Why Share in Peer-to-Peer Networks?

Lian Jian
School of Information
University of Michigan
Ann Arbor, MI 48109, USA
ljian@umich.edu

Jeffrey K. MacKie-Mason
School of Information, and Dept. of Economics
University of Michigan
Ann Arbor, MI 48109, USA
jmm@umich.edu

ABSTRACT

Prior theory and empirical work emphasize the enormous free-riding problem facing peer-to-peer (P2P) sharing networks. Nonetheless, many P2P networks thrive. We explore two possible explanations that do not rely on altruism or explicit mechanisms imposed on the network: direct and indirect private incentives for the provision of public goods. The direct incentive is a traffic redistribution effect that advantages the sharing peer. We find this incentive is likely insufficient to motivate equilibrium content sharing in large networks. We then approach P2P networks as a graph-theoretic problem and present sufficient conditions for sharing and free-riding to co-exist due to indirect incentives we call *generalized reciprocity*.

Categories and Subject Descriptors

C.2.4 [**Computer-Communication Networks**]: Distributed Systems; H.1.2 [**Models and Principles**]: User/Machine Systems—*human factors*; H.4.m [**Information Systems**]: Miscellaneous

General Terms

Design, Human Factors, Theory, Economics

Keywords

peer-to-peer, p2p, networks, file-sharing

1. INTRODUCTION

Studies of peer-to-peer (P2P) networks as static games predict these systems should suffer from enormous free-riding—peers download but do not upload—in the absence of altruism or explicit incentive mechanisms to encourage content uploading [11]. The fact that many peers free ride is also empirically confirmed [13, 1, 7]. However, in practice P2P networks such as eDonkey, KaZaa, and Gnutella, persist and flourish despite free-riding. One possible explanation for this

puzzling phenomenon is that altruism might sustain these networks. Rather than rely on this deus ex machina, we explore two alternative explanations: direct and indirect incentives for the private provision of public goods.[1][2] Our immediate goal is to understand economic conditions under which networks of self-interested participants might be sustainable despite equilibrium free-riding. Our ultimate goal is to develop a plausible model of P2P behavior in order to evaluate various proposed mechanisms to increase sharing, and to develop our own mechanism for the same.

In the next section, we investigate a direct private incentive to provide public goods proposed by Krishnan et al. [9], who suggest that sharing redistributes traffic in the network to the advantage of the sharing peer. We explore, in Section 3, generalized reciprocity as an indirect incentive explanation of both sharing and free riding on P2P networks. We close with a discussion of limitations in our work, and plans to continue this research.

2. DIRECT PRIVATE INCENTIVES

Providing files (*sharing*) to a P2P network is an instance of the *private provision of public goods* [14].[3] Sharing provides direct benefits to others for which, in the absence of an explicit incentive mechanism, the sharing peer is not compensated. One suggested explanation for the nonetheless observed sharing is that in the process of providing a benefit to other users, a sharing peer is simultaneously obtaining a direct private benefit, similar to the personal incentive to donate, for example, to support a public radio station. Krishnan et al. [9] model a particular form of this, which we call the "offload effect": sharing redistributes traffic in the network to the advantage of the sharing peer. In a P2P network, suppose peers A and B each want a different file, but both files are available from peer C. If peer A has the file

---

[1]We are not claiming that altruism does not exist or is unimportant. Rather, taken as a primitive, it is not susceptible to analysis, and does not help answer design questions. For example, if sharing occurs solely due to axiomatic tastes for altruism, we will have nothing to say about how to encourage increased sharing, unless we have a story about *why* people want to be altruistic: that is, what incentives do they have for sharing?

[2]Some P2P protocols impose "altruistic" (always on) sharing as a default setting. It may be a good design principle to encourage people to act as if they were altruistic, but that leaves open the incentives question: why do they not change the default setting?

[3]We adopt the widely-used convention of referring to uploading as *sharing*.

that B desires, by offering to share with B agent A may get her file sooner from C, by offloading some of the demand on C.[4]

Krishnan et al. show that an offload effect could support a network in which all peers share is a dominant strategy equilibrium. After modeling the traffic redistribution and network congestion more precisely, we find it implausible that the offload effect alone is sufficient to motivate the amount of sharing seen on successful P2P networks.

2.1 Modeling offloading

We construct a four-period game. In period 1, $n \geq 3$ players join the network. In period 2, each player chooses whether to share or not, at a fixed cost of c or zero, respectively. Capacity is fixed and scaled so that a player can share at most one unit in a sharing period. In period 3, each player requests a unit of content from the network. The network protocol randomly assigns each request to one player who has decided to share in period 2. For consistency we adopt the important simplifying assumption made by Krishnan et al. that every node has at least one file wanted by any other node. Suppose $k \geq 2$ players have decided to share their content in period 2; then the probability that player i's request is assigned to sharing player s is $\frac{1}{k}$. If multiple requests are assigned to a sharing player, she randomly chooses one to serve. In period 4, files are shared, and payoffs are realized.

Suppose in period 3, i's request, together with m other requests, has been assigned to player s. s will pick i's request to serve with probability $\frac{1}{m+1}$. Given that the event of any peer's request being assigned to a sharing peer is $\frac{1}{k}$, this event follows a Bernoulli distribution, and the event that m other players will be assigned to node s follows a binomial distribution, $m \sim b(n-2, p)$.[5]

Consider an arbitrary node i deciding whether to share. She calculates the expected value from sharing or not (u_i^S, u_i^N). These values are defined as the probability of obtaining one unit of content. If i shares, the total number of sharing nodes is $k + 1$; if she doesn't, it's k. Thus, the probability that another peer will choose the same source node as i is $p = \frac{1}{k+1}$ if i shares, and $q = \frac{1}{k}$ if i does not share. Now we need to calculate the expected value for i of being served by s, or $\mathrm{E}[(m+1)^{-1}]$. We calculate this as sum of the probabilities of m taking on each possible value on $\{0, \ldots, n-2\}$, times the probability that i gets a file from s when there are exactly m other demanders on s. Thus, the expected values, u_i^S and u_i^N, are:[6]

$$
\begin{aligned}
u_i^S(n,p) &= \Sigma_{m=0}^{n-2} C_{n-2}^m p^m (1-p)^{n-2-m} \frac{1}{m+1} \\
&= \frac{1 - (1-p)^{n-1}}{(n-1)p} \quad\quad (1)
\end{aligned}
$$

and similarly,

---

[4]We take a game-theoretic approach to studying incentives in P2P networks, and will use graph theory in the next section, so we use *peer*, *node* and *player* interchangeably.

[5]The number of trials is $n - 2$ because there are n downloading agents, but the set of possible *other* agents than i downloading from s excludes i and s.

[6]Equations (1) and (2) are a simplified approximation. The difference is qualitatively unimportant; see section Lemma 3 - explaining footnote 6 in the appendix.

$$
u_i^N(n,q) = \frac{1 - (1-q)^{n-1}}{(n-1)q}, \quad\quad (2)
$$

where C_x^y is the number of combinations "x choose y". We define the marginal benefit of sharing (MBS) as the difference between u_i^S and u_i^N:

$$
MBS_i(n,p,q) = u_i^S(n,p) - u_i^N(n,q). \quad\quad (3)
$$

2.2 Privately provided public good

If the sharing cost is low enough, nodes will share to obtain the offloading benefit.

LEMMA 1. $MBS_i(n,p,q) > 0$.

PROOF. See appendix. □

Lemma 1 implies if the cost of sharing, c, is less than $MBS(n,p,q)$, there is a dominant strategy equilibrium in which all peers choose to share. This verifies that our model is consistent with the results in [9].

2.3 How large is the offloading benefit?

It is straightforward to show that MBS is decreasing in the number of other sharing nodes, k, so the incentive for a marginal node to share decreases in larger networks. But the equilibrium outcome of the game depends on the relative values of c and MBS. Without an empirical estimate for c, it is difficult to determine whether the offload effect is meaningful for a P2P network. We can, however, gain an appreciation for the magnitude of the offloading effect by analyzing it as a percentage increase in a non-sharing peer's utility. Denote this increase by G_i:

$$
G_i(n,p,q) = \frac{MBS(n,p,q)}{u_i^N(n,q)} \quad\quad (4)
$$

Lemma 2 below characterizes the asymptotic properties of G_i. As $n \to \infty$, G_i converges to $\frac{1}{k}$. For example, when $k = 30$, by sharing her content, a player can only increase the probability of obtaining one unit of content by 3.3%. It seems implausible that in medium or large networks this small gain would motivate many peers to share their content. Further, since $k \leq n$ by definition, we see that G_i converges to zero as k increases, which means the benefit of sharing vanishes the more other peers are sharing.

LEMMA 2.

$$
\lim_{n \to \infty} G_i = \frac{1}{k}.
$$

PROOF. See appendix. □

In Figure 1 we plot G_i against n, for various small values of k. This illustrates our point that the gain of sharing becomes independent of the number of peers in large networks, and it decreases in the number of nodes that are sharing their content. We conclude that although the offloading effect may play some role in P2P networks, the private incentives it suggests are likely insufficient to motivate equilibrium content sharing in large networks.
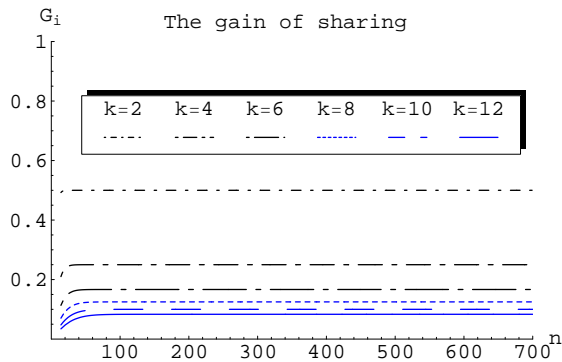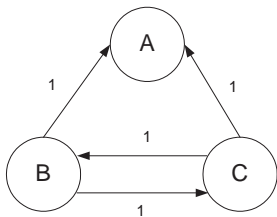
Figure 1: The gain of sharing for player i.

Figure 2: A simple example

3. INDIRECT PRIVATE INCENTIVES: GENERALIZED RECIPROCITY

We turn to a different candidate explanation for sharing: generalized reciprocity in a repeated game.[7] In BitTorrent, a form of direct reciprocity is implemented by embedding a tit-for-tat type of strategy in the client software [3]. This provides a form of direct incentive for uploading, similar to the offloading incentive we studied in the previous section. One type of indirect incentive for contributing to the public good is "generalized reciprocity" [10]:

> I'll do this for you without expecting anything specific back from you, in the confident expectation that someone else will do something for me down the road. (p. 21)

Generalized reciprocity also plays a role in motivating knowledge sharing [8]. In a P2P network, generalized reciprocity may be loosely described as a cycle in the directed graph in which each peer contributes to second peer, but receives a contribution from a third peer. We shall formally characterize conditions on the topology of the graph such that it has an equilibrium in which some self-interested peers contribute while others free-ride. Generalized reciprocity can arise when direct reciprocity is impossible, for example when demands between node pairs are very asymmetric.

---

[7]Contribution to public goods may be motivated by various indirect incentives other than generalized reciprocity, e.g., reputation, self-image, and popularity in the community. To implement such motivators, the system requires some kind of individual identity, which is often not available in peer-to-peer file-sharing systems. Participants usually stay anonymous. For systems in which these motivations are available, we speculate that there will be more contributors than in systems for which no such motivations exist.

We suppose there is no private benefit from sharing (i.e., no altruism, and no offloading effect), but that peers are interconnected through a network topology, and anticipate participating for an indefinite length of time. Feigenbaum and Shenker [5] suggested graph theory to model incentives in network problems because peer incentives might affect the formation of the graph. We follow Shneidman and Parkes [12] who suggest graph configuration may affect incentive structures. In particular, we characterize a family of graphs that support a generalized reciprocity equilibrium. Like us, Afergan and Sami [2] use repeated games theory to study problems on network topologies.

We illustrate with the simple graph in Figure 2. Each labeled, directed link represents the direction and volume of the traffic between the two end nodes of the link. In a repeated game of indefinite duration, B and C download a file from each other and both upload a file to A. A, however, is a free rider. Suppose peers restrict themselves to either sharing with every node or not sharing at all.[8] Suppose further that the benefit of receiving one unit of content is significantly higher than the cost of sharing it. With these assumptions B and C sharing is an equilibrium as long as they are receiving enough content. If say, B stops sharing, C will find it unprofitable to share, hence will also stop sharing. Thus the network breaks down due to B's deviation.

3.1 Definitions and assumptions

We model peers' interactions as an infinitely repeated game with a fixed time discount factor δ adopted by all peers.[9] A set of demand relationships among the n peers in the network is given exogenously, and remains constant through out. These relationships can be represented as a directed graph, D. Loosely speaking, a directed graph is a set of *nodes* connected by directed *edges* [4]. A *path* is a sequence of consecutive nodes and edges, with no nodes repeated. A path which ends at the node it begins is called a *cycle*. Two cycles are *independent* if they do not share any nodes in common. A graph is *connected* if there is an undirected path connecting every pair of nodes. A graph that is not connected can be divided into connected *components*, each of which is a connected subgraph. For example, Figure 3 is a directed connected graph with independent cycles $(1, a_{12}, 2, a_{23}, 3, a_{34}, 4, a_{41})$ and $(10, a_{10_11}, 11, a_{11_10})$. The link label values are the demand quantities. The graph has an equivalent representation as an $n \times n$ adjacency matrix with each element σ_{ij} the demand from peer j to peer i.

3.2 Game setup

For simplicity, we define matrix B as a binary demand matrix obtained by converting the positive link intensities in D into 1, with element b_{ij} the demand from j to i. We focus on an arbitrary *component* of the graph with the adjacency matrix K associated with it. K is thus connected and consists of k players. K remains constant through out the game and its member peers have complete information of K. In each stage game, permissible actions for player i, a_{ij}, are defined as follows,

---

[8]We remove this assumption for Proposition 2.

[9]With an appropriate increase to the discount rate, we can accommodate a finite but random time in the network, rather than an infinite horizon.
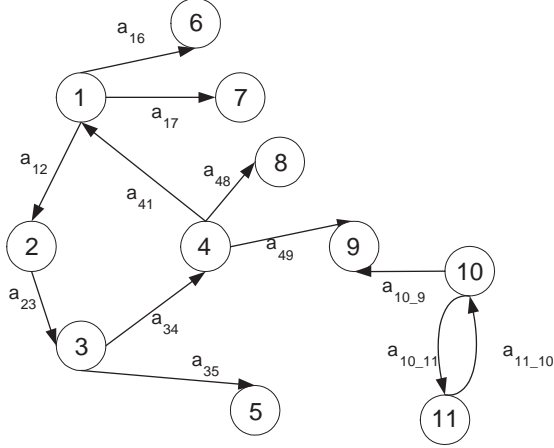
Figure 3: An example of equilibrium action graph

$$a_{ij} = \begin{cases} 1 & \text{if } \sigma_{ij} \neq 0 \text{ and } i \text{ shares to } j \\ 0 & \text{if } \sigma_{ij} = 0 \text{ or } i \text{ does not share to } j \end{cases}$$

In each round, the stage game is played and then payoffs are realized. Players observe all players' actions in the previous round before the next round starts. Each peer receives a positive value of v when she obtains a unit of file, and incurs a positive cost of c when she shares a unit of file to any other peer.

In round t, the actions chosen by the peers constitute a $k \times k$ action matrix, A, which again corresponds to a directed graph. For node i, $\Sigma_{i=1}^{k} a_{ij}$ is its out-degree and $\Sigma_{i=1}^{k} a_{ji}$ is its in-degree. When $\Sigma_{i=1}^{k} a_{ij} = 0$, which means peer i does not share her file to any other peer, she is considered a *free-rider*; otherwise she is considered a *sharing peer*. We also define a parameter ρ_i as the ratio between i's out-degree and in-degree, $\rho_i = \frac{\Sigma_{i=1}^{k} a_{ij}}{\Sigma_{i=1}^{k} a_{ji}}$.

3.3 Equilibrium analysis

We propose that an action matrix A^* can be sustained in equilibrium, if its corresponding graph satisfies the following properties,

P.1 No nodes have an in-degree of zero, and $\max_i \{\rho_i\} < \frac{\delta v}{c}$.

P.2 Any two cycles in the graph are independent.

P.3 Any leaf node is connected to a node that participates in a cycle.

Figure 3 shows a graph that satisfies properties P.1–P.3. The condition on the out-degree/in-degree ratio in P.1 limits consideration to networks in which users get enough net benefit that participating is better for them than is dropping out of the network. This condition can surely be relaxed to accommodate altruistic users. Property P.2 is purely to simplify the analysis, and we know from examples that there are networks in which users participate in more than one cycle and yet the result of our proposition still holds. We are
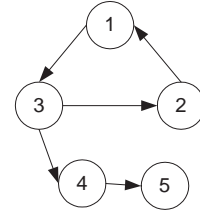

Figure 4: An example of an unsustainable graph

working to relax this condition in ongoing research. Property P.3 rules out a type of agents who are not leaf nodes and do not participate in any cycles, e.g., node 4 in Figure 4. Node 4 would always be better off to stop uploading to node 5 since there is no generalized return on the contribution.

We claim that free-riding on such a graph may exist in equilibrium even without altruistic players or the offloading effect. We formalize this intuition in Proposition 1, as a *subgame perfect Nash equilibrium* (SPNE). A profile of strategies constitutes a Nash equilibrium if given all other players' strategies, no player can benefit from deviating from her own strategy. A profile of strategies is a SPNE if it is a Nash equilibrium of the game itself, and if it induces a Nash equilibrium in every subgame [6].

PROPOSITION 1. *In an infinitely repeated game with the afore-mentioned stage game, if the action matrix A satisfies property P.1–P.3, there exists a SPNE which can have both sharing peers and free-riders.*

PROOF. Two types of nodes in any graph A satisfy properties P.1 $\sim$ P.3. We label the nodes on the cycle as *cycle nodes* and the nodes that do not have child nodes as *leaf nodes*. We restricts peers in each round to play either Share or Not Share with all demanding nodes: if i plays Share, $a_{ij} = b_{ij}, \forall j \neq i$; and if i plays Not Share, $a_{ij} = 0, \forall j$.

We consider two peer strategies, Not Share and the local grim trigger strategy (LGTS). In LGTS peer i plays Share in the first round and continues sharing as long as $\rho_i < \frac{\delta v}{c}$. We show that a strategy profile in which the cycle nodes play LGTS and the leaf nodes play Not Share is a SPNE. First, leaf nodes, by playing Not Share while receiving value from their parent nodes do not have any incentive to deviate.

Second, cycle node i can either follow LGTS or deviate by playing Not Share. We calculate the continuation payoffs of each from round t onwards, as u_i^t. If she follows the equilibrium strategy, LGTS, her continuation payoff is,

$$u_i^t = v\Sigma_{i=1}^{k} a_{ji} - c\Sigma_{i=1}^{k} a_{ij} \tag{5}$$

If player i deviates from LGTS in round t, the other nodes in the same cycle will know before round $t+1$ that she has deviated. Therefore in round $t+1$ no nodes will share to her. This is due to P.2, which implies that a peer belongs to no more than one cycle, such that once one peer deviates from LGTS, the cycle is going to be broken. Foreseeing this happening, no peer in the cycle will share in round $t+1$. Thus player i's continuation payoff is,

$$u_i^t = (1-\delta)v\Sigma_{i=1}^{k} a_{ji} \tag{6}$$

The cycle nodes will choose to follow LGTS if the following inequality holds,

$$v\Sigma_{i=1}^{k} a_{ji} - c\Sigma_{i=1}^{k} a_{ij} > (1-\delta)v\Sigma_{i=1}^{k} a_{ji}, \tag{7}$$
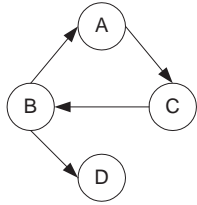
Figure 5: An equilibrium example

which is equivalent to ,

$$\rho_i < \frac{\delta v}{c}. \qquad (8)$$

Inequality (8) is satisfied by property P.1. For completeness, we can easily verify that once a cycle node or a leaf node has deviated, there is no incentive for her to return to the equilibrium strategy. In this equilibrium, the cycle nodes are sharing peers and the leaf nodes are free-riders.

If P.3 is not satisfied, then there will be at least one node that is neither a cycle nor a leaf node, such as node 4 in Figure 4.[10] A graph with such types of nodes will not be an SPNE because these nodes will be unambiguously better off not uploading any files, and thus will deviate from the proposed equilibrium. □

Thus even without altruistic peers, an offloading effect or an explicit incentive mechanism to encourage sharing, sharing can exist due to generalized reciprocity. Moreover, free-riding may exist too. The intuition is simple: peers care a lot about fulfilling their demands, and the cost of sharing is low, so they can tolerate free-riding to a certain extent. More free-riding does not occur because of the threat of a *local grim trigger strategy* (LGTS)[11]: if a node stops uploading to A, node A will leave the network forever, which through the generalized reciprocity cycle punishes the miscreant node, discouraging it from free-riding in the first place (see proof).

We derived Proposition 1 under restrictive conditions: all nodes could observe all flows (the *flow topology*), and nodes may only choose from a strategy space restricted to either upload to every requester, or upload to none. These two assumptions taken together are clearly not very general: if nodes know the entire flow topology then why punish all requesters when a single node deviates? Likewise, if node 4 knows node 8 is a free-rider in Figure 3, why not cut off only node 8 rather than all nodes?

We are working on a model of generalized reciprocity with incomplete information about the flow topology, and with an unrestricted space of strategies. These assumptions seem reasonable for the pseudonymous Internet. We have one preliminary result that illustrates how generalized reciprocity can support P2P networks with equilibrium free-riding in more general settings. We assume peers only know the flows in which they participate and each peer selectively shares to other peers to maximize her value.

PROPOSITION 2. *In an infinitely repeated game, with only local knowledge of the flow graph and selective strategies, the flow graph depicted in Figure 5 is a weak perfect Bayesian equilibrium.*

PROOF. See appendix. □

4. DISCUSSION

We have shown the existence of an equilibrium in a constrained family of network topologies, under two different game forms. Both cases are restrictive, and we would like to relax our assumptions to obtain more general results. We would also like to characterize the *set* of equilibria to assess the plausibility of outcomes with a mixture of sharing and free-riding. Further, we would like to characterize other families of network topologies, to uncover those features (size, connectedness, overlapping cycles, etc.) that affect the equilibrium configurations. Of course, we would also like to address the question we asked about the offloading effect: is generalized reciprocity important enough to explain the amount of sharing we see in large networks?

Our ultimate goal is to use the model as a principled foundation to explore the design and performance of various incentive mechanisms to encourage sharing in P2P networks.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] E. Adar and B. A. Huberman. Free riding on Gnutella. *First Monday*, 5(10), 2000.

[2] M. Afergan and R. Sami. Repeated-game modeling of multicast overlays. In *INFOCOM*. IEEE, 2006.

[3] B. Cohen. Incentives build robustness in bittorrent. In *Workshop on Economics of Peer-to-Peer systems*, June 2003.

[4] N. Deo. *Graph Theory with Applications to Engineering and Computer Science*. Prentice-Hall, Englewood Cliffs, N.J., 1974.

[5] J. Feigenbaum and S. Shenker. Distributed algorithmic mechanism design: Recent results and future directions. In *6th Intl Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*, pages 1–13, 2002.

[6] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, 1991.

[7] D. Hughes, G. Coulson, and J. Walkerdine. Free riding on gnutella revisited: the bell tolls? *Distributed Systems Online*, 6(6), 2005.

[8] A. Kankanhalli, B. C. Tan, and K.-K. Wei. Contributing knowledge to electronic knowledge repositories: An empirical investigation. *MIS Quarterly*, 29(1):113–143, 2005.

[9] R. Krishnan, M. D. Smith, Z. Tang, and R. Telang. The virtual commons: Why free-riding can be tolerated in file sharing networks? In *International Conference on Information Systems*, November 2002.

---

[10]Such a node may upload a file to another node, but at some point a node in that chain will be a terminal leaf node.

[11]Or others; LGTS is sufficient to support the equilibrium, but may not be unique.

[10] R. D. Putnam. *Bowling Alone : The Collapse and Revival of American Community*. Simon & Schuster, New York, 2000.

[11] K. Ranganathan, M. Ripeanu, A. Sarin, and I. Foster. To share or not to share : An analysis of incentives to contribute in collaborative file-sharing environments. In *Workshop on Economics of Peer-to-Peer Systems*, June 2003.

[12] J. Shneidman and D. C. Parkes. Rationality and self-interest in peer to peer networks. In *2nd Int. Workshop on Peer-to-Peer Systems*, 2003.

[13] S. Stefan, P. K. Gummadi, and S. D. Gribble. A measurement study of peer-to-peer file sharing systems. In *Multimedia Computing and Networking 2002*, January 2002.

[14] B. Theodore, L. Blume, and H. R. Varian. On the private provision of public goods. *Journal of Public Economics*, 29(1):25–49, 1986.

APPENDIX

Proof of Lemma 1

$$
\begin{aligned}
u_i^S(n,p) &= \frac{1-(1-p)^{n-1}}{(n-1)p} \quad (9)\\
&= \frac{1-(1-p)^{n-1}}{(n-1)(1-(1-p))}\\
&= \frac{1+(1-p)+\cdots+(1-p)^{n-2}}{n-1}.
\end{aligned}
$$

Here we used the sum of a geometric series $\sum_{k=0}^{n} r^k = \frac{1-r^{n+1}}{1-r}$. Let $I_{n,p} = 1+(1-p)+\cdots+(1-p)^{n-2}$. Then $u_i^S(n,p)$ simplifies to $u_i^S(n,p) = I_{n,p}/(n-1)$. Similarly, let $I_{n,q} = 1+(1-q)+\cdots+(1-q)^{n-2}$, then $u_i^N(n,q)$ can be written as $u_i^N(n,q) = I_{n,q}/(n-1)$. Since $p = \frac{1}{k+1} < q = \frac{1}{k}$, $1-p > 1-q$ and $(1-p)^x > (1-q)^x$, $x = 0, \cdots, n-2$. Therefore $MBS(n,p,q) > 0$.

Proof of Lemma 2

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{MBS(n,p,q)}{u_i^N(n,q)}\\
&= \lim_{n\to\infty} \frac{\frac{1-(1-p)^n}{p}}{\frac{1-(1-q)^n}{q}} - 1 = \frac{1}{k}.
\end{aligned}
$$

Lemma 3 – explaining footnote 6

Equation (1) and (2) are an approximation of $u_i^S(n,k)$ and $u_i^N(n,k)$, which do not take into account the payoff difference between sharing peers and free-riders. The precise expressions of $u_i^S(n,k)$ and $u_i^N(n,k)$ are as follows,

$$
\begin{aligned}
&u_i^S(n,k)\\
&= \Sigma_{s=0}^{k-1}\Sigma_{l=0}^{n-k-1} C_{k-1}^s C_{n-k-1}^l (\frac{1}{k})^s(1-\frac{1}{k})^{k-s-1}\\
&\quad (\frac{1}{k+1})^l(1-\frac{1}{k+1})^{n-k-l-1}\frac{1}{s+l+1}, \quad (10)
\end{aligned}
$$

and

$$
\begin{aligned}
&u_i^N(n,k)\\
&= \Sigma_{s=0}^{k-1}\Sigma_{l=0}^{n-k-1} C_{k-1}^s C_{n-k-1}^l (\frac{1}{k})^l(1-\frac{1}{k})^{n-k-l-1}\\
&\quad (\frac{1}{k-1})^s(1-\frac{1}{k-1})^{k-s-1}\frac{1}{s+l+1}. \quad (11)
\end{aligned}
$$

The marginal benefit of sharing for node i follows as,

$$
MBS_i(n,k) = u_i^S(n,k) - u_i^N(n,k). \quad (12)
$$

And Lemma 1 becomes

LEMMA 3. $\forall\, n > 3$, and $\forall\, k > 2(k < n)$, $MBS_i(n,k) > 0$.

PROOF. We consider a hypothetical scenario, in which when node i shares her files, only some nodes know that she does. That is, those who do not know that i shares will continue as if i did not. Suppose X $(0 \le X \le k)^{12}$ out of the k sharing nodes and Y $(0 \le Y \le n-k-1)$ out of the $n-k-1$ free riding nodes know that i is sharing and the rest $k-X$ and $n-k-Y-1$ nodes do not know it. The payoff of node i is then a function of n, k, X, and Y,

$$
\begin{aligned}
&u_i(n,k,X,Y)\\
&= \Sigma_{s=0}^{k-X-1}\Sigma_{x=0}^{X}\Sigma_{l=0}^{n-k-Y-1}\Sigma_{y=0}^{Y}\\
&\quad C_{k-X-1}^s C_X^x C_{n-k-Y-1}^l C_Y^y\\
&\quad (\frac{1}{k-1})^s(1-\frac{1}{k-1})^{k-X-s-1}(\frac{1}{k})^x(1-\frac{1}{k})^{X-x}\\
&\quad (\frac{1}{k})^l(1-\frac{1}{k})^{n-k-Y-l-1}(\frac{1}{k+1})^y(1-\frac{1}{k+1})^{Y-y}\\
&\quad \frac{1}{s+l+x+y+1}. \quad (13)
\end{aligned}
$$

In this scenario, u_i^N could be re-interpreted as if i shared but no other nodes knew that she did. Similarly, u_i^S could be re-interpreted as if i shared and all other nodes knew that she did. That is,

$$
u_i^S(n,k) = u_i(n,k,k,n-k-1); \quad (14)
$$

$$
u_i^N(n,k) = u_i(n,k,0,0). \quad (15)
$$

If we could show that

$$
u_i(n,k,X+1,Y) > u_i(n,k,X,Y), \quad (16)
$$

and

$$
u_i(n,k,X,Y+1) > u_i(n,k,X,Y), \quad (17)
$$

we would have effectively shown that

$$
u_i(n,k,k,n-k-1) > u_i(n,k,0,0). \quad (18)
$$

According to equation (14) and (15), we would also have shown that

$$
u_i^S(n,k) > u_i^N(n,k).
$$

Now, we show the proofs for inequality (16) and (17). For simplicity in notation, we use Ω to denote the following part of equation (13),

$$
\begin{aligned}
\Omega &= \Sigma_{x=0}^{X}\Sigma_{l=0}^{n-k-Y-1}\Sigma_{y=0}^{Y} C_X^x C_{n-k-Y-1}^l C_Y^y\\
&\quad (\frac{1}{k})^x(1-\frac{1}{k})^{X-x}(\frac{1}{k})^l(1-\frac{1}{k})^{n-k-Y-l-1}\\
&\quad (\frac{1}{k+1})^y(1-\frac{1}{k+1})^{Y-y}.
\end{aligned}
$$

---

[12] In equation (13), the maximum of X is $k-1$ instead of k. This is because the node chosen by i for downloading is not counted, since its downloading choice does not affect i's payoff.

and re-write equation (13) as

$$u_i(n, k, X, Y)$$
$$= \Sigma_{s=0}^{k-X-1} C_{k-X-1}^s (\frac{1}{k-1})^s (1 - \frac{1}{k-1})^{k-X-s-1}$$
$$\Omega \frac{1}{s+l+x+y+1}$$
$$= \Sigma_{s=0}^0 C_{k-X-1}^s (\frac{1}{k-1})^s (1 - \frac{1}{k-1})^{k-X-s-1}$$
$$\Omega \frac{1}{s+l+x+y+1}$$
$$+ \Sigma_{s=1}^{k-X-2} C_{k-X-2}^s (\frac{1}{k-1})^s (1 - \frac{1}{k-1})^{k-X-s-1}$$
$$\Omega \frac{1}{s+l+x+y+1}$$
$$+ \Sigma_{s=1}^{k-X-2} C_{k-X-2}^{s-1} (\frac{1}{k-1})^s (1 - \frac{1}{k-1})^{k-X-s-1}$$
$$\Omega \frac{1}{s+l+x+y+1}$$
$$+ \Sigma_{s=k-X-1}^{k-X-1} C_{k-X-1}^s (\frac{1}{k-1})^s (1 - \frac{1}{k-1})^{k-X-s-1}$$
$$\Omega \frac{1}{s+l+x+y+1}. \quad (19)$$

The last step uses the fact that $C_{k-X-1}^s = C_{k-X-2}^s + C_{k-X-2}^{s-1}$. Combining the first two terms and the last two terms in the previous equation, we have

$$u_i(n, k, X, Y)$$
$$= \Sigma_{s=0}^{k-X-2} C_{k-X-2}^s (\frac{1}{k-1})^s (1 - \frac{1}{k-1})^{k-X-s-1}$$
$$\Omega \frac{1}{s+l+x+y+1}$$
$$+ \Sigma_{s=1}^{k-X-1} C_{k-X-2}^{s-1} (\frac{1}{k-1})^s (1 - \frac{1}{k-1})^{k-X-s-1}$$
$$\Omega \frac{1}{s+l+x+y+1}$$
$$= (1 - \frac{1}{k-1}) \Sigma_{s=0}^{k-X-2} C_{k-X-2}^s (\frac{1}{k-1})^s$$
$$(1 - \frac{1}{k-1})^{k-X-s-2} \Omega \frac{1}{s+l+x+y+1}$$
$$+ (\frac{1}{k-1}) \Sigma_{s=0}^{k-X-2} C_{k-X-2}^s (\frac{1}{k-1})^s$$
$$(1 - \frac{1}{k-1})^{k-X-s-2} \Omega \frac{1}{s+l+x+y+2}. \quad (20)$$

The last step involves substituting t for $s-1$ and then s for t. Using the same method, we obtain $u_i(n, k, X+1, Y)$,

$$u_i(n, k, X+1, Y)$$
$$= (1 - \frac{1}{k}) \Sigma_{s=0}^{k-X-2} C_{k-X-2}^s (\frac{1}{k-1})^s$$
$$(1 - \frac{1}{k-1})^{k-X-s-2} \Omega \frac{1}{s+l+x+y+1}$$
$$+ \frac{1}{k} \Sigma_{s=0}^{k-X-2} C_{k-X-2}^s (\frac{1}{k-1})^s$$
$$(1 - \frac{1}{k-1})^{k-X-s-2} \Omega \frac{1}{s+l+x+y+2}. \quad (21)$$

Finally,

$$u_i(n, k, X+1, Y) - u_i(n, k, X, Y)$$
$$= (\frac{1}{k-1} - \frac{1}{k}) \Sigma_{s=0}^{k-X-2} C_{k-X-2}^s (\frac{1}{k-1})^s$$
$$(1 - \frac{1}{k-1})^{k-X-s-2} \Omega \frac{1}{s+l+x+y+1}$$
$$- (\frac{1}{k-1} - \frac{1}{k}) \Sigma_{s=0}^{k-X-2} C_{k-X-2}^s (\frac{1}{k-1})^s$$
$$(1 - \frac{1}{k-1})^{k-X-s-2} \Omega \frac{1}{s+l+x+y+2}. \quad (22)$$

Notice that $\frac{1}{k-1} - \frac{1}{k} > 0$ and each term in the first term is weighed more than the second term, since $\frac{1}{s+l+x+y+1} > \frac{1}{s+l+x+y+2}$. We conclude that

$$u_i(n, k, X+1, Y) - u_i(n, k, X, Y) > 0. \quad (23)$$

Applying the same logic to $u_i(n, k, X, Y+1) - u_i(n, k, X, Y)$, we can also show

$$u_i(n, k, X, Y+1) - u_i(n, k, X, Y) > 0. \quad (24)$$

□

Proof of Proposition 2 (sketch)

Suppose all the peers adopt *individual grim-trigger strategy* (IGTS). Whenever a pair of peers each demand one unit of content from each other, IGTS requires they start by sharing with each other and stop sharing forever if one has deviated in the previous period. Suppose further that a node has diffuse (uniform) priors over the distribution of possible flow topologies (each possibility is equally likely). These beliefs will be sustained in equilibrium because there are no moves by nature and the problem is stationary so there are no changes in flows that are informative about the unobserved links. The proof follows 5 steps:

Step.1 Show that node B in Figure 6.(a) will choose to stop sharing to C. This can be done by examining Figure 6.(c), (d), (e), and (f), which represent all of B's possible beliefs. Given that each peer is individually rational, in (c), (e), and (f) B can gain from cutting off C. In (d), B can gain from cutting off either A or C. Thus it is profitable for B to stop sharing to C.

Step.2 Following the same logic as in Step. 1, show that node B in Figure 6.(b) will choose not to deviate and continue sharing to A.

Step.3 To show that Figure 5 can be sustained in a weak perfect Bayesian equilibrium, we only need to show that node B in both Figure 7.(a) and (b) will not deviate, since these two cases represent scenarios for all nodes in Figure 5.

Step.4 In Figure 7.(a), B only knows the links that she participates in, and the total number of nodes in the graph. Table 1[13] lists all the 64 possible flow graphs that B may believe to be true, out of which only 28 satisfy individual rationality according to the results of Step.1 and 2. These 28 graphs are marked with "Yes" in the "Possible?" column. The rest 36 graphs are eliminated as impossible due to the reasons listed in the "Reason" column. For example, in No. 5, node C is sharing to both B and A but not receiving any files from any nodes. Node C would benefit from cutting off both A and B. Thus the graph in No. 5 is not possible and hence B should eliminate it as a possible belief. After eliminating all impossible beliefs, 28 equally likely flow graphs are left. These 28 beliefs are further analyzed in Table 2, each row of which lists

---

[13]Table 1-4 are put online at http://www-personal.si.umich.edu/~ljian/papers/icec08-supp-appendix.pdf.

a particular belief, whether B can profit by deviating from its current equilibrium actions, and the resulted flow graph if it deviates. This analysis shows that in 9 cases, B profits by cutting off D; in 9 other cases B profits by cutting off A; in 8 cases B profits by cutting off either A or D; and in the rest 2 cases, B could cut off both A and D and free-ride. Thus if B cuts off D, with probability $\frac{2+8+9}{28} = \frac{19}{28}$ this is a profitable deviation and it receives $(v-c)$ from period t onwards. And with probability $\frac{9}{28}$ node C will eventually stop sharing to B, and B would drop out and receive a payoff of zero. With a discount factor of δ, node B's continuation payoff when cutting off D in period t, U_t^D, is calculated as follows,

$$U^D = \frac{19(v-c)}{28} + \frac{9(1-\delta)}{28} \quad (25)$$
$$(\underbrace{v-c}_{period\ t} + \underbrace{\delta(v-c)}_{period\ t+1} + \underbrace{\delta^2(-c)}_{period\ t+2})$$

Equation 25 shows B's payoffs in period t, $t+1$, and $t+2$ if she cuts off D when she should have cut off A. From period $t+3$ onwards node B drops out and receives zero payoff. Collecting terms, equation 25 simplifies to

$$U_t^D = v - c - \frac{9\delta^2 v}{28} + \frac{9\delta^3 c}{28} \quad (26)$$

By symmetry, $U_t^A = U_t^D$. If B decides to cut off both A and D, her payoff U_t^{AD} is

$$U_t^{AD} = \frac{2v}{28} + \frac{26(v + \delta v)(1-\delta)}{28} \quad (27)$$

On the other hand, B receives $U_t^e = v - 2c$ if she maintains the current equilibrium strategy. Therefore, the conditions for node B not to deviate are:

$$U_t^D < U_t^e, \quad (28)$$

and

$$U_t^{AD} < U_t^e, \quad (29)$$

which simplify to

$$c < \frac{9\delta^2 v}{28 + 9\delta^3}, \quad (30)$$

and

$$c < \frac{13\delta^2 v}{28}. \quad (31)$$

Combining inequalities 30 and 31 we conclude that if $c < \frac{9\delta^2 v}{28 + 9\delta^3}$, node B will choose not to deviate from the equilibrium strategy.

Step.5 Similar logic applies to Figure 7.(b), in which if $c < \frac{6\delta^2 v}{7}$ node B will choose not to deviate from the equilibrium strategy. Similar to Table 1 and 2, Table 3 and 4 list node B's possible beliefs and strategies.
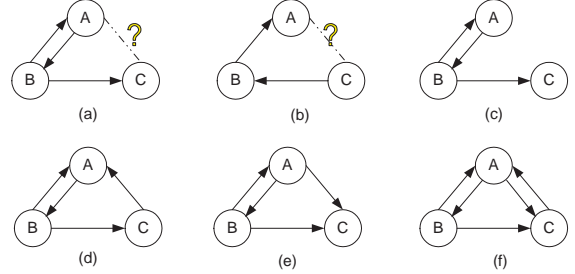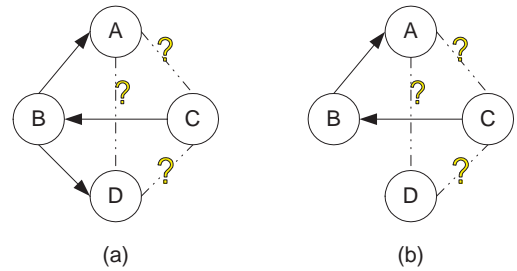

Figure 6: Three Node Scenarios

Figure 7: Node B's Beliefs