5-1-2008

The Virtual Commons: Why Free-Riding Can Be Tolerated in File Sharing Networks

Ramayya Krishnan
Carnegie Mellon University, rk2x@andrew.cmu.edu

Michael D. Smith
Carnegie Mellon University, mds@andrew.cmu.edu

Zhulei Tang
Carnegie Mellon University

Rahul Telang
Carnegie Mellon University, rtelang@andrew.cmu.edu

Published In
.

The Virtual Commons:
Understanding Content Provision in Peer-to-Peer File Sharing Networks

Ramayya Krishnan[†], Michael D. Smith[†], Zhulei Tang[‡], Rahul Telang[†]

{rk2x, mds, rtelang, zhuleit}@andrew.cmu.edu

This Version: November 2004

Acknowledgements: We thank participants at the 2004 Hawaii International Conference on Systems Sciences, the 2003 Workshop on Information Systems Economics, the 2002 International Conference on Information Systems, the University of California at Berkeley SIMS and Haas Schools, the Management Science department at the University of Washington, and the Carnegie Mellon University Graduate School for Industrial Administration for valuable comments on this research. Financial support was provided by the National Science Foundation through grant IIS-0118767.

† H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, Pittsburgh, PA, 15213. (Smith: 412-268-5978)

‡ Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, 15213.

The Virtual Commons:

Understanding Content Provision in Peer-to-Peer File Sharing Networks

ABSTRACT

Peer-to-peer networks have emerged as a popular alternative to traditional client-server architectures for the distribution of information goods. Recent academic studies have observed high levels of free-riding in various peer-to-peer networks, leading some to predict the imminent collapse of communities hosted on these networks.

Our research develops static and dynamic analytic models to analyze the behavior of peer-to-peer networks in the presence of free-riding. In contrast to previous predictions we find that free-riding is sustainable in equilibrium and in some cases occurs as part of the socially optimal outcome. We also show that network performance can be improved by grouping peers into communities that provide similar value and by providing external incentives to peers to share. With regard to the latter finding, we show that without external incentives, the level of free-riding in peer-to-peer networks will be higher than the socially optimal level, but that quality of service tied to the contribution of content can be used as a lever to induce users to share and thereby achieve the socially optimal network outcome. Our results are robust to consideration of a dynamic game where content is duplicated in the network over time.

1. Introduction

Peer-to-peer (P2P) networks have gained a great deal of exposure recently as a way for a distributed group of users to easily share resources. Napster was the first such network to gain prominence for sharing consumer information goods, but numerous other similar networks including Gnutella, OpenNap, and Kazaa have followed it. However, P2P architectures also have the potential to enable widespread adoption of sharing in other application domains and of a wide variety of resources. For example, P2P networks have recently been developed for sharing streaming media content (e.g., Allcast, Blue Falcon, Kontiki, Uprizer), computing power (e.g., SETI@Home), and spam filtering (e.g., Couldmark); and are being deployed in business settings for remote collaboration (e.g., Groove Networks), enterprise information sharing (e.g., Bad Blue, Nextpage), and decentralized data storage and archival (e.g., Publius, FreeHaven).

One widely discussed characteristic of peer-to-peer networks is that the economic characteristics of sharing in these network can lead to an under provision of content. P2P networks rely on the willingness of individual users to provide resources (e.g., content, processing power, storage) to other users. However, users can, and frequently do, turn off their resource sharing. Consuming network resources through downloading without providing resources to the network in return is commonly known as free-riding, and is widespread in P2P networks. For example, Adar and Huberman (2000) found in August 2000 that 66% of Gnutella 0.4 users do not contribute content to the system and the top 1% of users contribute over 50% of the total amount of content. Similarly, Asvanund et al. (2004) found that 42% of Gnutella 0.6 users were free-riders in September 2002. In noting the inefficiency of these high levels of free-riding, Adar and Huberman (2000) observe “Free-riding leads to degradation of the system performance…if this trend continues copyright issues might become moot compared to the possible collapse of such systems.”

Since these networks are gaining popularity across various application domains, our ability to understand user participation, content evolution and free-riding is critical. It is important for a variety of managerial audiences including network designers designing protocols with incentives to optimize user participation , entrepreneurs designing enterprise-level P2P content management systems, businesses deploying these systems, and copyright holders seeking to reduce piracy on P2P networks.

Specifically, we seek to understand how and when users will contribute to these networks? Can P2P networks persist in spite of high levels of free-riding? How do heterogeneity in user content and cost change their decision to participate? How does duplication of content, which is a unique feature of these networks, change the value of these networks? How does this affect the decision by the peers to participate? And finally, how can we create practical economic incentives to improve network performance?

We address these questions using static and dynamic models of user participation in P2P networks in the presence of free-riding. While the issue of free riding in the context of the provision of public goods has been well studied by both economists and sociologists (e.g., Butler 2001, Faraj and Wasko 2002, Groves and Ledyard 1977, Kollock 1999, Kraut et al 1998, Sandler 1992, Sharp 1997) P2P networks have several unique characteristics that differ from a typical public goods context. For example, the endowment of public goods are typically discrete and fixed with regard to contribution levels: Either there is a sufficient level of participation and the good is provided or the good is not provided. In contrast, the size of the endowment of resources in a P2P network varies based on the level of contribution. Each peer who shares resources increases the size of the endowment to all other members. Public and club goods are also typically provided through a monetary outlay that is independent from the consumption of the resource. In contrast, in P2P networks, provision occurs entirely through resource allocation on the part of network participants. Finally, in P2P networks consumers of the P2P resources are also the providers of the content in subsequent periods through sharing, making content replication both a unique characteristic of provision in P2P networks and a factor which can potentially alter the quality and quantity of available resources and the dynamics of user participation over time.

We incorporate these unique characteristics into analytic models of resource provision and network performance in P2P networks. We find that in contrast to Adar and Huberman's prediction that free-riding will lead to the collapse of P2P networks, both sharing and free-riding can be observed in equilibrium in P2P networks. We also find that sharing in these networks occurs even in the absence of altruism or external incentives. Sharing is individually rational in our model because sharing content serves to offload traffic at other servers the sharing peer is interested in accessing. We also find that when the cost of sharing is the same across peers, high value content is shared before lower value content. These results are reinforced in a dynamic game because users who download high value content will share it in future periods — ensuring that high value content will propagate on the network even if the originating peer exits the network. Finally, while our model shows that free-riding is sustainable in equilibrium, without external incentives the level of free-riding is higher than the optimal level for the network — degrading performance for all network users. To address this issue, we propose an easily implemented mechanism where, by differentiating the quality of service provided to sharers and non-sharers, network operators can influence the level of free-riding and thereby achieve the socially optimal network outcome.

The remainder of this paper proceeds as follows. In Section 2 we review the relevant economic and social networks literatures related to our question and review the emerging literature on the

P2P networks. In Section 3 we present our model and derive analytic results. In Section 4 we derive results for network performance under several incentive schemes designed to reduce free-riding. We extend our model in Section 5 to analyze the impact of heterogeneity in sharing costs and user value on sharing. In Section 6 we analyze how content duplication in P2P networks leads changes user sharing behavior. Finally, we provide concluding remarks and discuss some avenues for future research in Section 7.

2. Literature Review

Our analysis combines the economics literature relating to public and club goods with the computer science literature relating to improving efficiency in P2P networks. Following the definition of Mas-Colell et al., a public good is "a commodity for which use of a unit of the good by one agent does not preclude its use by other agents" (1995, p. 359). One well-known characteristic of public goods is that the self-interested consumption may deplete the overall public utility. This is commonly known as the "tragedy of the commons" (Hardin 1968). The tragedy of the commons problem occurs because public goods are under-provided and over-consumed due to inadequate incentives on individual behavior.

Typically, in a decentralized setting, we are more concerned about the private provision of public goods. Not surprisingly, various papers in the literature have analyzed free-riding in the context of public goods. One interesting stream of this research shows that free-riding worseness as group size increases (Olson 1968; Palfrey and Rosenthal 1984, 1988; Hindricks and Pancs 2001; Dixit and Skeath 1999). Other relevant papers in this literature examine the private provision of public goods in a static complete information setting (e.g., Bergstrom, et al. 1986; Gradstein and Nitzan 1990). Later work extends this stream of research by introducing incomplete information (Nitzan and Romano 1990) and time dynamics (Gradstein (1992; Konrad 1994).

Research by Asvanund et al (2004) and Krishnan et al. (2003) has found that resources provided over P2P networks share many of the characteristics associated with the private provision of public goods. In particular, in many P2P networks, peers do not need to provide resources as a precondition for consuming resources. Because providing resources imposes a cost on peers (Feldman et al. 2003) many peers choose to "free-ride" by consuming network resources without providing resources back to the network in return. In fact, much work on P2P networks is based on the premise that without "altruism" all users will be free-riders (Ranganathan 2003). However, this view ignores the possibility that sharing could reduce congestion on these networks, which may increase an individual peer's utility of using the network — providing a rationale for sharing even in the absence of "altruism."

Given the prevalence of free-riding, much of the prior work in P2P networks has explored mechanisms to reduce free-riding. These mechanisms typically use either monetary transfers or membership rules based on persistent identity. For example, Chandan and Hogendorn (2001) discuss monetary solutions to the free-riding problem in a wireless P2P application. Similarly, Golle et al. (2001) propose to use P2P micropayments to encourage efficient use of scarce network resources. In many P2P settings, however, solutions based on payments may be impractical. In consumer file-sharing networks for example, peers are globally distributed and in many cases anonymous for all practical purposes, making monetary incentive schemes extremely difficult to implement. Likewise, monetary transfers across organizational boundaries may be impractical for P2P knowledge management solutions deployed within enterprises.

Other researchers have proposed membership rules based on persistent user identity to encourage peers to share. For example, Vishnumurthy et al. (2003) propose a system for tracking both a user's contribution to and consumption of P2P network resources and then using this information to govern future consumption of network resources. Kamvar et al. (2003) propose a similar scheme that allows peers to purchase the right to issue queries to other peers in the network based on their past contribution of network resources. However, in many P2P file sharing settings identity can be easily changed. Thus approaches based on static identity are difficult to implement in our setting.

In contrast to the approaches to limit free-riding mentioned above, we propose to provide peers with incentives to share their content by differentiating the quality of service provided to peers based on whether they share content. Since, the number of files shared by a peer is readily observable in most current P2P networks, such an approach has practical feasibility and does not require changes to the existing protocol.

3. A Model of Contribution of Content in Peer-to-Peer Network

3.1. Static Contribution with Heterogeneous Content

In this baseline model we consider a network in which each user[1] has a single endowment of unique content, and where each user independently decides whether to share content based on her private utility. Sharing implies a cost to the user by reducing the user's private bandwidth available for downloading. However, sharing deflects traffic from other peers in the network to the sharing peer, thereby increasing the chance that the user will be able to get her desired content from other peers on the network. Thus, it is possible for a peer to increase her private utility through sharing. In effect, the sharing decision serves as a means to redistribute traffic in our model. It provides a rational incentive for some users to share content even without altruism or external incentives.

[1] We use the terms user, peer and node interchangeably in this paper.

It is important to note that this setup extends the standard modeling of the provision of public goods by explicitly modeling the impact of peers' sharing decisions on the size of the endowment. In a typical public or club goods environment, the size of the endowment is discrete and fixed exogenously, while in a P2P environment the size of the endowment is a function of the number of users. (In Section 5, we also model replication, another unique aspect of peer-to-peer networks vis-à-vis public goods.)

To formalize this setup, consider a static game where, in period 1, n users join the network simultaneously, each with a unique endowment of content (see Figure 1). Further let the value of this content to other peers on the network be heterogeneous. Specifically, we assume that the value of the content V is a random variable, which has a distribution on the integers $v_1, v_2, ..., v_n$. Without loss of generality, we assume that we can rank order the value of this content such that $(v_1 \geq v_2 \geq v_3 \geq .... \geq v_{n-1} \geq v_n)$. [2]

In period 2, each user independently decides whether to share their content, and each user demands one unit of content randomly from the other n-1 users. If a user does not share, she can still download content from other users. We refer to downloading without sharing as *free-riding*. Sharing content incurs a lump-sum cost c, which is common knowledge. Initially, we assume that this cost is fixed and independent of number of other users downloading from the peer and same for all users; however, we relax this assumption in subsequent model development.[3]

In period 3, each user realizes some utility level based on their sharing strategy, the congestion in the network, and the variety of content available on the network.

[2] We assume that all users in the network have the same preferences on the value of the contents and have the same rank ordering.
[3] For example, in case of a DSL connection, there is a fixed upload bandwidth. It would not matter to the user how many users download from him for his download rate will remain unchanged. Users may also incur fixed costs of content independent of number of downloaders.

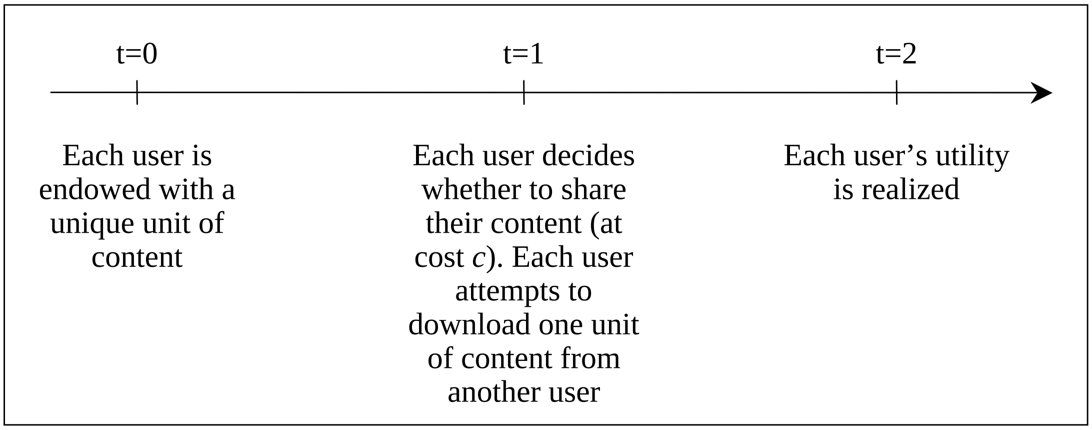

Figure 1: Timing of the Game.

Users experience congestion when they download content from a peer. The congestion arises due to other users trying to download content from the same peer. Clearly, the higher the value of the content carried by that peer, the more the congestion at that peer *ceteris paribus*. Since the up-load bandwidth on a peer is finite, when more users are trying to download from a peer, each will experience more congestion and delay.

Based on the discussion in the preceding paragraph, if U_{ij} is the utility peer *i* gets when it down-loads from a peer *j* then the utility function satisfies the following assumption: $\frac{\partial U_{ij}}{\partial v} > 0, \frac{\partial^2 U_{ij}}{\partial v^2} < 0$ and $\frac{\partial U_{ij}}{\partial n} < 0, \frac{\partial^2 U_{ij}}{\partial v^2} > 0$.[4]

This means that the utility of peer *i* is increasing in the value of peer *j*'s content v_j, but at a de-creasing rate; and that utility is decreasing, in the number of users trying to download from peer *j*, n_j, but at an increasing rate. Each of these assumptions is consistent with the empirical findings in Asvanund et al (2004).

[4] We can also use difference function to represent these properties as *n* is an integer. Therefore, $\Delta U_{ij}^{v}(.) > 0,\ \Delta^2 U_{ij}^{v}(.) < 0;$ and $\Delta U_{ij}^{n}(.) < 0,\ \Delta^2 U_{ij}^{n}(.) > 0$. However, for notational convenience we will continue to use derivatives.

In particular, all else equal one might expect that more users would download from a high value peer. But as noted above, when more users are trying to download from a peer, they will experience more congestion. Therefore, even though high value content is available, higher congestion may reduce each user's utility such that users may end up downloading from a lower value peer where they will experience less congestion. This means a user will always face trade-offs between the value of the content they are downloading and the congestion of the peer from which he is trying to download content.

Therefore, we can write a user's utility as her ability to get the desired content from a peer without suffering undue congestion. If *k* users are sharing their content in the network, then the gross utility of peer *i* downloading content from peer *j* is

$$U_{ij} = f\left(v_j, n_j(v_j, k; N)\right)$$

where, n_j is the number of other users downloading the content from peer *j*.

In the following lemma we show that each sharing peer *j* = 1, 2,…, *k* should provide the same value to the whole network.

Lemma 1: *In equilibrium, all peers that share content provide the same utility to the other peers in the network*

Proof: See the Appendix.

The intuition for Lemma 1 is the following: A higher value peer attracts more downloaders, which increases congestion and reduces its attractiveness; a lower value peer attracts fewer downloaders, which decreases congestion and increases its attractiveness. As a result, in equilibrium, downloading peers will adjust their download "targets" until all peers that share their content provide the same utility.

Using Lemma 1, if peers $j = 1, 2, ...k$ are sharing content, we can write that

$$U_1 = U_2 = ... = U_k = U_0^k$$

where U_0^k is the utility provided by the network to any downloader when k peers are sharing their content. We now focus our attention on U_0^k for the subsequent analysis.

First consider how U_0^k changes as the number of sharers on the network changes. Consider another peer k' who decides to share content of some value $v_{k'}$. As long as $f(v_{k'},1) > U_0^k$, at least one of the downloaders from peer k (or any other sharing node) has an incentive to switch to peer k'.[5] Since $\frac{\partial f}{\partial n} < 0$, the utility of peer k should increase, which in turn will lead to redistribution of traffic such that $U_0^{k'} \geq U_0^k$. Therefore, it must be that $\frac{dU_0^k}{dk} > 0$, where k is the number of sharers. Hence, the total utility of the network increases when more users share their content.

The above discussion provides the utility structure of our model. Now, we focus on a peer's decision of whether to share its content or not. Suppose a network of k sharers provides some utility U_0^k. Consider the sharing decision of two peers with value v_x and v_y (where, without loss of generality, $v_x > v_y$) who are not in the subset of k sharers. In equilibrium, if peer v_x shares then it will attract more downloaders than peer v_y does. Since the total number of users in the network N is unchanged, when v_x begins to share her content, she is able to shift more downloaders from other peers to herself, and thus the marginal utility that v_x brings to the network is higher than the marginal utility that v_y brings to the network. Said another way, $U_0^{v_x} - U_0^k > U_0^{v_y} - U_0^k$, where

[5] More precisely, some proportion of downloaders from each peer will shift to peer k'.

$U_0^{v_x}$ is the new utility of the network when peer v_x starts sharing, and similarly, $U_0^{v_y}$ is the new utility when peer v_y starts sharing.

In summary, the marginal benefit from sharing for a higher value peer is higher than the marginal benefit of a lower value peer. If the cost of sharing is fixed *ceteris paribus*, in equilibrium a higher value peer should share first.

Lemma 2: *Given a fixed cost of sharing, in equilibrium, if peers share, then higher value peers should share before the lower value peers.*

The proof follows from our discussion above. Because the sharing costs are the same, a higher value peer will receive a higher benefit from sharing than the low value peer does, which gives the higher value peer a larger incentive to share.

Because of this, if k users share in equilibrium, it must be that $v_1, v_2, v_3 .... v_k$ share while the other $v_{k+1}, v_{k+2}, ... v_N$ peers free-ride.[6] Thus, U_0^k becomes the utility obtained from the network when top k peers share.

Now we can turn our attention to the equilibrium number of sharers. Consider the decision of the k^{th} peer to share, when the top k-1 peers share. Given the fixed sharing cost c, the k^{th} peer will share its content of value v_k if

$$U_0^k - U_0^{k-1} > c \tag{1}$$

Similarly, consider the decision of the $k+1^{st}$ peer to share when k peers are sharing. Given the fixed sharing cost c, the $k+1^{th}$ peer will not share its content of value $v_{k+1} < v_k$, if

$$U_0^{k+1} - U_0^k < c \tag{2}$$

[6] In case of a tie in values, either both or one of them at random will share.

Combining (1) and (2), we get

$$U_0^{k+1} - U_0^k < c < U_0^k - U_0^{k-1} \tag{3}$$

Since v is decreasing in k, U_0 is a concave function in k, and we can characterize the sharing decision of the peers as follows.

Characterization of Equilibrium

Proposition 1*: There exists a unique pure strategy equilibrium, where every player contributes if* $c < U_0^N - U_0^{N-1}$*. There exists a unique pure strategy equilibrium where none of the players contributes if* $c > U_0^2 - U_0^1$*. For intermediate values of c, there exists a unique pure strategy equilibrium such that out of N original users, the top k users with the highest value of content* $v_1, v_2, v_3 .... v_k$ *will contribute, while the bottom N - k users free-ride.*

Proof: See the Appendix.

Proposition 1 formalizes the optimal user strategy for all potential ranges of cost c. The most interesting case is the intermediate range of cost c where, in spite of the existence of free-riding, there are always k^* users willing to share fully. By sharing, these users are able to reduce the congestion on other peers that share, making it easier for them to download their desired content from those peers. Therefore, sharing leads to redistribution of traffic. As long as the costs are not too high, some users will rationally find it optimal to share, while other peers will continue to free-ride. More interestingly, we show that high value users are the ones with the incentive to share first, which further increases the value provided by the network.

It is important to note that in an equilibrium where all users are symmetric (i.e. $v_1 = v_2 = v_3 = ... = v_{n-1} = v_n$), k^* users may still contribute while the others free-ride, because U_0 is concave in k. Because users are symmetric in value, *a priori*, one cannot say which k* users will share. That is, some permutation of *k* out of *n* users will share and rest will free-ride. Therefore, in this case we would observe multiple pure strategy equilibria. This is formalized below.

Corollary 1: *When all users are symmetric in the value of their content, there exists a unique pure strategy equilibrium, where all users contribute, if* $c < U_0^N - U_0^{N-1}$. *There exists a unique pure strategy equilibrium, where none of the users contribute if* $c > U_0^2 - U_0^1$. *For intermediate values of c, there exist multiple pure strategy equilibria, where out of N original users, any* k^* *users will contribute while the other N - k users free-ride.*

As mentioned previously, much of the work in the P2P literature points to the fact that without altruism there will be no sharers and the game degenerates into something akin to prisoner' dilemma (e.g., Ranganathan et al. 2003, Strahilevitz 2002). It is important to note that this result is a special case of our framework when costs are very high. Specifically, from proposition 1 it is immediately clear that for sufficiently high sharing cost *c*, no one shares and the network collapses. However, while this is possible, in practice the prevalence of P2P networks with both sharers and free-riders suggests that the sharing cost is likely to fall in the intermediate range of values between full sharing and network collapse. More importantly, our results show that high value content is shared before low value content.

The relationship between the size of the network and cost of contribution

We have shown that in a network in which each participant voluntarily chooses to contribute, a pure strategy equilibrium exists where out of *n* original users, the *k* $(0 \le k \le n)$ users with the highest value content contribute. In this Section we address how this equilibrium changes as the cost of contribution *c* changes; specifically showing that an increase in the cost of contribution causes a decrease in the number of contributors. The following lemma establishes this property of equilibrium behavior.

***Lemma 3*:** *As the cost of providing content increases, fewer people will contribute.*

Proof: See the Appendix.

4. Mechanisms to Limit Free-riding

4.1. Social optimality

In Section 3, we showed that peer-to-peer networks persist in equilibrium in spite of free riding. In this Section, we apply our static model from Section 3 to show that the equilibrium outcome in the absence of external user incentives is less efficient than the socially optimal outcome.

To solve for the social optimal level of sharing, first note that to maximize social utility one must maximize the sum of the individual utilities for the network participants. That is, the socially optimal number of contributors *k* solves the following problem (i.e., the social planner's problem):

$$\underset{k}{Max}\ nU_0^k - kc \qquad \text{(SP)}$$

Proposition 2. *In the absence of external incentives to encourage sharing, the optimal number of sharers in a network is below the socially optimal level for any n and c.*

Proof: See the Appendix.

Proposition 2 demonstrates that the standard result, that public goods are underprovided in equilibrium, holds in the P2P environment. However, it is significant to note that while much of the discussion around free-riding in P2P networks has focused on eliminating all free-riding, in our result the social optimum does not necessarily imply that $n=k$. Specifically, if $\frac{c}{n} > U_0^n - U_0^{n-1}$, some users will be free-riders even in the socially optimal solution. The intuition for this somewhat counter-intuitive result is that whenever the cost an individual peer would incur to share does not justify the value she would provide to the network, the socially optimal outcome is that this peer should not share her content.

4.2. Providing incentives through Quality of Service Differentiation

Since the equilibrium number of sharers and the socially optimal number of sharers differ, a natural question is: how we can motivate peers to contribute more content so that social optimality can be obtained?

In this section, we propose an incentive mechanism to allow us to achieve the socially optimal outcome. As noted above, a variety of solutions to the free-rider problem have been proposed based on monetary payments and/or membership rules (e.g., Chandan and Hogendorn (2001), Golle and Mironov (2001)). In contrast, our mechanism relies on varying the quality of service provided to users based on whether they share resources. In our previous model development, we treated contributors and non-contributors equally in the sense that contributors and non-contributors have the same probability to obtain their desired content. In this section, we drop this requirement and allow contributors to have a higher probability of obtaining their desired content. This approach is similar to the approach used by some current P2P networks (e.g., Kazaa and Bit Torrent).

To operationalize this incentive structure, we assume that a peer can distinguish between download requests from contributors and from non-contributors. This capability is available in most current P2P networks either directly[7] or indirectly by first querying a potential downloader's content before starting the download, and thus this capability does not rely on persistent identifi-

[7] E.g., in Gnutella 0.6 ultrapeers know the content shared by all the peers they are connected to, and thus can tell which of these peers are free-riding.

ers. Once a peer knows the potential downloader to be a sharer, contributors could always be ranked ahead of non-contributors with regard to ordering response packets.

Further, our mechanism ensures that non-contributors will have a probability p ($0 < p < 1$) to obtain their desired content before their request packets are dropped. Denoting the equilibrium number of sharers in the presence of incentives as k_p^*, we can modify (3) to get the revised equilibrium condition:

$$U_0^{k_p^*+1} - pU_0^{k_p^*} \le c \le U_0^{k_p^*} - pU_0^{k_p^*-1} \qquad (4)$$

Proposition 3: *For any cost* $c < U_0^n - pU_0^{n-1}$*, the number of sharers under this incentive structure will be strictly larger than the number of sharers in the absence of incentives.*

Proof: See the Appendix.

Thus, this quality of service mechanism will encourage more users to share as long as some users were free-riding in the absence of incentives (which, of course, is the only case where such incentives would be necessary).

Hence, the probability p is a tuning parameter. As p decreases, sharing increases as non-sharers consider the likelihood of not being able to access content because of dropped packets. At some level of p the socially optimal level of sharing can be reached. In effect, the probability that a packet is dropped functions as an additional implicit cost in the decision for non-sharers.

5. The Impact of Clustering on Sharing

Recent work (e.g., Asvanund et al. 2004; Li et al. 2004) has argued that creating communities of interest within peer-to-peer networks can improve information retrieval within the network. In this section, we extend this argument to show that clustering users into groups where user's provide similar value to each other can also increase their propensity to share versus a more heterogeneous network. To do this, we relax our previous assumption that the cost of sharing is a constant *c* for all users and is independent of number of downloaders. Instead, in this section we allow peers to face different costs because, it is likely that when more peers are trying to download from it, its cost may increase. For example, Feldman et al. (2003) have shown that because of SYN/ACK packets in the TCP/IP protocol, even with duplex connections, when more users upload content from a peer, its ability to download content decreases. In other words, the peer incurs a longer delay in downloading as more users download content from it because of its inability to acknowledge the packets it is downloading. It is also possible that a peer which generates a lot of traffic on its node may face higher legal costs.

In general, the cost *c* faced by a peer should be a monotonically increasing function of the number of peers downloading content from it. If *c* were to be a continuous function of *n*, then extending the analysis of Section 3.1 will require specific functional form assumptions on values of *v* and *c*(*n*), and a general proof of peer participation is difficult. In practice, empirical evidence suggests that *c* takes on a discrete form (Feldman et al 2003; Asvanund et al. 2004), where the cost of sharing is low below the capacity of its connection, and increases rapidly to infinity as that capacity is approached. We can use this finding to simply our assumption. Specifically, we assume the following step function cost structure:

$c = c_0$ if $n \le n^*$

$c \to \infty$ if $n > n^*$

That is, a peer will never share if the number of users downloading from that peer exceeds n^*.

This cost structure has important implications on the equilibrium of the network. Recall that peers with high value content will attract the most number of users. High value users generate more benefit to the network, because they displace more users from other peers. A high number of downloaders, however, can have adverse implications for the high value peer. In particular, if $n > n^*$ then the high value peer will not find it optimal to share. Thus, the marginal benefit a peer *i* provides is $U_0{}^v - U_0^k$ where U_0^k is the utility of the network without peer *i* sharing, and $U_0{}^v$ is the utility of the network with peer *i* sharing. The larger this gap, the more users will switch from other sharing peers to peer *i*. If we define $G = U_0{}^v - U_0^k$ and let n_i be the number of downloaders on peer *i* when it chooses to share, then $\frac{\partial n_i}{\partial G} > 0$. That is, the higher the marginal value provided by peer *i* to the network, the more users will download from *i*, and the higher the chance that $n > n^*$. By assumption, however, if $n > n^*$, peer *i* will not share. The following proposition formalizes this intuition.

Proposition 4: *The larger the gap between the content valuation of a peer and the rest of the network's content, the lower the chance that a high value peer will share its content.*

The proof follows from the fact that a peer with very highly valued content attracts more downloaders, which in turn prevents a peer from sharing if the number of downloaders is above the threshold n^*. This proposition also suggests that highly popular content will not be introduced in a low value network because of the larger value of *G*.

Further, the equilibrium level of sharing on the network depends on the value of n^*. Unlike Section 3.1, where the cost of sharing is the same across users, in the case of heterogeneous sharing costs it is not necessarily the case that the highest value users will share in the equilibrium. In fact, it is conceivable that if peers are highly heterogeneous in their content, then the P2P network will collapse because no peers will find it beneficial to share due to high congestion. Therefore, P2P networks with peers that have similarly valued content will likely encourage higher participation and in turn a more stable network.

6. Dynamic Game: The Effect of Duplication and Redistribution of Content

One unique aspect of P2P networks is the duplication and distribution of content; users who consume network resources by downloading can become providers of those network resources in future periods. This characteristic of peer-to-peer networks distinguishes it from typical public and club goods settings where the endowment is typically discrete and fixed. In this section, we extend the static model developed above to study the effect that duplication of content has on a peer's decision to share her content. We do this by extending the one-period static game of Section 3.1 into a two-period dynamic game.

In a single period game, peers have incentives to share to spatially distribute the demand other peers place on its resources. In a dynamic game with replication, peers have added incentives to share to temporally distribute the demand other peers place on its resources. Thus, we extend our basic model in Section 3.1 in two ways. First, we allow peers to enter and exit the network at the end of the first period. Second, we allow duplication and redistribution of the content.

We first start with a case where there is no duplication of the content but where some proportion of the peers leave and new peers join the network. As in the static model, in period 1, peers with value $v_1 \geq v_2 \geq v_3 \geq .... \geq v_{n-1} \geq v_n$ make a choice of whether to share their content. In period 2, an α proportion of the peers leave the network and are replaced by new peers. The new peers carry content whose value is also drawn from a uniform distribution on the integers $v_1, v_2, ..., v_n$. We assume that the sharing cost c is such that, in first period the top k^* peers share while the rest of the peers free-ride.[8] In the second period, new peers join the network, and all peers again decide whether to share their content. We show that, in this situation, the total value of the network does not change and the same number of users share in both periods.

Proposition 5: *In expectation, the total value provided by the network remains unchanged in the second period, when α proportion of the peers have been replaced by new peers.*

Proof: See the Appendix.

We now consider what happens with both entry and exit of peers and duplication of the content. To do this, we make the following assumptions:

i) User preferences are constant across periods.

ii) In period 1, only one user will be able to download content from a peer, even though n_j downloaders are trying to download the content from peer $j = 1, \dots k$. Therefore, at most two copies of the same content will be available on the network in period 2.

iii) When peers have two units of content in period 2, they share only one of them (which will be the higher value content in equilibrium, because it adds more value to the network than the lower value content does).

Under these assumptions we show that duplication increases network value because low value content will be replaced by higher value content. This redistribution process occurs as follows: Since, by definition, v_1 is the highest value content, any peer j $(j \neq 1)$ that downloads v_1 and does not exit will always share v_1 instead of v_j. On the other hand, v_2 will be shared by all peers except peer 1, who already has $v_1 > v_2$ and hence will share only v_1. Since any peer i can only share one copy of content in period 2, and since the other N -1 users demand that content, content i will be

[8] In other words, we ignore the corner solutions of the equilibrium.

duplicated on any peer j = 1, 2, …, i-1, i+1, … , N with probability $1/(N-1)$. But such a duplication is useful only if the downloaded content $v_i > v_j$ for j = 1, 2, …, i-1, i+1, … , N.

Therefore after period one, the expected distribution of the content is

$$v_1, v_2(1-\frac{1}{N-1})+\frac{1}{N-1}(v_1), v_3(1-\frac{2}{N-1})+\frac{1}{N-1}(v_1+v_2), \ldots, v_n(1-\frac{k}{N-1})+\frac{1}{N-1}(v_1+v_2+..v_k)$$

The second term above calculates the expected value at peer 2 after duplication of the content. In this expression, $\frac{1}{N-1}(v_1)$ means that with probability 1/(N-1) peer 2 will download v_1 and share it in period 2, while $v_2(1-\frac{1}{N-1})$ means that with probability $(1-\frac{1}{N-1})$ peer 2 does not get v_1, so the best she can share in period 2 is v_2. Similar logic holds for the remaining terms.

Because only the top k contents are shared in period 1, only those pieces of content have a chance to be duplicated. Therefore, the expected value at peer k+1 and onwards (until N) will be:

$$v_{k+1}(1-\frac{k}{N-1})+\frac{1}{N-1}(v_1+v_2+..v_k)$$

which reflects the fact that the content v_{k+1} is never duplicated for all $k = 1, \ldots N-1$.

At the end of period 2, α proportion of the peers leave the network and are replaced by new peers. The game can be extended in similar ways to more than two periods.

What happens to the overall value provided by the network when duplication takes place? In proposition 6 we show that duplication allows the network to retain high value content even if the original peer providing the content exits the network. Thus, over time, the total value of the network increases.

Proposition 6: *When content is duplicated and redistributed in a peer-to-peer network and when peers can enter and exit, the value of the network increases in period 2.*

Proof: See the Appendix.

The intuition for this result is that although the original peer carrying the high value content may exit the network, other peers who have downloaded the content from that peer will share the high value content in the next period. Over time, the higher value content will be redistributed to low value peers, and as a result, the network value increases.

Our results therefore suggest that longer tenured networks have an advantage over newer networks. Older networks provide more value (because they retain more of the higher value content) and hence offer more incentives for newer users to join and share their content. The result is somewhat akin to positive externality although the high value in our set-up is due to duplication.

7. Discussion

We examine content sharing in peer-to-peer networks using game theoretic models. In contrast to prior predictions in the literature, we find that networks can sustain free-riding in equilibrium. In spite of free-riding, some peers find it individually rational to share as a way to reduce congestion on other peers they are interested in accessing. We find that typically high value good content is shared before lower value content as long as sharing costs are not too high. However, if sharing costs are high, a peer with high value content will attract too many downloaders and will not want to share. This can happen when a peer carrying very popular content joins a network with mostly lower value content. Thus, our results show that networks where peers carry content of similar value are more stable *ceteris paribus* which is consistent with recent work suggesting that P2P networks should be organized into communities of interest (Asvanund et al. 2004).

We also show that in the absence of external incentives, free-riding on P2P networks will be above the socially optimal level. In this case, reducing the level of free-riding would increase the aggregate utility provided to the network's users. However, in contrast to much of the literature on reducing free-riding, in our model it is not necessarily socially optimal to eliminate free-riding altogether. In some cases, the cost to the individual user from sharing does not justify the benefit they provide to the network.

We consider the effect of external incentives on free-riding behavior and show that the socially optimal level of sharing can be achieved by varying the quality of service provided to users based on whether they share content. Providing better quality of service to users who share content increases the incentives for free-riders to share content. Further, quality of service differentiation acts as a tuning parameter. As quality of service differentiation increases, more users will find it optimal to share and eventually the socially optimal number of sharers can be reached.

Finally, we consider the duplication and redistribution of content on P2P networks and find that such duplication is very beneficial to the network. It allows the low value content to be replaced by high value content which increases the value of the network over time.

Our analysis should have value for entrepreneurs and designers of P2P networks, for copyright holders seeking to protect their copyrights in the presence of P2P networks, and for businesses implementing P2P content management systems. For network designers, our results suggest that, in contrast to some predictions in the literature, networks can persist in spite of free-riding. This suggests that while altruism may play a role in user behavior in these networks (e.g., Gu and Jarvenpaa 2003), some users will share and networks will function even in the absence of altruistic motivations. However, we also show that the aggregate utility of network would be improved in the presence of incentives to encourage more users to share their content. Further, we show that non-priced incentives that impose an implicit cost on non-sharers can proxy for price-based incentives in settings where monetary transfers are infeasible. For copyright holders, our results suggest that increasing the cost of sharing can reduce the number of sharers and above a certain point can lead to network collapse. One way copyright holders appear to be increasing the implicit cost of sharing is by increasing the legal risks to individual network users from sharing copyrighted information, as noted by the recent prosecution of several college students found to be sharing copyrighted materials online. For enterprises implementing P2P network products, our results suggest that forcing all network members to share may not be the socially optimal outcome. The cost of sharing to an individual user (in terms of reduced bandwidth, storage, or processing power) may not justify the value this user would provide to other network members.

This research can be extended in a variety of ways. From an empirical standpoint, one could explore how recent changes to P2P networks have affected network performance vis-à-vis the predictions of our model.[9] Similarly, it would be interesting to analyze the impact of the RIAA's much publicized efforts to raise legal risks from sharing and increase search costs (through fake content) have had on network performance across the board. Further, the Neo Modus network allows individual network operators to determine their own network membership rules. This could create an opportunity for researchers to compare network outcomes as a function of the particular membership rules chosen by a diverse set of network operators. Finally, our model provides an empirically testable hypothesis that users with lower cost of sharing tend to share more. To the extent that an important component of user costs is available bandwidth, one future avenue for research could be to test whether users with broadband connections (e.g., leased line, cable modem, or DSL) are more likely to share content than dial-up users.

[9] For example, Kazaa has recently implemented systems to differentiate quality of service levels based on whether users are sharing content.

References

Adar, E, B. A. Huberman. 2000. Free-riding on Gnutella. First Monday 5(10).

Asvanund, Atip, Karen Clay, Ramayya Krishnan, Michael D. Smith. 2004. An Empirical Analysis of Network Externalities in Peer-To-Peer Music Sharing Networks. Information Systems Research, 15(2) 155-174.

Bergstrom, T., Blume, L. and Varian, H. 1986. On The Private Provision of Public Goods. Journal of Public Economics, 29, 25-49.

Butler, B.S. 2001. Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures. Information Systems Research 12(4) 346-362.

Chandan, S., Hogendorn C. 2001. The Bucket Brigade: Pricing and Network Externalities in Peer-to-Peer Communications Networks. Telecommunications Policy Research Conference, Alexandria, VA, October 27-29.

Dixit, A and Skeath, S. 1999. *Games of Strategy*. New York. W. W. Norton &Co.

araj, S., McLure Wasko, M. 2002. The Web of Knowledge: An Investigation of Knowledge Exchange in Networks of Practice. Working Paper, University of Maryland, College Park, MD.

Feldman, M., K. Lai, J. Chuang, and I. Stoica (2003), “Quantifying Disincentives in Peer-to-Peer Networks”, Workshop on Economics of Peer to Peer Network.

Golle, P., K. Leyton-Brown, I. Mironov. 2001. Incentives for Sharing in Peer-to-Peer Networks. Working Paper, Stanford University, Palo Alto, CA.

Gradstein, M., Nitzan, S. 1990. Binary Participation and Incremental Provision of Public Goods. *Social Choice and Welfare*, 7, 171-92.

Gradstein, M. 1992. Time Dynamics and Incomplete Information in the Private Provsion of Public Goods. *Journal of Political Economy*, 100, 581-97.

Groves, T., Ledyard, J. 1977. Optimal Allocation of Public Goods: A Solution to the ‘Free Rider’ Problem. *Econometrica*, 45, 783-809.

Gu, B. and S. Jarvenpaa. 2003. Are Contributions to P2P Technical Forums Private or Public Goods? - An Empirical Investigation. Working Paper, University of Texas at Austin, Austin, TX.

Hardin, G. 1968. The Tragedy of the Commons. *Science* **162** 1243-48.

Hindriks, J. and Pancs, R. 2001. Free Riding on Altruism and Group Size. Working Paper, University of London.

Kamvar, Sepandar, Beverly Yang, Hector Garcia-Molina. 2003. Addressing the Non-Cooperation Problem in Competitive P2P Systems. Working Paper, Stanford University, Palo Alto, CA.

Kollock, P. 1999. The Economies of Online Cooperation: gifts and Public Goods in Cyberspace. Chapter 9, *Communities in Cyberspace*, M.A. Smith and P. Kollock eds. London, Routledge.

Konrad, K.A.. 1994. The Strategic Advantage of Being Poor: Private and Public Provision of Public Goods. *Economica*, 61, 79-92.

Kraut, R. Steinfield, C. Chan, A. Butler, B. and Hoag, A. 1998. Coordination and Virtualization: The Role of Electronic Networks and Personal Relationships, *Journal of Computer Mediated Communication* **3**(4).

Krishnan R., Smith, M., and Telang, R. 2003. The Economics of Peer-To-Peer Networks. *Journal of Information Technology Theory and Applications*, **5**(3) 31-44.

Li, Yung-Ming, Yong Tan, Prabuddha De. 2004. Self-Organized Formation and Evolution of Peer-to-Peer Networks. Working Paper. University of Washington, Seattle, WA.

Mas-Colell, A., M.D. Whinston, J.R.Green. 1995. *Microeconomic Thoery*. Oxford University Press, Oxford, New York.

Nitzman, S. and Romano, R.E. 1990. Private Provision of a Discrete Public Good with Uncertain Cost. *Journal of Public Economics*, 42, 357-70.

Olson, M. 1965. *The Logic of Collective Action.* Harvard University Press, Cambridge, MA.

Palfrey, T. and Rosentral, H. 1984. Participation and the Provision of Discrete Public Goods: A Strategic Analysis. *Journal of Public Economics*, 24, 171-193.

Palfrey, T. and Rosentral, H. 1988. Private Incentives in Social Dilemmas: The Effects of Incomplete Information and Altruism, *Journal of Public Economics* 35 309-332.

Ranganathan, K, Ripeanu M, Sarin A, and Foster I. (2003), “To Share or Not to Share: An Analysis of Incentives to Contribute in Collaborative File Sharing Environments”, *Working paper*, University of Chicago.

Sandler, T. 1992. *Collective Action: Theory and Applications*, University of Michigan Press, Ann Arbor, Michigan.

Sharp, J. 1997. Communities of Practice: A Review of the Literature. March.

Strahilevitz, L.J. 2002. Charismatic Code, Social Norms, and the Emergence of Cooperation on the File-Swapping Networks. John M. Olin Law and Economics Working Paper, Number 162, University of Chicago, Chicago, IL, September.

Vishnumurthy, V., S. Chandrakumar, E.G. Sirer. 2003. KARMA: A Secure Economic Framework for Peer-to-Peer Resource Sharing. Working Paper, Cornell University, Ithaca, NY.

Appendix: Proofs of Propositions

Proof of Lemma 1:

Suppose in equilibrium peer i is attempting to download from peer j and gets utility $U_{ij} = f\big(v_j, n_j(v_j, k; N)\big)$. Suppose another peer j' provides utility $U_{ij'} = f\big(v_{j'}, n_{j'}(v_{j'}, k; N)\big)$. Here v_j and $v_{j'}$ are the value of their content and n_j and $n_{j'}$ are the number of peers trying to download from them. We prove by contradiction. Suppose that $U_{ij} \neq U_{ij'}$. Without loss of generality, let $U_{ij} < U_{ij'}$, then it must be that peer i should shift to peer j' because it will get higher utility. Since $\frac{\partial U}{\partial n} < 0$, U_j should increase and $U_{j'}$ should decrease until $U_j \cong U_{j'}$.[10] In short, $U_{ij} \neq U_{ij'}$ cannot be an equilibrium.

Proof of Proposition 1:

Consider the decision of the k^{th} user who decides to contribute content. For this user to share, it must be the case that the increased probability of successfully downloading content from the other k-1 sharers (i.e. the first term on the right hand side of (A1)) net the cost of sharing is greater than the utility from not sharing (i.e. the right hand side of (A1)). Thus, denoting $k^* = \sum_{i=1}^{n} S_i^*$ and $k^* - 1 = \sum_{\substack{j=1 \\ j \neq i}}^{n} S_j^*$, we have,

$$U_0^{k*} - c \geq U_0^{k*-1} \tag{A1}$$

Further, consider the decision problem of the $(k^* + 1)^{th}$ user who does not contribute her content, we have,

---

[10] Strictly speaking, peers do not provide exactly the same utility in equilibrium. However, as long as $f\big(v_j, n_j(v_j, k; N) + 1\big) < f\big(v_{j'}, n_{j'}(v_{j'}, k; N)\big)$ and $f\big(v_j, n_j(v_j, k; N)\big) > f\big(v_{j'}, n_{j'}(v_{j'}, k; N) + 1\big)$, no peer will have an incentive to switch. For large k, N and n, $U_j = U_{j'}$.

$$U_0^{k^*+1} - c < U_0^{k^*} \tag{A2}$$

Combining (A1) and (A2), we get,[11]

$$U_0^{k^*+1} - U_0^{k^*} < c < U_0^{k^*} - U_0^{k^*-1} \tag{A3}$$

Thus, a pure strategy equilibrium always exists if and only if

$$U_0^{k^*+1} - U_0^{k^*} \le U_0^{k^*} - U_0^{k^*-1}$$

The assumption of concavity of U_o^k with respect to k is necessary and sufficient for (A3) to hold. Noting that $0 \le k \le n$, there are three different scenarios in which the network can attain equilibrium: When $c < U_0^N - U_0^{N-1}$, there is a unique pure strategy equilibrium, where everyone contributes; that is $(s_1^*, s_2^*, ..., s_n^*)$ = (1,1,…,1). Similarly when $c > U_0^2 - U_0^1$, there is a unique pure strategy equilibrium, where $(s_1^*, s_2^*, ..., s_n^*)$ = (0,0,…,0), that is nobody contributes.[12] For levels of c between these two values (i.e., $U_0^{k+1} - U_0^k < c < U_0^k - U_0^{k-1}$) there exists a k^* satisfying (A3). Since higher value peers contribute before lower value peers do, this is a pure strategy equilibrium. In case of a strict inequality in the value of content, this is a unique equilibrium as well.

Proof of Lemma 3: Let $H(k) \equiv U_0^k - U_0^{k-1}$. Then equilibrium condition (3) becomes $H(k^*+1) \le c \le H(k^*)$. By the Inverse Function Theorem, $\frac{\partial k^*}{\partial c} = \frac{\partial H^{-1}(c)}{\partial c} = \frac{1}{H'(H^{-1}(c))}$. And this is negative because of the concavity of U_0 with respect to k.

<u>Proof of Proposition 2:</u>

Denote the optimal k in this case as k_s^*. Using logic analogous to that used in section 3.1, the first order condition (FOC) of the social planner's problem (SP) implies that

[11] By (2), at least one of the inequalities in (A4) will hold strictly and thus the interior solution is unique.

[12] Note that k=1 is not a stable equilibrium, since the peer will be better off not sharing if she is the only contributor.

$$U_0^{k_s^*+1} - U_0^{k_s^*} \le \frac{c}{n} \le U_0^{k_s^*} - U_0^{k_s^*-1} \tag{A4}$$

Recall again the equilibrium of individual optimization (IO) is determined by (A3)

$$U_0^{k+1} - U_0^{k} \le c \le U_0^{k} - U_0^{k-1}$$

Thus, by Lemma 1 it is then clear that $k^* \le k_s^*$. That is, content is underprovided compared to the social optimum.

<u>Proof of Proposition 3:</u>

First, note that in the presence of incentives, when $c < U_0^n - U_0^{n-1}$ all users will share. Comparing this condition to the condition in proposition 1 where all users share, and noting that $p<1$, it is straightforward to show that the point at which all users share under our incentive structure occurs at a higher cost level than in the absence of incentives. Therefore, for any given cost level such that some users free-ride, $k_p^* > k^*$.

<u>Proof of Proposition 5:</u>

Since k^* peers share in period 1, the total utility provided by the network is simply $U_0^{k^*}$ as shown in proposition 1. Now consider the case where each peer has α probability of leaving the network and being replaced by a new peer. We want to find the distribution of the content after this peers' replacement. That is, we are interested in finding the expected number of peers who will carry content v_1, v_2, ..,v_n after replacement.

First we calculate the expected number of peers who will have content v_1. Originally peer 1 has the content. Since the probability of a replaced peer having content v_1 is ($1/N$), the number of peers having content v_1 is binomially distributed. The mean of the binomial distribution is

$Np = N\left(\frac{1}{N}\right) = 1$. Since each peer has α chance of being replaced, the expected number of peers having content v_1 is $E(.) = (1-\alpha) + \alpha\left(\frac{1}{N}\right)N = 1$

Here $(1-\alpha)$ is the probability that peer 1, which originally carries content v_1, is not replaced. Analogous logic shows that the expected number of peers having content v_1, v_2, v_3, … v_n is 1 as well.

Therefore, in expectation, the distribution of value across peers remains unchanged and hence in equilibrium, top k^* peers to share their content and the network provides utility $U_0^{k^*}$, which is the same as before replacement.

Proof of Proposition 6:

With probability α each peer exits the network. With probability (1-α), each peer remains in the network. When the peer exits, an incoming peer will have a value drawn from U~[v_1, v_n]. Following the previous proof, we want to find the expected number of peers having content v_1, v_2, ..,v_n after replacement where if more peers have higher value content then the overall value of the network will be higher.

Recall that after duplication the value of the content across peers is

$$v_1, v_2(1-\frac{1}{N-1}) + \frac{1}{N-1}(v_1), v_3(1-\frac{2}{N-1}) + \frac{1}{N-1}(v_1+v_2), \ldots, v_n(1-\frac{k}{N-1}) + \frac{1}{N-1}(v_1+v_2+..v_k) \quad \text{(A5)}$$

Consider content v_1. The expected number of peers having content v_1 at the end of period 1 is:

$$\left[1 + \frac{1}{N-1} + \frac{1}{N-1} + \ldots \frac{1}{N-1}\right] = 2$$

For peers 2 to N, there is a probability $(1/(N\text{-}1))$ that they download content v_1 in the first period, which yields the terms in the expression above. With probability α, a peer exits and is replaced by a new peer. Therefore the expected number of peers having content v_1 after replacement is $(\alpha)\left[\frac{1}{N}+\frac{1}{N}+\frac{1}{N}+\ldots\frac{1}{N}\right]=\alpha$. Each term in the bracket is $(1/N)$ because when the peer exits and is replaced by a new peer, there is only $1/N$ probability that the peer will have content v_1. Therefore the expected number of peers having content v_1 is the sum of $(1-\alpha)\left[1+\frac{1}{N-1}+\frac{1}{N-1}+\ldots\frac{1}{N-1}\right]$ and $(\alpha)\left[\frac{1}{N}+\frac{1}{N}+\frac{1}{N}+\ldots\frac{1}{N}\right]$ which is $2 - \alpha$. Note that since $0 < \alpha < 1$, more peers now have a high value content v_1 than before.

Following the same logic we can find the expected number of peers having content v_2 as

$$(1-\alpha)\left[0+(1-\frac{1}{N-1})+\frac{1}{N-1}+\ldots\frac{1}{N-1}\right]+\alpha\left[\frac{1}{N}+\frac{1}{N}+..+\frac{1}{N}\right]=(1-\alpha)\left(1+\frac{N-3}{N-1}\right)+\alpha$$

The first term in the first bracket is 0, because peer v_1 will never share value v_2 unless it exits and is replaced by a new peer. The second term in the first bracket $(1-\frac{1}{N-1})$ comes from (A5). The expression above simplifies to $1+(1-\alpha)\frac{N-3}{N-1}$. Therefore, we can write the expected number of peers having content v_j as E(.) = $1+(1-\alpha)\frac{N-3(j-1)}{N-1}$ where j can be 2.,…,k. Note that the expected number of peers having content v_j is declining in j. This is intuitive because the lower the value of the peer, the lower chance that it will be duplicated. After j ($k > j > N$), the expected number of peers having content v_j becomes less than 1.

Also note that only the top k content is duplicated (because only the top k peers share their content). Therefore the expected number of peers having content $j = k+1,..,N$ is simply

$$(1-\alpha)\left[1-\frac{k}{N-1}\right]+\alpha\left[\frac{1}{N}+\frac{1}{N}+\ldots\frac{1}{N}\right]=(1-\alpha)\left[1-\frac{k}{N-1}\right]+\alpha=1-(1-\alpha)\frac{k}{N-1}$$

In sum, after duplication, more peers carry higher value content and fewer peers carry low value content.

Recall that in the case of no duplication, k^* peers share and the total value of the network is U_0^{k*}. Now consider the equilibrium sharing level, and hence the total value of the network, in the case of duplication. In the case of duplication, the value of the content v_j has not changed but there are more peers carrying high value content. If the same k^* peers share, then $U_0^{k*}(d) > U_0^{k*}(nd)$, where *d* and *nd* represent *duplication* and *no duplication*, respectively. This inequality must hold because for a fixed k^*, and the top *k* peers in the duplication case have higher value and we know that U_0^{k*} is increasing in *v*.

It is also possible that in the case of duplication, $\hat{k} < k^*$ number of peers share while the rest free-ride. But even in this case, $U_0^{\hat{k}}(d) > U_0^{k*}(nd)$. To see this, consider peer k^* in the case of no duplication. Since peer k^* shares its content, the marginal benefit of sharing is higher than the cost $U_0^{k*} - U_0^{k*-1} > c$. Now consider peer k^* in the case of duplication. Peer k^* does not share because the marginal benefit of sharing is $U_0^{\hat{k}}(k*) - U_0^{\hat{k}} < c$, where $U_0^{\hat{k}}(k^*)$ is the value when k^* chooses to share and $\hat{k}$ is already sharing. Since the value of the $k^{*\text{ th}}$ peer is higher in the duplication case than in the no duplication case, $U_0^{\hat{k}}(k*) - U_0^{\hat{k}} < c$ will hold only if $U_0^{\hat{k}}(d) > U_0^{k*}(nd)$.

Therefore in any equilibrium we have $U_0(d) > U_0(nd)$.