literature, the free-riding effect is significant in larger networks.

Another stream of literature is on the incentives to contribute to online communities. In the Information Systems area, research on contribution to online messaged-based communities identifies altruism and reciprocity as the primary drivers of contribution (Wasko and Faraj 2000, Gu and Javenpaa 2003, Peddibhotla and Subramani 2005). In a non-verbal setting such as P2P networks, researchers seem to agree that excessive free riding behavior should be discouraged and have devised mechanisms to encourage sharing (Ranganathan et al. 2004, Krishnan et al. 2004). Another setting also similar to our sharing community is open source software development, where programmers voluntarily write code to serve the needs of themselves and other potential users, the majority of whom rarely communicate verbally with the developers. There is a large body of literature that studies the incentives to develop when there is little direct financial benefit. Again, besides satisfying user needs (Raymond 2000, Lakhani and Wolf 2005), reciprocity (Kollock 1999, Shah 2006) and altruism (Torvalds 1998, Hars and Ou 2001) are the next most cited motivations to contribute. It must also be noted that most of the research is based on surveys that inevitably may be biased.

In our setting, we anticipate that both reciprocity and altruism serve as main incentives for sharers to contribute. Even though our focus is not on the contribution incentives, we can still examine the effect of the actual downloading behavior of other users to the sharers, and vice versa. In other words, our intention is not to study "*why* do people contribute?" but rather "*how* are people's contributing behavior affected by those using it?" and "how is the size of all users affected by the size of the contributors and free riders?"

3. Theory and Hypotheses

3.1 Two Types of Users

In voluntary, anonymous online sharing communities, inevitably, there will always be users who contribute and who only consume content. Extant research in online communities recognizes different types of users and observes their disparate behaviors. For example, Adar (2000) detected that "almost 70% of Gnutella users share no files, and nearly 50% of all responses are returned by the top 1% of sharing hosts". Yet, there is no systematic empirical study that tests the differences between user groups in their influence to the community as well as how the community impacts their behavior.

In sharing networks, there are two distinct types of users: those who make their content available for downloading by other users, and those who do not. We call the former user group sharers, and the latter free riders.

Sharers, through the software installed on their computer, often provide a collection of files and announce their availability to other users. In some networks, sharers are able to configure how their content can be retrieved. At the aggregate level, they can decide whether to make the content available and for whom. For example, they can restrict the download speed, the group of user requests they respond to and total bandwidth reserved for other users. Like free riders, sharers also download from other sharers themselves.

Much of the extant literature *assumed* that free riders only have negative impacts on the community. By definition, they do not share any content with other members of the community. Therefore they only consume the resource without contributing to it (Butler 2001). Because their downloading behavior causes more congestion, they are the source of negative network externalities (Asvanund et al. 2004). However, it is not known if and how their behavior has any positive impact on the community. In the extant literature on online communities, due to the lack of long-term observation of individual activities, the exact impact of free riders on the community, especially on sharers, could not be measured.

3.2 Extending Existing Research

In the literature review, we mentioned two papers in the Information Systems area that study online communities. Our theory will integrate and extend their view.

Butler (2001) employs the resource-based view to study the sustainability of online communities. He takes membership size and communication activities as measurements of community resources and recognizes that they have both positive and negative impacts on the community. However, because of the aggregate nature of the model, the exact cause of the positive and negative impacts of the two variables cannot be pinpointed.

Asvanund et al. (2004) examine the marginal positive and negative network externalities each additional user provides in a peer-to-peer (P2P) network. They posit that positive network externalities arise when “users who choose to share their content” bring new content, replicas of existing content, or other shared resources to the network, while P2P users impose negative network externalities when “consuming scarce network resources’’ and “without providing resources back to the network in return”. Because of the decentralized nature of the content and user activities, they only observe the status of sharing networks, including availability and congestion, but not actual user activities. Moreover, since they do not distinguish sharers and free riders, they cannot identify the sources of positive and negative network externalities and the exact impact of different types of users.

While the above two papers study seemingly unrelated phenomena, the underlying question is the same: how do users and their activities affect the community and how does the community affect the users? Based on the resource-based model of Butler’s, sharers contribute resources to the community as well as consume resources. Free riders, on the other hand, only drain resources of the community. In other words, free riders provide negative network externalities, while sharers’ network externalities are the result of the two opposing forces. Asvanund et al.’s (2004) results show that on average, the benefit from an additional user (sharer or free rider) to the community decreases as the network size increases, and the cost to the community increases. Since only sharers provide benefits and free riders far outnumber sharers in P2P networks, the estimated benefit of *any* additional user is well below the actual value brought by a sharer. As to negative network externalities, the impacts come from both sharers and free riders, as both groups cause congestion when downloading. With the help of detailed log data of user activities, our model aims to integrate both models and identify the exact contribution and cost of sharers and free riders to the community, as well as to tease out the composition of the overall effect of user size and activities.

Before analyzing individual member activities data, we first test the same set of hypotheses from Butler (2001) on the aggregate data. Even though unlike a mailing list, there is little verbal communication in IRC sharing channels, the concept of resources still applies. Corresponding to mailing list subscriber size and activities are the total user size and downloading activities, both of which were observed. Like the impact of subscriber size in a mailing list, we expect that the total user size has a positive effect on both membership gain and loss. The larger the number of users in the community, the more resources, in this case the number of files, each user has access to. This makes it easier to attract new users. At the same time, more users also mean more traffic and congestion for the existing users, thereby reducing their benefit. For some users, when the size grows to a level at which it is no longer worth it to stay in the community, they will choose to leave.

Based on Butler's (2001) model, we have the following set of hypotheses:

Hypothesis 1a: The total user gain, i.e. the number of new users that joined in each period, is positively affected by total number of users.

Hypothesis 1b: The total user gain is positively affected by total downloading activities.

Hypothesis 1c: The total user loss, i.e. the number of users that left after each period, is positively affected by total number of users.

Hypothesis 1d: The total user loss is positively affected by total downloading activities.

Butler treats the total member size as a resource of a community. However, we choose to differentiate those who provide content and those who only consume content when studying member impact on the community. The number of free riders should not be counted as resources, because they not only do not contribute content, but also often come just to find content and leave immediately after. For example, from many free riders' perspective, IRC sharing channels are just another tool to download music. The sharers, in contrast, provide resources to the community, and their number should be a measure of the total resources. In addition to the size of different user types, their activities are also a good measure of community resources and usage.

Our next group of hypotheses characterizes the dynamics of sharers. A sharer in one period can change their behavior in the next period. They may choose to become a free rider, who no longer contributes content, or leave the community altogether. This switching decision is affected by many factors, such as community size and resources, individual payoff, sharing cost, and community recognition.

To become a sharer in a non-verbal, voluntary sharing community, the user has to have a genuine interest in the subject matter and enjoy sharing his own content with others. Sharers contribute and consume content, even without verbal communication. They typically have a strong sense of community and view themselves as an important part. Their view of the free riders is two-fold. On the one hand, they do not mind the existence of free riders, as part of the sharers' utility is derived directly from how much and how often his/her contribution is consumed. In fact, it is often the case that the more people download from a sharer, the more satisfied the sharer. On the other hand, as in the classic economic tale of the Tragedy of Commons (Hardin 1968), too many free riders may lead to excessive downloading activities, which in turn cause network congestion that makes sharing and downloading difficult, defeating the purpose of the community. Some sharers may be so upset by the negative impact of free riders destroying the balance of the community that they choose not to contribute any more. But they are still interested in staying in the community because of their interests in the subject matter. In other words, the conversion from a sharer to a free rider is endogenous to the community.

Hypothesis 2a: The more resources and value a share brings to a community, the more likely he will continue to share.

Hypothesis 2b: When the quality if a community deteriorates, sharers are less likely to continue to contribute, i.e. congestion has a negative impact on sharer retention.

Compared to active posters in a message-based community, sharers in a non-verbal community incur a lower cost of contributing. Once a copy of the sharing software is installed, there is virtually no cost to join and leave a sharing community. Therefore, users' short-term absence from the community is often affected by some random factors that are mostly exogenous to the community. For example, the departure of a sharer may be due to lack of interest, computer failures or just being too busy at the time it happened. The short-term user loss and gain have little to do with the endogenous resources and activities of the community. However, in the case of prolonged absence, we expect that it is due to the impact of factors endogenous to the community, such as the imbalance of benefits and costs from the perspective of a sharer. Hypothesis (2c) summarizes the difference between short-term and permanent sharer size change.

Hypothesis 2c: The temporary changes of sharer size and long-term sharer gain and loss are affected by different user activities.

Free riders, like lurkers in an online forum or mailing list (Nonnecke et al. 2006), may have different reasons for their non-contribution. They may not view the online-based sharing application as a community. They may have higher costs to share; or they simply do not see their contribution as necessary for the community. Therefore, although the majority of the population are free riders, more free riders cannot produce more sharers because they are inherently different users. Nevertheless, free riders' consumption of resources, i.e. downloading activities, does cause congestion and hence has a big impact on sharers' behavior.

Hypothesis 3a: The size of free riders has no positive impact on the gain of sharers.

Hypothesis 3b: The size of free riders has no positive impact on the loss of sharers.

4. Data and Empirical Model

4.1 Data Description

Music sharing is the most popular form of sharing communities. Using on different software, millions of users exchange music files every week. In this research, we focus on music sharing in Internet Relay Chat (IRC) networks – one of the popular music sharing applications.

IRC is a form of instant communication over the Internet. It consists of a collection of topic-oriented chat rooms (channels) that also support instant user-to-user communication. Users can, with the help of scripts such as SDFind and OmenServe, set up file servers to make music files available in a channel for others to download. When another user searches for a particular music file by keywords, IRC scripts broadcast the search request to all the users logged in the IRC channel. If a logged-in user turns on the sharing switch, his scripts will automatically respond to the request and sends the download information about matching files.

Compared to specialized file-sharing applications, such as Napster, Gnutella, and Kazaa, three unique characteristics make IRC music sharing a good representative of sharing communities. First, IRC channels have a very long history, which covers the lifespan of other competing file sharing paradigms, including server-centric sharing, Napster, and second generation of P2P sharing (Asvanund et al. 2004). People started using IRC to share files from the early 90s and there are still many active users today. Second, the IRC sharing mechanism has not been changed after its release. Therefore, the mechanism itself should have no impact on changes of user behavior in the time period we observed. Third, IRC sharing is not only a search and download tool, but also supports user communication. Within a channel, a user can use keyword search to locate and download specific MP3 files. Unlike other P2P file sharing networks, users can also chat with others, browse others' file collections, and check file servers' status. These activities strengthen connections among users and form an online community.

From March 2001 to May 2006, we monitored and logged messages exchanged in the second popular IRC music sharing channel – mp3passion. Since all messages exchanged in the channel are publicly available, we wrote a script that collected individual activities and statistics of community resources, including find commands, download requests, the number of files shared, total contribution, and bandwidth[2]. Figure 1 shows the number of biweekly downloads during the data collection period. At its peak during the year 2003, more than one million files were exchanged every two weeks, which amounted to 0.05% of the global music sharing volume then[3].

---

[2] Due to connection dropout, the total time span of our log data is 30,286 hours, covering 72% of the 1,750 days of the entire logged period. We designed a non-linear weighting algorithm to calculate aggregate variables. The detailed method is available upon request.

[3] Around one billion songs are downloaded per week at the end of 2003 (Wingfield 2003).

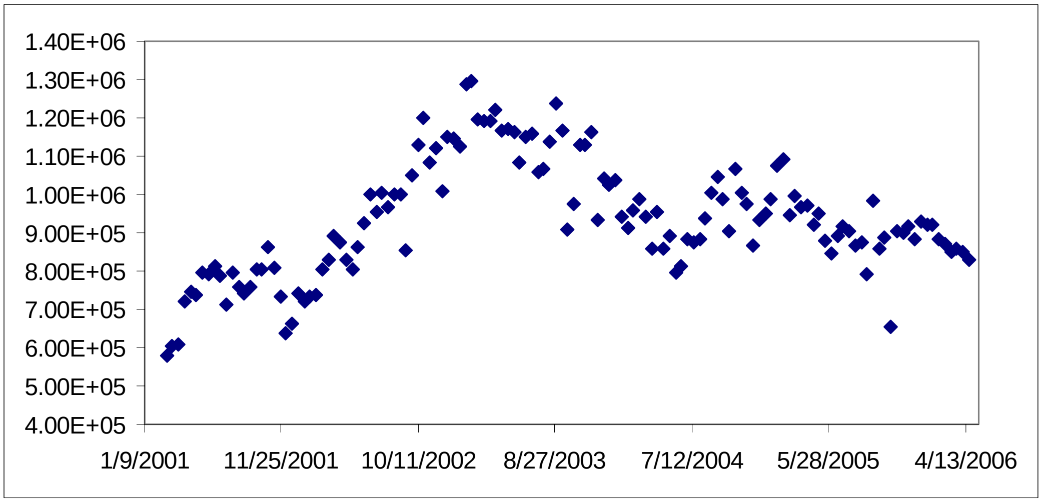

Figure 1: Numbers of downloads (biweekly)

With the status of the sharing function observable for each user, it is easy to distinguish sharers from free riders. Figures 2 and 3 depict the change of the numbers of sharers and free riders in the channel during our data collection period. Figure 2 shows a very clear S-curve growth pattern for sharers (Gurbaxani 1990). As to their download activities, we can see from Table 3, the number of downloads by sharers is highly correlated with the number of sharers. Therefore sharers not only contribute resources to the community they are also major consumers themselves. At 73 files per week, a sharer on average downloads five times as free riders do.

It is not surprising that free riders far outnumber sharers throughout the observation period (Figure 3), at times twenty times more than the number of sharers. However, their number remained relatively stable during the five years and has a small correlation with their download activities (see Table 3). This evidence is consistent with our theory that sharers are most important members and determine the value and growth of a sharing community.

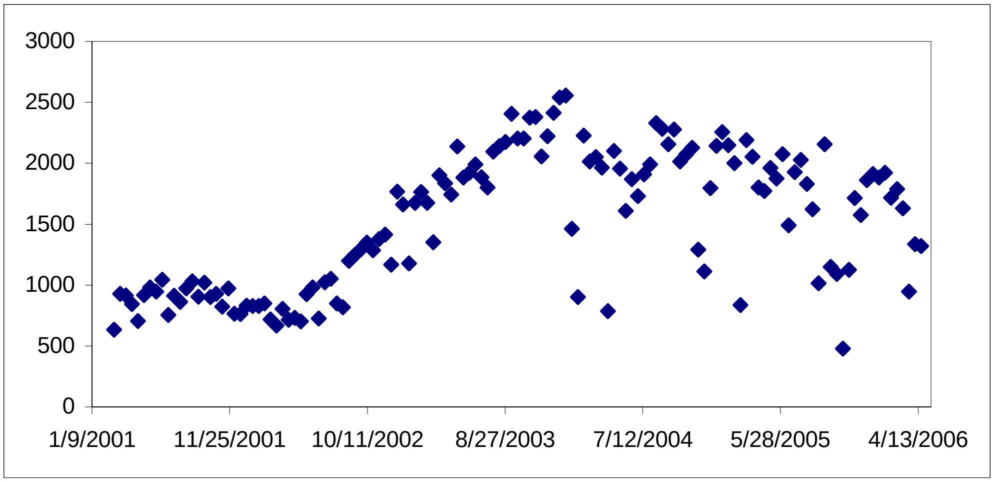

Figure 2: Numbers of sharers (biweekly)

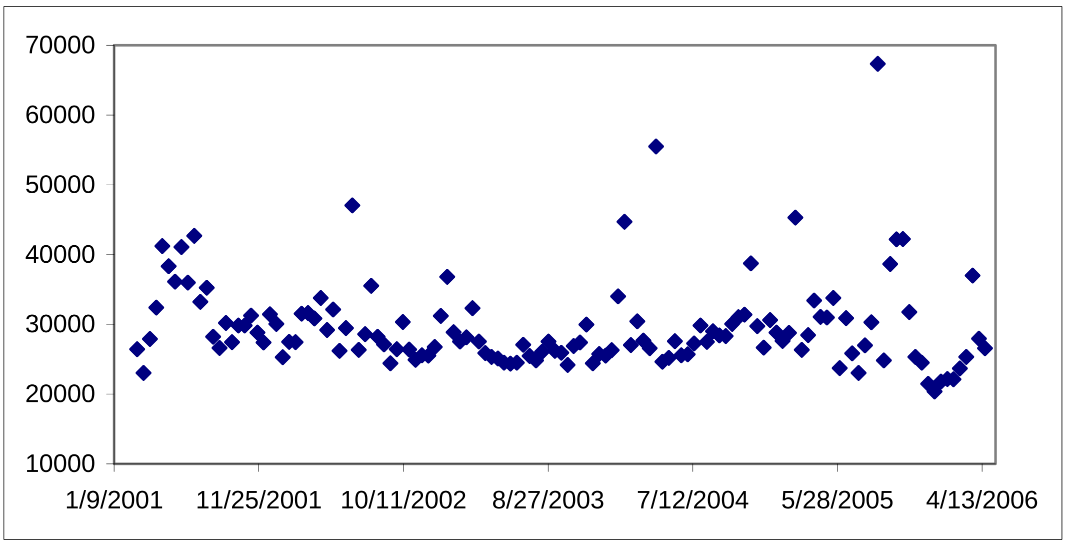

Figure 3: Numbers of free riders (biweekly)

Based on the log data, we compile lists of the identities (nick names) of sharers and free riders in every time period. Using these lists, we can observe the changes in user types and usage patterns, and could therefore track the gain and loss of each user group. For example, by comparing the lists of sharers in time period *t*-1 and time period *t*, we can identify the new sharers appeared in period *t* but not in period *t-1*. We record the number of these new sharers as $Sharer\_Gain_t$. Using a similar comparison, we can also identify the number of disappeared sharers as $Sharer\_Loss_{t-1}$.

Based on biweekly aggregation, we calculate the numbers of downloads, users, find requests, and member gain and loss. We also construct the same set of variables for sharers and free riders respectively. Tables 1, 2 and 3 show the description of key variables, their descriptive statistics, and the correlation matrix.

Table 1. Description of Key Variables

Variable	**Description**
Aggregate:	
$User\_Size_t$	The total number of users in the channel during period t
$Sharer\_Size_t$	The number of users who share music during period t
$Freerider\_Size_t$	The number of users who do not share music during period t
$User\_Download_t$	The total number of files downloaded during period t
$Sharer\_Download_t$	The total number of files downloaded by sharers during period t
$Freerider\_Download_t$	The total number of files downloaded by free riders during period t
$User\_Gain_t$	The total number of users in period t who were not present during period t-1
$User\_Loss_t$	The total number of users in period t who were not present during period t+1
$User\_Gain\_Cum_t$	The total number of new users who entered for the first time during period t
$Sharer\_Gain\_Cum_t$	The total number of new sharers who entered for the first time during period t
$Sharer\_Gain_t$	The total number of sharers in period t who did not appear during period t-1
$Sharer\_Loss_t$	The total number of sharers in period t who were not present during period t+1
$Sharer\_Gain\_Permanent_t$	The total number of sharers in period t who did not share in the previous six time periods
$Sharer\_Loss\_Permanent_t$	The total number of sharers in period t who did not share in the next six time periods
$Sharer\_Sharer_t$	The number of sharers in period t who continue to share during period t+1
$Sharer\_Freerider_t$	The number of sharers in period t who become free riders during period t+1
Individual sharer:	
$Decision_{it}$	Sharer *i*'s choice of action after period t: Share, Free ride, or Leave
$Download_{it}$	The total number of files downloaded by sharer *i* during period t
$Commands_{it}$	The total number of commands used by sharer *i* during period t
$Contribute_{it}$	The total number of files downloaded from sharer *i* during period t
$Active\_Time_{it}$	The total time sharer *i* had been present in the channel during period t
$Valued\_Member_{it}$	A dummy variable indicating whether sharer *i* has been marked as a valued member during period t
$Time\_Period_{itl}$	Time dummy variables: $Time\_Period_{itt}$=1 and $Time\_Period_{itl}$=0 if $l \neq t$ $t = 1,...,134$

Table 2. Descriptive Statistics of the Key Variables					
Variable	N	Mean	SD	Min	Max
Aggregate:					
User_Size	135	32,160.41	7,377.87	22,410.09	74,363.39
Sharer_Size	135	1,514.63	558.05	479.00	2,556.00
Freerider_Size	135	29,795.24	6,465.58	20,484.68	67,652.37
User_Download	135	938,594.44	152,112.05	580,763.09	1,302,609.58
Sharer_Download	135	221,876.30	82,159.29	79,707.11	378,545.63
Freerider_Download	135	716,718.13	92,150.05	452,337.39	958,370.72
User_Gain	134	15,246.48	4,139.47	1,008.80	34,128.33
User_Loss	134	14,951.93	5,018.88	165.02	40,820.36
User_Gain_Cum	134	11,060.07	3,824.73	6,281.75	36,180.00
Sharer_Gain_Cum	134	451.42	179.66	54.00	903.00
Sharer_Gain	134	640.76	277.19	176.00	1,403.00
Sharer_Loss	134	633.34	270.47	135.00	1,464.00
Sharer_Gain_Permanent	134	533.25	205.39	72.00	1,000.00
Sharer_Loss_Permanent	134	528.13	205.91	95.00	1,098.00
Sharer_Sharer	134	741.51	296.84	282.00	1,238.00
Sharer_Freerider	134	133.82	60.10	12.00	272.00
Individual:					
Download	8516	106.63	330.62	0	9587
Commends	8516	117.34	314.99	0	4173
Contribute	8516	315.82	908.74	0	15735
Active_Time	8516	260.59	498.31	0	5777
Value_Member	8516	0.87	0.33	0	1

Table 3. Correlation Matrix of the Key Aggregate Variables						
Variable	1	2	3	4	5	6
1. User_Size	1.00					
2. Sharer_Size	-0.41	1.00				
3. Freerider_Size	0.99	-0.48	1.00			
4. User_Download	-0.16	0.58	-0.19	1.00		
5. Sharer_Download	-0.19	0.81	-0.27	0.86	1.00	
6. Freerider_Download	-0.10	0.23	-0.07	0.89	0.52	1.00

4.2 Aggregate Models and Results

To analyze the data at the aggregate level, we propose four different measurements to describe community dynamics: cumulative user gain/loss, user gain/loss, sharer gain/loss, and sharers' decision. To examine how the dynamics is affected by community resources and activities, we construct four groups of systems of equations. Considering the variables used in our data are dynamically correlated over different time periods, we employ the Seemingly Unrelated Regression (SUR) technique and collectively estimate the system of equations. SUR assumes correlation across error terms in different equations which can provide links among different equations during estimation (Wooldridge 2002). SUR interrelates different equations and impose much fewer constraints than the single equation Ordinary Least Square (OLS) estimation, which fits well into our data of longitudinal observations.

In the first system of equations, as suggested by Butler (2001), we investigate the relationship between total user change variables (gain and loss) and user size and user downloads to see if the results for a message-based community carry over to its non-verbal counterpart (Models (1a) and (1b)):

$$User\_Gain_t = a_{1a} + b_{1a} User\_Size_t + c_{1a} User\_Download_t + \varepsilon_{1a} \qquad (1a)$$

$$User\_Loss_t = a_{1b} + b_{1b} User\_Size_t + c_{1b} User\_Download_t + \varepsilon_{1b} \qquad (1b)$$

We then extend our empirical analyses to the more detailed level. In Models (1c) and (1d), in order to see the possibly different impacts of sharers and free riders on the community, we replace the independent variables of aggregate user number and usage measurements with those of sharers' and free riders'. Since sharer download is highly correlated with the number of sharers (see Table 3), we remove it from the equation. We thus can estimation Models (1a)-(1d) simultaneously through the SUR estimation technique.

$$User\_Gain_t = a_{1c} + d_{1c} Sharer\_Size_t + e_{1c} Freerider\_Size_t + f_{1c} Freerider\_Download_t + \varepsilon_{1c} \quad (1c)$$

$$User\_Loss_t = a_{1d} + d_{1d} Sharer\_Size_t + e_{1d} Freerider\_Size_t + f_{1d} Freerider\_Download_t + \varepsilon_{1d} \quad (1d)$$

Table 4. Regression Results for Channel Size

	Dependent Variable			
Independent Variable	**User_Gain (1a)**	**User_Loss (1b)**	**User_Gain (1c)**	**User_Loss (1d)**
Constant	6794.69 (1844.71)	8905.87 (1462.74)	4421.65.64 (2113.00)	7441.25 (1658.31)
User_Size	0.16*** (0.04)	0.16*** (0.04)		
User_Download	0.004** (0.002)	0.0009 (0.001)		
Sharer_Size			0.83*** (0.23)	0.43*** (0.16)
Freerider_Size			0.21*** (0.04)	0.20*** (0.05)
Freerider_Download			0.005** (0.002)	0.001 (0.001)
Adj. R Sq	0.16	0.10	0.21	0.13

Note: ***, **, and * denote significance at 1%, 5%, and 10% respectively. Standard errors are in the parenthesis.

Estimation results for Models (1a)-(1d) are presented in Table 4. Results of Models (1a) and (1b) are in the most part consistent with Butler (2001). User size has positive impact on both overall user gain and loss; user download has positive impact on user gain (Hypotheses 1a, 1b, and 1c). These results suggest that both total activities and size are double-edged swords – they strengthen the community's ability in attracting new users but at the same time weaken its capacity in retaining users. In addition, in Models (1c) and (1d), we observe that both sharers and free riders have the same positive contribution to total user gain. However, Hypothesis (1d) is not supported because the coefficient of user download in Model (1b) is positive but not significant. Therefore, at the aggregate level, download activities, as the source of network congestion, do not play a role in user loss.

The second measurement of the growth of a community is the number of cumulative membership. This measurement is especially popular in the industry--when referring to size, companies running community web sites often cite the number of their registered users, a cumulative measure that never decreases. For example, a Flickr.com spokesman once cited the picture sharing site had 775,000 registered users before it was acquired by Yahoo in March 2005 (Kuchinskas 2005). We run a set of tests similar to Models (1a) and (1c) using cumulative gain as the dependent variable. Moreover, we also test the impact of size and activity on the numbers of cumulative sharer gain (Models (2c) and (2d)). The estimations results are demonstrated in Table 5.

$$User\_Gain\_Cum_t = a_{2a} + b_{2a}User\_Size_t + c_{2a}User\_Download_t + \varepsilon_{2a} \quad (2a)$$

$$User\_Gain\_Cum_t = a_{2b} + d_{2b}Sharer\_Size_t + e_{2b}Freerider\_Size_t + f_{2b}Freerider\_Download_t + \varepsilon_{2b} \quad (2b)$$

$$Sharer\_Gain\_Cum_t = a_{2c} + b_{2c}User\_Size_t + c_{2c}User\_Download_t + \varepsilon_{2c} \quad (2c)$$

$$Sharer\_Gain\_Cum_t = a_{2d} + d_{2d}Sharer\_Size_t + e_{2d}Freerider\_Size_t + f_{2d}Freerider\_Download_t + \varepsilon_{2d} \quad (2d)$$

Table 5. Regression Results for Channel Size

	Dependent Variable			
Independent Variable	**User_Gain_Cum (2a)**	**User_Gain_Cum (2b)**	**Sharer_Gain_Cum (2c)**	**Sharer_Gain_Cum (2d)**
Constant	6687.02 (1201.36)	7329.77 (1210.24)	297.30 (71.94)	169.25 (65.67)
User_Size	0.31*** (0.03)		-0.01*** (0.001)	
User_Download	-0.006*** (0.001)		0.0005*** (0.00006)	
Sharer_Size		-1.30*** (0.14)		0.13*** (0.01)
Freerider_Size		0.32*** (0.03)		-0.009*** (0.001)
Freerider_Download		-0.006*** (0.001)		0.0005*** (0.00007)
Adj. R Sq	0.45	0.53	0.48	0.60

Note: ***, **, and * denote significance at 1%, 5%, and 10% respectively. Standard errors are in the parenthesis.

Comparing Model (2a) to Model (1a), although the coefficient of user size remains positive, user download now has negative impact on cumulative user gain. Therefore user size and activities provide opposite forces for cumulative user gain: more users stimulate cumulative user gain, whereas more activities have a weakening effect. In other words, total user size is a source of positive network externalities, but their download activities are a source of negative network externalities for the community. This result complements Asvanund et al.'s (2004) finding that both positive and negative externalities exist in P2P music sharing networks and explicitly identify the sources of such externalities.

Due to the nature of the sharing networks, open and anonymous access encourages more free riding behavior compared to the mailing lists Butler studied, and therefore, the community growth in size is mainly due to the free riders, who do not contribute content. This creates network congestion, which in turn deters user gain. Furthermore, because sharer size and their download activities are highly correlated, the negative coefficient of Sharer_Size in Model (2b) suggests that sharers' activities actually dampen community's cumulative growth. In other words, more sharer downloads lead to slower user gain.

Another interesting finding is the opposite effects of sharer size and free rider size in Models (2b) and (2d), which suggest that we must be careful in assessing the impact of user size on the community. By separating sharers and free riders, we can pinpoint how each user type contributes to user gain and loss. Not surprisingly, the coefficient for sharer size is positive and significant in predicting cumulative sharer gain. On the other hand, more sharers and associated activities would produce a certain level of exclusiveness, which drives free riders away.

It is clear that sharers contribute to the positive network externalities of the community, while free riders are responsible for the negative network externalities. Free riders hinder the growth of the community by slowing down sharer gain. Sharers bring resources to attract more sharers and diminish the growth of free riders. These results extend Asvanund et al.'s (2004) work by identifying the source of the positive and negative network externalities.

Because sharers are the only contributors in the community as well as the more active participants, sharer gain and loss provide a more precise measurement of the community. We estimate Models (3a) and (3b) to explore how short-term sharer gain and loss can be influenced by the two types of participants and their activities in the community. In Models (3c) and (3d), we measure sharer gain and loss in a much longer period (12 weeks) (see description of Sharer_Gain_Permanent and Sharer_Loss_Permanent in Table 1). Estimation results for models (3a)-(3d) are presented in Table 6

$$Sharer\_Gain_t = a_{3a} + d_{3a}Sharer\_Size_t + e_{3a}Freerider\_Size_t + f_{3a}Freerider\_Download_t + \varepsilon_{3a} \quad (3a)$$

$$Sharer\_Loss_t = a_{3b} + d_{3b}Sharer\_Size_t + e_{3b}Freerider\_Size_t + f_{3b}Freerider\_Download_t + \varepsilon_{3b} \quad (3b)$$

$$Sharer\_Gain\_Permanent_t = a_{3c} + d_{3c}Sharer\_Size_t + e_{3c}Freerider\_Size_t + f_{3c}Freerider\_Download_t + \varepsilon_{3c} \quad (3c)$$

$$Sharer\_Loss\_Permanent_t = a_{3d} + d_{3d}Sharer\_Size_t + e_{3d}Freerider\_Size_t + f_{3d}Freerider\_Download_t + \varepsilon_{3d} \quad (3d)$$

Table 6. Regression Results for Sharer Size

	Dependent Variable			
Independent Variable	**Sharer_Gain (3a)**	**Sharer_Loss (3b)**	**Sharer_Gain_Permanent (3c)**	**Sharer_Loss_Permanent (3d)**
Constant	166.98 (109.05)	61.05 (105.87)	40.09 (85.37)	78.17 (75.38)
Sharer_Size	0.42*** (0.02)	0.44*** (0.02)	0.27*** (0.02)	0.30*** (0.02)
Freerider_Size	-0.003* (0.002)	-0.002 (0.002)	-0.006*** (0.001)	-0.005*** (0.001)
Freerider_Download	-0.00009 (0.0001)	-0.00006 (0.0001)	0.0003*** (0.0001)	0.0002**(0.0001)
Adj. R Sq	0.80	0.82	0.79	0.84

Note: ***, **, and * denote significance at 1%, 5%, and 10% respectively. Standard errors are in the parenthesis.

Equations (3a)-(3d) examine how the same set of independent variables as we used before affect the dynamics of sharers. The results suggest several findings. First, sharer size has a dominant effect on both sharer gain and loss. Second, free rider size and download activities actually have different effects on sharer dynamics. More free riders in number slow down both sharer gain and loss, which support Hypotheses (3a) and (3b), while more free riders download activities lead to higher sharer turnover.

Considering that free riders dominate the community both in download traffic (three times of that of sharers) and size (twenty times of that of sharers), we can claim that the community size and activities have different effects on its sharer growth. This extends Butler's (2001) results by separating and pinpointing the different effects from community size and activities.

Compared to Models (3a)-(3b), free rider size and download have significant affects on sharer gain and loss defined in a longer time period (Models (3c) and (3d)). As we proposed in Hypothesis (2c), long-term changes of sharers are affected by free riders' behavior but temporary changes are more random.

The fourth way to study the growth of the community is to check sharers' choice of actions in the following period: whether they keep sharing, change to a free rider, or leave the community altogether. Existing literature has shown the positive network externalities of member size (Asvanund et al. 2004), i.e. the greater the member size, the more benefits the community provides to its members, and the more likely members will contribute to it as well. It is, therefore, interesting to assess the relationship between sharer conversion and user size and download volume (Models (4a) and (4b)). The estimation results for Models (4a) and (4b) are shown in Table 7.

$$Sharer\_Sharer_t = a_{4a} + d_{4a}Sharer\_Size_t + e_{4a}Freerider\_Size_t + f_{4a}Freerider\_Download_t + \varepsilon_{4a} \quad (4a)$$

$$Sharer\_Freerider_t = a_{4b} + d_{4b}Sharer\_Size_t + e_{4b}Freerider\_Size_t + f_{4b}Freerider\_Download_t + \varepsilon_{4b} \quad (4b)$$

Table 7. Regression Results for Sharer Size		
	Dependent Variable	
Independent Variable	**Sharer_Sharer (4a)**	**Sharer_Freerider (4b)**
Constant	-21.18 (102.76)	-29.10 (36.95)
Sharer_Size	0.51*** (0.02)	0.05*** (0.008)
Freerider_Size	0.004** (0.002)	-0.002*** (0.0006)
Freerider_Download	-0.0002 (0.0001)	0.0002*** (0.00004)
Adj. R Sq	0.84	0.51
Note: ***, **, and * denote significance at 1%, 5%, and 10% respectively. Standard errors are in the parenthesis.		

The coefficients of Freerider_size in Models (4a) and (4b) suggest that by increasing the number of sharers who stay as sharers and reducing the number of sharers who become free riders, the number of free riders has a "moderating" effect on the turn-over of sharers. Interestingly, the force of free rider download is almost opposite: it discourages, although insignificantly, sharers to stay as sharers and encourages sharers to convert to free riders. In other words, free rider download hurts sharers' willingness to share (Hypothesis (2b)).

4.3 Individual Model and Results

Our individual level analysis complements the aggregate level results. In the previous sections, we propose four measurements of community dynamics and examine the related aggregate models. However, the aggregate level measurement of one variable usually is composed of multiple components, therefore making it difficult to differentiate the sources of impacts. For example, sharer size represents three different community factors: total resources, sharer download (their own benefit), and congestion generated by sharers. When a positive impact of sharer size on both sharer gain and sharer loss is observed, we cannot distinguish which one of the three factors is in play. Previous results presented in this paper demonstrated that sharers are the "core" part of the community. To generate deeper insights in understanding the factors that influence sharer's behavior in the community, we perform a micro level analysis investigating each individual sharer's behavior. Every sharer faces three choices at every stage of their residence in the community: to stay as a sharer, to become a free rider, or to leave the community. From our data, we randomly select 1000 sharers who were active in at least two time periods during our data collection. Still using biweekly time interval, at every time period, we analyze how a sharer's choice is influenced by activities of her own as well as the community overall. We run the standard Multinomial Logistic discrete choice model on our unbalanced individual level data (Wooldridge 2002). The dependent variable is the choice of each individual sharer at different time periods and the independent variables include both individual and aggregate level measurement of community activities. The results are shown in Table 8.

Table 8. Multinomial Logistic Regression Results for Sharer Decision

	Dependent Variables		
Independent Variable	**Sharer_Share Base choice: Sharer_Leave**	**Sharer_Freeride Base choice: Sharer_Leave**	**Sharer_Freeride Base choice: Sharer_Share**
Constant	0.29 (0.28)	-1.10 (0.51)	-1.39 (0.49)
Download	0.001*** (0.0002)	0.002*** (0.0002)	0.0005*** (0.0001)
Commands	0.0005*** (0.0001)	0.0004** (0.0002)	-0.00007 (0.0002)
Contribute	0.0005*** (0.00008)	0.0001 (0.0001)	-0.0004*** (0.0001)
Active_Time	0.001*** (0.0001)	-0.002*** (0.0003)	-0.003*** (0.0003)
Valued_Member	0.26*** (0.08)	-0.35*** (0.13)	-0.61** (0.13)
Sharer_Size	-0.0003*** (0.00006)	-0.0006*** (0.0001)	-0.0004*** (0.0001)
Freerider_Size	0.00001* (0.00001)	-0.00003*** (0.00001)	-0.00004*** (0.00001)
Freerider_Download	-2.0e-07 (3.0e-07)	2.3e-06*** (5.2e-07)	2.5e-06*** (5.0e-07)
	Number of observations = 8516, Log likelihood = -6694.68, Pseudo R Sq = 0.08		

Note: ***, **, and * denote significance at 1%, 5%, and 10% respectively. Standard errors are in the parenthesis.

Compared to Models (4a) and (4b), the individual level analysis provides the same evidence that free rider size is a “moderating” factor and free rider download represents part of community congestion. However, unlike the aggregate level results, the individual level analysis indicates that larger sharer size makes current sharers more likely to free ride or leave the community. Our interpretation for this finding is that, for the individual level model, the user-perceived resources and their benefit has been captured by the individual download (the variable Download). Therefore, Sharer_Size in this context captures solely the influence of network congestion generated by sharers. It is also noticed that the coefficients of sharer size are an order of magnitude bigger than those of free riders. This result is in line with our observation that on average a sharer downloads much more than a free rider, thus inducing much more congestion than a free rider does.

This discrete choice model also gives us two major results showing how individual characteristics influence sharer dynamics. First, Contribute, Active_Time, and Valued_Member all have a positive and significant impact on sharer’s decision to share. The more a sharer is involved in a community, e.g. contributing more files to others, staying longer, and being recognized as a valued member, the more he is willing to keep sharing. This confirms Hypothesis (2a). Second, Download has a positive and significant impact on the odd of sharing over leaving and free riding over sharing. In other words, when a sharer can download more files, he is more likely to stay in the community (either a sharer or a free rider). However, between the choices of sharing and free riding, more files downloaded make it more likely for the sharer to free ride. The intuition of this result could be that if downloading is too easy, users will treat a sharing network as a tool instead of a community.

5. Discussion and Future Research

The number of non-verbal online sharing communities has quietly taken off recently, thanks mainly to the ever lower technical barrier to create such a community. However, despite a growing body of literature on message-based online communities, there is little research on the emerging non-verbal sharing communities. How do different member types influence one another and the community without verbal communication is a fascinating question that Information Systems researchers need to study in order to understand the ever-popular form of online communities.

This research bridges our understanding in message-based communities and peer-to-peer networks and extends it to a new context. We run the same tests from Butler's (2001) work on the sustainability of message-based communities using a resource-based model. We find that, at the aggregate level, Butler's results do carry over and apply in sharing communities without verbal communications. Our paper also extends Asvanund et al.'s (2004) work on the network externalities of P2P networks by identifying the source of positive and negative network externalities. At the aggregate level, total user size provides positive network externalities as the larger the user size, the more resources available to community users, whereas download activities contribute to negative network externalities as the more download activities, the more congestion, which is a cost to every user.

Another important feature of our models is the two levels of analysis, i.e. at the aggregate-level for all users of a certain type, as well as at the individual-level for randomly selected individual users. The results from both perspectives are consistent with each other, confirming the robustness of our results.

In non-verbal sharing communities, naturally there are two types of users, sharers and free riders. We discover these two different types of users have different roles in the community. We argue that sharers are the core part of the community for several reasons. First, the value of a community is in its resource, which is solely provided by sharers. The disproportional network congestion caused by sharers' own intensive use shows that sharers also contribute to negative network externalities of a community. Moreover, the more sharers are involved in a community (through sharing and downloading activities), the more likely the sharers continue to contribute content.

What is perhaps most surprising is the role of free riders. We are able to decompose free riders' impact into that of the *number* of free riders and that of their *download activities*. And their impacts are at the exact opposites in retaining sharers. The larger the number of free riders, the more likely a sharer will continue to share, whereas the more free rider download activities, the more likely a sharer will stop sharing. While seemingly surprising, this result actually provides an illuminating breakdown of a typical voluntary sharer's utility function. On the one hand, more free riders in *number* imply a larger user base for the sharer's contribution, a greater exposure any voluntary sharer would be happy to see. On the other hand, more free rider *traffic* means increasing congestion in downloading for the sharers, who are themselves heavy users of the content. This result is confirmed at both the aggregate and individual user level analysis and therefore is robust.

What is also interesting is the role of sharer size in retaining sharers. Similar to previous studies, we recognize sharer size is responsible for both sharer gain and loss. However, we note that the dual effect may be due to the fact that sharer size represents three conflicting community factors: total resources available, sharer download (their own benefit), and congestion generated by sharers. When we disentangle the effect of sharer download from sharer size through individual-level analysis, we find sharer size only has a negative impact on sharing decision—the positive effect of sharer size in the aggregate model is explained by the sharer download. This also reveals an interesting fact – sharers are much heavier users than free riders and they may produce more congestion than people expected.

Our findings have several significant managerial implications. First and foremost, the role of free riders is not as clear-cut as people believe. Their impacts on the community, especially on sharers, are both positive and negative. For entrepreneurs, when designing non-verbal communities, the creator should encourage all types of users to join, even if some of them may not contribute any content—their mere existence will be a reason for potential contributors to join. At the same time, the community operator should carefully monitor free riders' download activities, maybe to go as far as imposing an upper limit to free riders' download volume or bandwidth, so as to ensure that the sharers' downloads are smooth and not affected by the often excessive free riders' traffic. To retain sharers, it is also helpful for the community to provide more features for sharers to interact with other users. For example, recognition of some sort always has a positive impact on sharer retention, even if it is as simple as adding a special mark in front of a user ID. However, one thing the community does not want to make too easy for a sharer to do is downloading. Based on our analysis, everything else being intact, sharers have a stronger tendency to become a free rider if it is easier to download content. They would just treat it as another application to download content and ignore the community aspect. For people who need to evaluate non-verbal sharing communities such as investors, they should not only look at the total number of registered users in a community, but also different user types and their often conflicting forces, even within one's own type. While sharers should be valued highly for community sustainability and growth, free riders should also be credited for their part of the contribution, if not in the form of actual content.

For researchers, our results extend academic understanding of message-based communities to non-verbal communities. From the resource-based view perspective, even though user size is a resource, its components, sharers and free riders should be considered two different resources. Sharers provide content to the community, which directly benefits all community members; Free riders, despite their uncompensated consumption of bandwidth, provide resources through their positive impact on sharer retention. Our results also shed light on the exact sources of positive and negative network externalities, expanding Asvanund et al.'s work on P2P music sharing networks.

For future research, one important issue to study is "when and why does a free rider start to share?" To answer this, we need to find the composition of sharers—did they share starting from the first day, or did they convert from free riders. And for the latter type, what factors affect their decision to change?

Complementary to the question, we can analyze what factors influence free riders' choice of a user type for the next period. We anticipate that community's features have little impact on free riders' choice of not sharing. For the small fraction of free riders who do choose to convert to sharers, it is possible that the community's aggregate measure have a positive impact.

References

Adar E., B. A. Huberman. 2000. Free riding on Gnutella. In *First Monday*, http://www.firstmonday.dk/issues/issue5_10/adar/index.html. Last retrieved on July 10, 2006.

Arguello, J, B. S. Butler, L. Joyce, R. Kraut, K. S. Ling, C. Rose, X. Wang. 2006. Talk to me: Foundations of successful individual-group interactions in online communities. Working paper. http://www.cs.cmu.edu/~jaime/Foundations_for_Success%20_Submission %20Copy_final.pdf. Last retrieved on April 6, 2006.

Asvanund, A., K. Clay, R. Krishnan, M. Smith. 2004. An empirical analysis of network externalities in peer-to-peer music sharing networks. *Information Systems Research* **15**(2) 155-174.

Bobala, B. 2001. Last breaths of theglobe.com? *The Motley Fool*, August 6, http://www.fool.com/news/2001/tglo010806.htm. Last retrieved on June 1, 2006.

Butler, B. S. 2001. Membership Size, communication activity, and sustainability: a resource-based model of online social structures. *Information Systems Research* **12**(4) 347-362.

Gu, B., S. Javenpaa. 2003. Online discussion boards for technical support: The effect of token recognition on customer contributions. Salvatore T. March, Anne Massey, Janice I. DeGross, eds. *Proc. 2003 Internat. Conf. Inform. Systems*, Seattle, WA (Dec. 14—17).

Gurbaxani, V. 1990. Diffusion in computing networks: the case of BITNET, *Communications of the ACM* **33** 65-75.

Hardin, G. 1968. The tragedy of the commons. The population problem has no technical solution; it requires a fundamental extension in morality. *Science*, 162(859):1243-8.

Hars, A., S. Ou. 2000. Why is Open Source software viable? - a study of intrinsic motivation, personal needs, and future returns. In *Proceedings of the Sixth Americas Conference on Information Systems* (AMCIS 2000), 486-490, Long Beach, California.

Jones, Q., G. Ravid, S. Rafaeli. 2004. Information overload and the message dynamics of online interaction spaces: a theoretical model and empirical exploration. *Information Systems Research* **15**(2) 194–210.

Kollock, P. 1999. The economies of online cooperation: Gifts and public goods in cyberspace, in Marc Smith and Peter Kollock eds. *Communities in Cyberspace*, Routledge, London.

Krishnan, R., M. D. Smith, R. Telang. 2003. The economics of peer-to-peer networks. *Journal of Information Technology Theory and Applications* **5**(3) 31-44.

Krishnan, R., M. D. Smith, Z. Tang and R. Telang, 2004. The virtual commons: why free-riding can be tolerated in file sharing networks". SSRN working paper, available at SSRN: http://ssrn.com/abstract=450241 or DOI: 10.2139/ssrn.450241

Kuchinskas, S. 2005. Flickr to add print to photo service. Internet News, June 15. Available at http://www.internetnews.com/ec-news/article.php/3512866. Last retrieved on July 10, 2006.

Lakhani, K., R.G. Wolf. 2003. Why hackers do what they do: understanding motivation and effort in Free/Open Source Software projects, working paper, MIT Sloan School of Management, September.

Liedtke, M. 2006. Now starring on the Internet: YouTube.com. http://www.usatoday.com/tech/news/techinnovations/2006-04-09-youtube-popularity_x.htm, Last retrieved on April 6, 2006.

Nonnecke, B., D. Andrews, J. Preece. 2006. Non-public and public online community participation: Needs, attitudes and behavior, *Electron Commerce Res.* **6** 7-20.

Peddibhotla, N., M. Subramani. 2005. Do the same motivations drive all contributors to public document repositories? An empirical study. University of Minnesota working paper.

Ranganathan, K., M. Ripeanu, A. Sarin, I. Foster. 2004. "To share or not to share": An analysis of incentives to contribute in collaborative file sharing environment. Working paper. University of Chicago.

Raymond, E. 2000. *The Cathedral and the Bazaar*: *Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly and Associates, Sebastopol, California.

Shah, S. 2006. Motivation, governance and the viability of hybrid forms in open source software, *Management Science* **52** 1000-1014.

Torvalds, L. 1998. "What motivates free developers?", Interview in First Monday 33, available at: http://firstmonday.org/issues/issue3_3/torvalds/index.html. Last retrieved on July 10, 2006.

Wasko, M. M., S. Faraj. 2000. Is it what one does: Why people participate and help others in electronic communities of practice. *Journal of Strategic Information Systems* 9(2–3) 155–173.

Wingfield, N., E. Smith. 2003. Crowded house: with the web shaking up music, a free-for-all in online songs; companies race to stake out turf in fledgling market; a shakeout on the way? *Wall Street Journal*, Nov. 19, pA1.

Wooldridge, J. M. 2002. Econometric Analysis of Cross Section and Panel Data, the MIT Press, Cambridge, MA.