Information Systems Research
Vol. 15, No. 2, June 2004, pp. 155–174
ISSN 1047-7047 | EISSN 1526-5536 | 04 | 1502 | 0155
informs®
DOI 10.1287/isre.1040.0020

An Empirical Analysis of Network Externalities in Peer-to-Peer Music-Sharing Networks

Atip Asvanund, Karen Clay, Ramayya Krishnan, Michael D. Smith
H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213
{atip@andrew.cmu.edu, kclay@andrew.cmu.edu, rk2x@andrew.cmu.edu, mds@andrew.cmu.edu}

Peer-to-peer (P2P) file sharing networks are an important medium for the distribution of information goods. However, there is little empirical research into the optimal design of these networks under real-world conditions. Early speculation about the behavior of P2P networks has focused on the role that positive network externalities play in improving performance as the network grows. However, negative network externalities also arise in P2P networks because of the consumption of scarce network resources or an increased propensity of users to free ride in larger networks, and the impact of these negative network externalities—while potentially important—has received far less attention.

Our research addresses this gap in understanding by measuring the impact of both positive and negative network externalities on the optimal size of P2P networks. Our research uses a unique dataset collected from the six most popular OpenNap P2P networks between December 19, 2000, and April 22, 2001. We find that users contribute additional value to the network at a decreasing rate and impose costs on the network at an increasing rate, while the network increases in size. Our results also suggest that users are less likely to contribute resources to the network as the network size increases. Together, these results suggest that the optimal size of these centralized P2P networks is bounded—At some point the costs that a marginal user imposes on the network will exceed the value they provide to the network. This finding is in contrast to early predictions that larger P2P networks would always provide more value to users than smaller networks. Finally, these results also highlight the importance of considering user incentives—an important determinant of resource sharing in P2P networks—in network design.

Key words: peer-to-peer networks; Napster; network externalities; empirical; incentives; size limitation; efficiency; network design
History: Cynthia Beath, Senior Editor. This paper was received on October 15, 2002, and was with the authors 9 months for 3 revisions.

1. Introduction

At their core, peer-to-peer (P2P) network architectures enable resource sharing directly between autonomous individual network users, also known as peers. Currently, these resources are most commonly files containing digitized information content such as music, movies, pictures, software, or text. However these resources can also include storage capacity, or bandwidth or computing power. A defining characteristic of these networks is that resource availability and consumption patterns on the network are determined by individual user (peer) behavior. Thus, P2P networking is different than the traditional client-server architecture where all network resources are contained in and managed by a central server (Parameswaran et al. 2001).

P2P networking has its origins in the early design of many Internet architectures and protocols (Minar and Hedlund 2001), but its recent popularity began with the launch of the Napster network in May 1999. Napster enabled users worldwide to share music files compressed in MP3 format. By many accounts it was the fastest growing application in the Internet's history, expanding from 30 users to 25 million users in its first 12 months of operation (Strahilevitz 2002). Numerous P2P file sharing systems have followed Napster, including OpenNap, Scour, iMesh, Gnutella, eDonkey, FreeNet, BitTorrent, and DirectConnect. As of June 2004 the most popular such network is Kazaa, which according to Download.com has been downloaded over 350 million times since its introduction in July 2000.

Building on the success of P2P file sharing networks, entrepreneurs and programmers have recently developed P2P-based networks in other application domains. Notable examples include streaming media distribution (e.g., Allcast, Blue Falcon Networks, Kontiki, Uprizer), distributed computing (e.g., SETI@Home), remote collaboration (e.g., Groove Networks), enterprise information sharing (e.g., Bad Blue, Nextpage), spam filtering (e.g., Cloudmark), copyright-friendly content distribution (e.g., Altnet), and decentralized data storage and archiving (e.g., Publius, FreeHaven).

Despite P2P's potential as an efficient tool for digital content distribution and distributed resource sharing, there has been little academic work analyzing the impact of user behavior on its design and real-world operation. Systematic research to address these questions is important for a variety of constituencies, including engineers determining the right configuration and provisioning of P2P networking equipment, protocol designers and network planners designing user incentives to optimize network performance, entrepreneurs developing user adoption forecasts to support P2P business models, and intellectual property holders seeking to develop systems to minimize the incidence of copyright violations on P2P networks.

In this paper we study one component of P2P network operation: the interplay between positive and negative network externalities in a real-world environment. A network externality is the marginal effect that an additional user of a network has on existing users, where the impact of this marginal effect is not fully internalized by the additional user.

Most of the discussion around network externalities in P2P environments has focused on positive network externalities—the marginal value marginal users provide to the network. Positive network externalities arise when users who choose to share their content bring new content, replicas of existing content, or other shared resources to the network. Viewed in isolation, these positive externalities mean that larger networks will provide more value to users than smaller networks (e.g., Strahilevitz 2002). The importance of positive externalities on the scalability of P2P networks has been widely discussed (e.g., Saloner and Spence 2002, p. 54). For example, Hibbard (2001) opines, "In conventional content delivery, every PC that requests a file bogs down a server's performance. In peer-to-peer delivery, every PC that joins the network improves download speeds by adding another available cache." Hibbard further quotes Ian Clarke, Founder and Chief Technology Officer of Uprizer as saying, "the bigger the network gets, the more efficiently it is able to deliver content" (p. 62).

However, positive network externalities should not be considered in isolation. P2P users can also impose negative externalities on other members of the network by consuming scarce network resources or by an increased propensity in larger networks for users to consume network resources without providing resources back to the network in return. The impact of these sources of negative externalities—while potentially important to network scalability—has received far less discussion than the impact of positive negative network externalities.

This study seeks to extend the understanding of P2P network scalability by measuring how both positive and negative externalities vary as a function of network size. We did this by gathering a unique dataset from the six most popular OpenNap networks from December 19, 2000, to April 22, 2001.[1] Our data include information on network congestion, and song availability and replication (number of copies of the song available for sharing on the network) for 170 randomly selected songs in 17 musical genres. These data are useful because they allow comparison across a set of networks with identical design but widely varying size.

We find that the marginal benefit an additional user provides to the network decreases with network size, while the marginal cost they impose on the network in terms of congestion on shared resources increases with network size. This suggests that the optimal size of this type of P2P network should be bounded in many common settings—At some point the marginal cost an additional user imposes on other network users will be larger than the marginal value they provide. To explore this relationship further, we apply the Erlang-B model to assess the impact of increasing network capacity on network congestion. This model

[1] OpenNap networks use the Napster protocol, which employs a centralized searchable catalog of content to allow peers to locate content on other peers' computers.

suggests that although increasing capacity may allow more users to participate on a network, at some point there may be little incentive for network operators to provision this capacity because for sufficiently large networks diminishing positive network externalities implies decreasing the benefits achieved by adding more capacity.

These findings contribute to the literature in three primary ways. First, we use a high-quality and unique panel data set to directly measure positive and negative network externalities. As observed by Varian, while network externalities are commonly discussed in the literature, "for most network goods, the frequency of data collection is too low to capture the interesting dynamics" (Varian 2003, p. 33). Second, while positive network externalities in networked environments have been commonly discussed, ours is one of the first papers to measure the role of negative network externalities in limiting network scalability. Third, our analysis has implications for the operation and design of P2P networks, an emerging and important architecture for distributing information goods and sharing other computing resources.

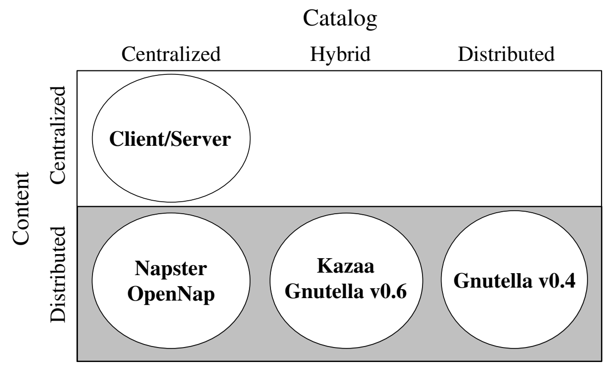
The remainder of this paper proceeds as follows. Section 2 provides background on P2P networks and reviews the relevant IS, computer science, economics, and social psychology/groups literatures as they relate to our study. Section 3 presents a model of positive and negative externalities in P2P networks, and develops sufficient conditions for the optimal size of P2P networks to be bounded. Section 4 discusses our methodology and data. Section 5 presents our empirical results and discusses the limitations and potential generalizability of our study. Section 6 concludes the paper and identifies areas for future research.

2. Background and Literature

2.1. Architecture

Network architectures can be summarized along two axes: the degree of decentralization of the network content and the degree of decentralization of the catalog of this content (Figure 1). The degree of decentralization of network *content* reflects whether the content is stored in a central location (improving direct management of the content), or is stored in a distributed manner separately by the individual nodes/peers (caching content within the network, eliminating a single point of failure for content distribution, and offloading bandwidth burden to the edge of the network). The degree of decentralization of the *catalog* of content reflects whether this catalog is stored in a central location (increasing the accuracy and reliability of the catalog), or is stored in a distributed manner separately by the individual nodes/peers (improving flexibility and eliminating a single point of failure for directory services).

Figure 1 Taxonomy of Content Distribution Architectures



By definition, P2P networks reside in the shaded region of Figure 1 corresponding to distributed content. As noted in the figure, P2P networks can be categorized into three general types depending on the degree of centralization in the catalog of content.[2] At one end of the spectrum, Napster and OpenNap networks have a single central catalog of content for the entire network. At the other end, in Gnutella Version 0.4, each node catalogs its own content; thus, the catalog is completely distributed within the network. The Kazaa and Gnutella Version 0.6 architectures fall in between—sharing design elements from both the centralized and distributed architectures. We describe each architecture category in more detail below.

Centralized P2P Architectures. The two most popular and well-known centralized P2P architectures are Napster and OpenNap. Shortly after Napster's

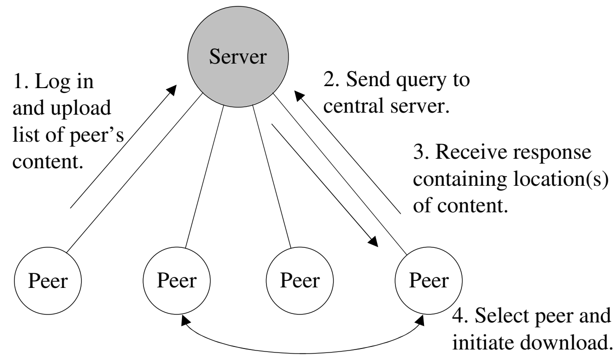
[2] Some emerging P2P network architectures could reasonably be cataloged as having a "hybrid" content distribution where content is served from both a central server and from the edges of the network. Centerspan's C-Star One content delivery network and Blue Falcon's stream media distribution systems are examples. For simplicity, these architectures are left out of our figure in favor of the more common and established network architectures.

introduction in May 1999, a group of programmers began work to reverse engineer the Napster protocol and create an open source implementation of the server. The first version of this open source Napster server, dubbed OpenNap, was made available in January 2000.[3] Individuals could use the server software to create their own "Napster-like" networks that could be accessed by a variety of clients conforming to the Napster protocol. Many individuals did so, creating hundreds of networks competing alongside Napster. These networks were distinct and separate from each other and from Napster's closed network.[4]

As noted above, OpenNap networks operated with a particular hierarchy. During our data collection period, each network contained one designated server that maintained a centralized catalog of all the content on the network. To gain access to the network, an OpenNap user operating a peer would choose a server from a directory of available OpenNap servers (e.g., those listed on Napigator.com). The peer maintains a connection to this central server for the duration of the user's presence on the network. Because these servers have limited capacity for simultaneous connections with peers, this creates a potential source of congestion.

Functionally, the protocol for both the Napster and OpenNap networks is nearly the same. After a user logged into the network, their peer would perform the following steps (Figure 2). First, it would upload a list of the names, sizes, and encoding speeds of the files it is sharing (if any), along with its IP-address and the (self-reported) speed of its connection. Any subsequent changes to its shared files are immediately uploaded, keeping the central catalog current. Second, to locate a file, the user places a keyword query against this catalog database. Third, the database returns a list of any matching results. This list includes the name, length, encoding speed, and provider for each file. The client program issues

Figure 2 **Napster/OpenNap Network Operation**



a ping request to each provider and sorts the list in ascending order by the amount of time it took to receive a pong (response) message from the provider, using this time as a proxy for the congestion at the peer. Fourth, the user chooses one of the peers from the list and initiates a download. This download request may be accepted or queued by the peer computer providing the content. Peers accept simultaneous downloads up to a user-specified limit and queue any additional download requests. Once the download request is accepted, the requesting peer computer downloads the content directly from the providing peer.

Two additional points are important to our analysis. First, under the default sharing settings, the requesting peer becomes a provider of all content they download (in addition to any content they initially brought to the network). In this manner, content is auto-replicated on the network in proportion to its popularity. More popular content will be downloaded by—and therefore available from—more peers than less popular content. Second, although Napster and OpenNap clients share files in their download directory by default, users can turn off this default setting. Users who turn off sharing consume network resources without providing resources in return. They hamper the auto-replication characteristic of these networks. We refer to these peers as free-riders in the remainder of the paper.

Decentralized P2P Architectures. Unlike centralized networks, decentralized P2P networks have no hierarchy. In Gnutella 0.4, the most popular decentralized architecture, peers are connected in a "web," with each peer connected to approximately three

[3] See http://opennap.sourceforge.net/.

[4] During our study period, a user searching one OpenNap network would not see users logged into the other networks. Most importantly, actions taken by Napster, in an effort to remove copyrighted materials late in our study, did not impact the availability of content on the OpenNap networks we studied.

other peers. Because there is no central server, peers maintain separate catalogs of their own content. To locate content in the network, peers pass a query to each of the peers to which they are connected. In turn, these peers pass the query to the peers to which they are connected (eliminating any peers who have already received the request). Each peer who receives the query checks to see if they have the desired content, and if so, returns a reply to the initiating peer along the original query's path.

In principle, this protocol could allow queries to reach every node in the network. However, the Gnutella protocol limits the depth that queries can propagate through the network by including a "time to live" (TTL) parameter in each query message. This TTL parameter takes on a maximum value of seven. Each peer who receives a query message decrements the TTL value and only forwards queries when the TTL is greater than zero. The TTL effectively limits the size of the network each node can reach to about 10,000 nodes.[5]

Hybrid P2P Architectures. Hybrid P2P architectures such as Gnutella 0.6 and Kazaa contain design elements from both centralized and decentralized architectures. As in centralized architectures, peers (a.k.a. leaf nodes) connect to "local" centralized servers (a.k.a. ultrapeers or supernodes). The connection between a leaf node and an ultrapeer is similar to the connection between peers and centralized servers in centralized P2P networks: Leaf nodes upload a list of the content they are sharing, ultrapeers maintain a catalog of content for all their leaf nodes, and queries from the leaf nodes are sent to the ultrapeer. However, unlike centralized P2P networks such as OpenNap, ultrapeers are connected to each other in a structure comparable to the decentralized networks. If an ultrapeer cannot adequately satisfy a query issued by one of its leaf nodes, it can forward this query to the ultrapeers it maintains connections to and they in turn can forward the query to their interconnected ultrapeers. While this design innovation increases the scalability of hybrid networks versus decentralized networks, the forwarding of queries among ultrapeers is still limited by a TTL parameter. Thus, as with decentralized networks, the number of leaf nodes that can be reached from any particular leaf node is explicitly limited.

[5] See the Gnutella 0.4 protocol at http://rfc-gnutella.sourceforge.net/ for more details.

2.2. Literature

The literature on network externalities has focused on two types of externalities—direct and indirect. The classic example of a direct network externality is a telephone network, where the utility of the network to the individual increases with the number of users that the individual can talk to. An indirect network externality arises when the utility of a product increases with the number of users because, for instance, the quality of the product is higher or there are more complementary products available (Katz and Shapiro 1986, 1994; Farrell and Saloner 1987).

While direct network externalities have been widely discussed in the context of telecommunications, they have been difficult to empirically measure in this setting because of the limitations of available data, which are typically either time-series or cross-sectional (Varian 2003). If the data are time series, then it can be difficult to disentangle network effects from other unrelated changes (e.g., falling equipment prices). If the data are cross-sectional, then it can be difficult to determine whether regional differences are attributable to peer group effects (differences in preferences) or network externalities.

In the context of indirect network externalities, most empirical studies have confirmed theoretical expectations with regard to their impact on consumer demand, product adoption, and switching costs. For example, several papers have found that positive network effects increase consumer willingness to pay and, as a result, increase market price. This has been demonstrated in the context of spreadsheet software (Gandal 1994, Brynjolfsson and Kemerer 1996) and video cassette recorders (Park 2003). Positive network effects have also been shown to influence switching costs and platform adoption of telecommunications equipment (Economides and Himmelberg 1995, Augereau et al. 2003, Forman and Chen 2003), banking networks (Saloner and Shepard 1995, Kauffman et al. 2000), and consumer electronics equipment (Dranove and Gandal 2003, Goolsbee and Klenow 2003, Gandal et al. 2000).

Our work differs from prior empirical work on indirect network externalities in three primary ways.

First, we empirically measure *negative* network externalities in addition to positive network externalities. To the best of our knowledge, we are the first to empirically analyze the impact of both positive and negative network externalities. Second, we are able to collect panel data where the cross-sectional observations are made across networks using exactly the same server software and protocol design. Thus, the nature of these data solves some of the data problems mentioned above for studies using only time-series or cross-sectional data alone. Third, we analyze network externalities in the context of P2P networks, which are a new and important architecture for distributing information goods and are differentiated from many of the existing contexts in which indirect network effects arise. Specifically, in P2P networks, participants take on dual roles of consumers and providers of resources, the consumption and provision of these scarce resources is heterogeneous across peers, and resources (e.g., CPU cycles, bandwidth, content) are served from the edges of the network. This is in contrast to a more traditional telecommunications environment where the scarce resources (bandwidth, switching capacity) are provided centrally and separately from consumption decisions, and where consumption of the scarce resource is more homogeneous across users.

Our work is related to three additional types of literature. The first type concerns the motivation of participants in online forums. This analyzes the benefits online groups can provide to each other in the form of ties to a community, social support, and access to community resources (e.g., Kraut and Attewell 1993, Constant et al. 1996). More recently, several papers have analyzed personal motivations for online group participation, concluding that motivations appear to be driven primarily by altruism and reciprocity (Wasko and Faraj 2000, Gu and Jarvenpaa 2003, Subramani and Peddibhotla 2002).

In the online forum literature, our paper is most closely relates to Butler (2001) who develops a resource-based model of social interaction in online communities. He applies this model to data from Listserv™ communities on the Internet, finding that larger networks have both advantages and disadvantages in attracting and retaining members. Advantages derive from increased potential interactions with other participants. Disadvantages derive from fewer opportunities to participate, reduced opportunities to form personal relationships, and lower levels of contribution.

Our work differs from this literature in three ways. First, unlike most online forums mentioned above, user identity in P2P networks is obscured from other users, thus eliminating explicit reciprocity as a primary motivating factor for content provision. Second, our data allow us to empirically compare the operation of similar networks with different numbers of users. Third, our data allow us to extend Butler's (2001) work to analyze in more detail how positive and negative externalities impact optimal group size.

The second type of literature relevant to our study is the economics literature pertaining to public and club goods (Samuelson 1954, Buchanan 1965). The services provided over P2P networks have some of the characteristics of public goods (Krishnan et al. 2003), so in general we would expect to see underprovision of content. The underprovision of public goods is likely to be exacerbated by the voluntary nature of contributions to P2P networks. In a typical public or club goods setting, individuals are compelled to contribute to support the public good through taxation and the club through payment of dues if they wish to join. In contrast, contributions in P2P networks have been voluntary contributions of content or resources by members.

The underprovision may be mitigated to some degree by another feature of the P2P environment—Consumers become contributors by default. In the public and club goods environments, indirect network effects arise from cost sharing of a fixed resource base with a greater number of people. In contrast, in P2P environments, additional consumers expand the resource base by bringing more content. Further, if free-riding is not an issue, the content available per person does not fall. This unique feature of P2P networks means that in an ideal case the heaviest consumers of network resources can also be among the most valuable contributors.

Given concerns about the underprovision of content caused by free-riding in P2P networks both in theory and in practice (e.g., Adar and Huberman 2000), a type of literature in computer science has arisen to address this problem. The focus has been on providing participants with explicit incentives to provide content and other resources. Researchers have demonstrated that network pricing (Cole et al. 2003), micropayment systems (Golle et al. 2001), reputation systems (Lai et al. 2003), autonomous club formation (Asvanund et al. 2003), and admission control systems (Kung and Wu 2003) are all feasible solutions. The drawback of all of these systems is that they may add considerable protocol overhead to the operation of P2P systems.

The third type of literature is closely related to the second, but focuses on addressing current problems of free-riding and associated congestion at the protocol level. Researchers have shown that network performance can be enhanced through improved indexing schemes (Stoica et al. 2001), the use of ultrapeers to reduce traffic load on low bandwidth peers (Kirk 2003), caching to improve the efficiency of content retrieval (Bhattacharjee et al. 2003), and intelligent linkage promotion based on similarity of interests (Sripanidkulchai et al. 2002). As with the explicit incentives discussed in the previous paragraph, each of these imposes additional overhead and so has its own cost as well.

Our work differs from these two literatures in its focus on measuring the extent of free-riding and associated congestion. In contrast, the second and third literatures take free-riding as a given and ask for a given level of free-riding whether or not changes in the protocol or the incentive structure can diminish or completely eliminate the problem. Without a more complete understanding of the empirical problem, however, the incentive or protocol approaches, if implemented, may not have the desired outcome.

3. Empirical Hypotheses

The central hypothesis that we explore in this paper is that positive and negative network externalities cause the optimal size of OpenNap networks to be bounded. In this section, we use an analytic model to derive sufficient conditions for this hypothesis to hold.

First, let N be the number of users in the network. Each user provides value to other network members by providing access to new songs or additional copies of songs already on the network, which ceteris paribus will increase variety and reduce the expected download time. Consistent with the definition of network externalities, positive network externalities arise because the range of content available and the number of copies of each piece of content is positively correlated with the number of users on the network. Likewise, users impose costs on other network members by increasing congestion in the form of expected login, query, and download times. Negative externalities arise as congestion is correlated with the number of users.

Thus, assume an individual user's utility from using the network is given by the sum of the utility from the availability and replication of a vector of content that they are interested in (F), and the (dis)utility of a vector of congestion effects they face (C):

$$U(F(N), C(N)) = U_F(F(N)) + U_C(C(N)). \qquad (1)$$

Modeling utility as separable in consumption and congestion is a good fit for P2P networks. For these networks, the value of the content can be modeled as independent of congestion, i.e., a song downloaded in 5 seconds will sound the same as a song downloaded in 30 seconds. This formulation would not, however, be applicable in a setting where the value of the content was time sensitive (e.g., stock quotes) or could be degraded due to congestion (e.g., streaming media). Network externalities in such settings would make useful areas for future research.

Consistent with the definitions of content and congestion, let users be better off when more content variety or more replicas of content are provided by the network, and worse off when network congestion increases:

$$\partial U/\partial f > 0 \qquad (2)$$

$$\partial U/\partial c < 0. \qquad (3)$$

In these equations, f is an element of the content vector F, and c is an element of the congestion effects vector C.

Furthermore, consistent with our discussion above, content and congestion will (weakly) increase in the number of network users N. Each user will bring either no content (free-riding), new content, or additional replicas of existing content to the network:

$$\partial f/\partial N \geq 0. \qquad (4)$$

Likewise, with regard to congestion, new users will (weakly) increase congestion measures, which in our setting include login time, query time, and download times:[6]

$$\partial c/\partial N \geq 0. \quad (5)$$

Finally, assume that U is concave in f and c:

$$\partial^2 U/\partial f^2 \leq 0 \quad (6)$$

$$\partial^2 U/\partial c^2 \leq 0. \quad (7)$$

Concavity in f would hold if users have a diminishing marginal utility from more content and more replicas of content. Concavity in c would hold if users had increasing marginal disutility from congestion, consistent with the standard intuition that the marginal utility of lost time increases as more is lost, because higher opportunity cost activities are precluded as the margin is pushed inward.[7]

Under these assumptions, we can characterize how utility varies with network size (i.e., $\partial U/\partial N$, $\partial^2 U/\partial N^2$) as a way to understand the impact of network externalities on optimal network size. Optimal network size will be bounded if $\partial U/\partial N$ is positive for small N (a necessary condition for networks to form at all), and $\partial^2 U/\partial N^2$ is strictly negative. Under these conditions, for sufficiently large networks, the marginal value a user provides to the network ($\partial U/\partial N$) will be negative, and thus the network would be better off if that user were not allowed to join.

To show that these two conditions hold, we first note that under our utility function (1),

$$\partial U/\partial N = \partial U/\partial f \cdot \partial f/\partial N + \partial U/\partial c \cdot \partial c/\partial N \quad (8)$$

$$\partial^2 U/\partial N^2 = \partial^2 U/\partial f^2 \cdot (\partial f/\partial N)^2 + \partial U/\partial f \cdot \partial^2 f/\partial N^2 + \partial^2 U/\partial c^2 \cdot (\partial c/\partial N)^2 + \partial U/\partial c \cdot \partial^2 c/\partial N^2. \quad (9)$$

Next, observe that by (2) and (4) the first term of (8) is positive, and by (3) and (5) the second term is negative. Finally, observe that by (2), (3), (6), and (7) for $\partial^2 U/\partial N^2$ to be strictly negative it is sufficient to show that the following are true, where (10) corresponds to the statement of Hypothesis 1 and (11) corresponds to our statement of Hypothesis 2:[8]

$$\partial^2 f/\partial N^2 < 0 \quad (10)$$

$$\partial^2 c/\partial N^2 > 0. \quad (11)$$

Thus, to show that optimal network size is bounded, it is sufficient to show that the following two hypotheses are true:

Hypothesis 1. *For all measures of value, the marginal value an additional user brings to the network decreases in network size.*

Hypothesis 2. *For all measures of cost, the marginal cost an additional user imposes on the network increases in network size.*

To test Hypothesis 1, we measure the collective content on the network in terms of availability and replication. Availability measures the number of unique songs that are provided on the network. Replication measures the number of copies of each song that is available on the network. Replication is a particularly important measure of network behavior. As noted above, OpenNap clients were configured to share downloaded songs by default, auto-replicating the song for the network.

Auto-replication allows a P2P network to efficiently meet download demand from users because the replication of songs on the network will scale in proportion to the song's popularity. The value of replication is that it helps distribute the load on the providers if multiple users choose to download songs simultaneously. It is important for replication to scale consistently with network size for download performance to scale well. However, to the extent that users

[6] Our model could be extended to include congestion from fake files on P2P networks. In May 2002, various artists and recording companies started flooding P2P networks with "fake" files labeled as popular music (Avery 2002, Warner 2002). While this practice falls outside of our data collection period, it could be included in our model as a congestion effect. An increase in the number of fake files on the network would reduce each user's utility because they would have to initiate more downloads to find their intended content, thus Equation (3) should hold. To the extent that record companies (weakly) target larger networks for more copies of fake files, Equation (5) would hold and our model would retain the same interpretation.

[7] We thank an anonymous referee for this insight.

[8] In fact only one of the inequalities in (10) and (11) needs to hold strictly for $\partial^2 U/\partial N^2$ to be strictly negative.

change the default sharing setting of OpenNap clients to disable, file sharing network scalability will be reduced.

To test Hypothesis 2, we measure the cost of accessing content on the network in terms of login congestion, query congestion, download attempt congestion, and download speed congestion. These measures of the negative network externalities reflect the steps in user interaction with centralized P2P networks where the congestion or delays may take place (Figure 2). These variables are discussed in more detail below.

4. Data

To empirically test these hypotheses, we collected data from six OpenNap networks on network congestion characteristics and content availability for 170 songs. As noted earlier, these networks use an open source version of the protocol used by the Napster network. Apart from that, these networks were entirely separate from each other and from the Napster network.

The OpenNap networks used in the data collection were the most popular networks listed by Napigator.com at the beginning of our collection period. We selected six networks because below this rank the size of the listed networks dropped significantly. The 170 songs were selected at random from the full repertoire of all popular artists in 17 separate genres listed at Amazon.com. We used Amazon.com's listings after determining that it had one of the most comprehensive publicly available databases of music content available on the Internet. Our data were collected every 18 hours from December 19, 2000, to April 23, 2001, and include user count, server count, login congestion, query congestion, song availability, song replication, and broadband song replication (Table 1).

User count measures the number of users on the network at the point in time when our data collection agent logged into the network. Server count measures the number of mirrored central catalog servers used by the network. Each of these pieces of information are passed to clients at login.

Login congestion measures the difficulty of logging on to the network. We measure login congestion as the amount of time it took to successfully log in to the network. As noted in §2.1, the server has a fixed capacity for simultaneous connections with clients. Because of this, we expect login congestion to be low initially and to quickly rise as network size approaches server capacity.

Query congestion measures the amount of time it takes to receive a query response from the central server. When users perform search queries for a file, they place traffic demand on the centralized servers that perform database searches, potentially degrading network performance for other users. This may happen in two ways: Having more users may increase the size of the database containing users' file listings, and having more users may generate more simultaneous search queries that the centralized servers must process.

Table 1 Summary Statistics

Variable	Obs.	Mean	St. dev.	Min	Max
Login congestion, query congestion, and song availability					
User count	323	3,118.00	2,283.00	68.00	8,618
Server count	323	7.00	3.40	1.00	15
Song availability	83,640	0.54	0.50	0.00	1
Song availability (broadband connection)	83,640	0.45	0.50	0.00	1
Song replication (number of copies of a song)	83,640	11.00	30.00	0.00	555
Song replication (broadband connections)	83,640	6.00	21.00	0.00	460
Login congestion (login time, seconds)	323	3.00	8.00	0.00	71
Query congestion (query time, seconds)	323	10.00	17.00	13.00	90
Download attempts and speed					
User count	13	2,620.00	687.00	1,458.00	3,588
Download attempt congestion (number of attempts before download starts)	582	2.85	4.37	1.00	45
Download speed congestion (average download speed, kbps)	582	32.00	33.00	0.00	200

These query responses allowed us to measure song availability and replication. Song availability is a binary variable for whether any copies of a particular song are available on the network, and broadband song availability is a binary variable measuring whether any copies of a particular song are available from peers reporting that they have a broadband connection. Similarly, song replication is the number of different peers who have copies of a particular song, and broadband song replication is the number of different peers who have copies of a particular song available over a broadband connection.

These data were collected using an automated software agent written for this purpose. The agent implemented the OpenNap protocol and was specifically designed to mimic the actions of typical users. This agent was also designed to have a negligible impact on network performance by spreading out content queries over time and by only downloading a small portion of songs when determining download speeds.

We chose popular artists because song availability was very low for a random selection of songs from all artists. The main drawback of this approach is that some tracks may become less popular over our data collection. However, the content was selected from the full repertoire of the artist (not just their most recent album) and only a few tracks were recent releases.[9]

We also developed control variables to mark significant announcements made by Napster during our study period (see Table 2). In our regressions, Time Period I refers to the period from December 19, 2000, to January 29, 2001, when Napster announced that they were planning to start a subscription service sometime during the summer of 2001. Time Period II refers to the period from January 29, 2001, to March 2, 2001, when Napster started filtering copyrighted tracks from its service. Time Period III refers to the period from March 2, 2001, to April 23, 2001. It is important to note that the aforementioned

Table 2 Key Data Collection Dates

December 19, 2000	Initial data collection starts (login congestion, query congestion, song availability/replication)
January 29, 2001	Napster announces subscription service, planned to start in summer 2001 (End of Time Period I)
March 2, 2001	Napster starts filtering copyrighted tracks from its servers in response to court order (End of Time Period II)
March 28, 2001	Collection of download attempt and download speed congestion starts
April 23, 2001	End of data collection

announcements by Napster had no impact on the operation of the OpenNap networks in our study. However, these announcements did have a secondary impact in that many former Napster users joined OpenNap networks immediately following these announcements, and these variables allowed us to control for such changes.

We collected an additional dataset on download congestion and speed from March 28, 2001, to April 19, 2001 (Tables 1 and 2). This dataset includes information on the size of the network and two additional measures of the congestion a user would face when trying to download a song. The first measure, download attempt congestion, is the number of download attempts our agent had to make (starting with the listing with the lowest ping time) before finding a peer who did not queue the download request. As noted in §2.1, P2P peers can define a maximum number of simultaneous downloads they are willing to serve. Requests above this value are then queued for subsequent processing.[10] The second measure is the download speed (in kbps) our agent observed when downloading the song. Download speed will vary, in part, based on the speed of the peer's connection to the network and the number of simultaneous downloads it is serving.

5. Empirical Analysis of the Network Externalities

5.1. Positive Network Externalities

In this section, we empirically investigate how both availability and replication vary with the number of users on a network. Our regression results for

[9] For all genres except emerging artists the list of best-selling artists did not change over the data collection period. We further checked the sensitivity of our results to changes in popularity by referencing the 36 most popular album charts tracked by Billboard at the beginning and end of our sample period. We found five songs that were contained in albums that moved off the Billboard album charts during our sample period. Eliminating these songs from our analysis did not change any of our results.

[10] Note that peers still respond to ping messages even after reaching the maximum number of simultaneous downloads.

Table 3 Regression Results for Positive Network Externalities

	1	2	3	4	5	6
Regression model	Logit-linear	Logit-log	Logit-polynomial	OLS-linear	OLS-natural log	OLS-polynomial
Dependent variable		Availability			Replication	
ln(user_count)		0.467 (8.05e-03)			4.63 (0.0735)	
user_count	1.82e-04 (4.05e-06)			1.74e-03 (3.27e-05)		
user_count′			4.72e-05 (1.27-e05)			0.003 (4.17e-05)
user_count′^{2}			−1.51e-07 (6.47e-09)			−2.77E-07 (1.64e-08)
user_count′^{3}			9.03e-12 (5.04e-13)			*−1.71e-12* (*1.27e-12*)
broadband	−0.481 (0.0117)	−0.486 (0.0118)	−0.485 (0.00118)	−4.74 (0.111)	−4.73 (0.111)	−4.74 (0.11)
Time [2]	Yes	Yes	Yes	Yes	Yes	Yes
Genre [16]	Yes	Yes	Yes	Yes	Yes	Yes
Network [5]	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	166,770	166,770	166,770	166,770	166,770	166,770
(Pseudo) R^2	0.247	0.253	0.252	0.186	0.191	0.199

Notes. Standard errors are in parentheses. Values in brackets denote the number of fixed effect variables. The ′ character denotes that the variable is centered. Italicized coefficients are insignificant ($P = 0.05$).

availability are presented in the first three columns of Table 3. Availability is measured as a binary value, where zero indicates that the song is not available and one indicates that that song is available. With binary data, ordinary least squares will not, in general, produce estimates confined to the 0-1 interval, making the results unreliable and difficult to interpret. Either the probit model, based on the normal distribution, or the logit model, based on the logistic distribution, is typically used in such circumstances. Given that the mean of the dependent variable is near 0.5, the estimates from the two approaches yield similar results. We present estimates of these equations using the logit model.

Our specifications for availability are given below for song i on network j at time t. These specifications represent three common functional forms: linear, logarithmic, and polynomial. The three functional forms allow us to compare model fit for different relationships between the number of users and availability.

Logit-linear: $Availability_{ijt} = user\_count_{jt} + Dbroadband_{ijt} + Dtime\_II_t + Dtime\_III_t + Dgenre_i + Dnetwork_j + \varepsilon_{ijt}$.

Logit-log: $Availability_{ijt} = \ln(user\_count)_{jt} + Dbroadband_{ijt} + Dtime\_II_t + Dtime\_III_t + Dgenre_i + Dnetwork_j + \varepsilon_{ijt}$.

Logit-polynomial: $Availability_{ijt} = user\_count'_{jt} + user\_count'^2_{jt} + user\_count'^3_{jt} + Dbroadband_{ijt} + Dtime\_II_t + Dtime\_III_t + Dgenre_i + Dnetwork_j + \varepsilon_{ijt}$.

In the foregoing equations, $user\_count_{jt}$ is the user count on network j at time t. In the polynomial specification $user\_count'_{jt}$ indicates the centered user count.[11] We performed centering to reduce multicollinearity that may occur among the polynomials. $Dbroadband_{ijt}$ is a 0-1 indicator of whether song i is available via a broadband connection on network j at time t. $Dtime\_II_t$ and $Dtime\_III_t$ are indicator variables for Time periods II and III. $Dgenre_i$ and $Dnetwork_j$ are dummy variables for the 17 song genres and the 6 networks, respectively.

The replication equations are similar to the availability equations; however, because replication is a count of the number of copies, the equations are estimated using ordinary least squares.

[11] $user\_count'_{jt} = \overline{user\_count} - user\_count_{jt}$.

OLS-linear: $Replication_{ijt} = user\_count_{jt} + Dbroadband_{ijt} + Dtime\_II_t + Dtime\_III_t + Dgenre_i + Dnetwork_j + \varepsilon_{ijt}$.

OLS-log: $Replication_{ijt} = \ln(user\_count)_{jt} + Dbroadband_{ijt} + Dtime\_II_t + Dtime\_III_t + Dgenre_i + Dnetwork_j + \varepsilon_{ijt}$.

OLS-polynomial: $Replication_{ijt} = user\_count'_{jt} + user\_count'^2_{jt} + user\_count'^3_{jt} + Dbroadband_{ijt} + Dtime\_II_t + Dtime\_III_t + Dgenre_i + Dnetwork_j + \varepsilon_{ijt}$.

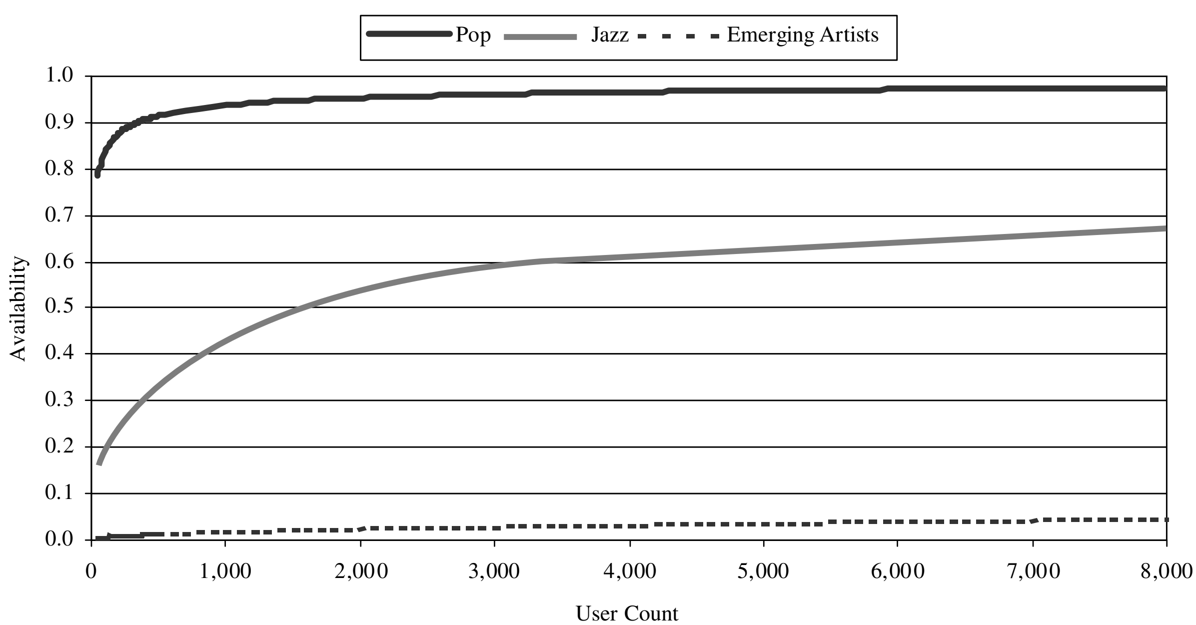
Results for both sets of regressions are presented in Table 3. In all cases, the coefficient on user count is positive, indicating that both availability and replications are increasing with the number of users. Note that the log and polynomial specifications have higher R^2 measures than the linear specifications, suggesting that they offer a better fit.

The concavity of availability in network size is consistent with expectations and suggests that the probability that a user contributes to network resources is either constant or decreasing in network size. Figures 3 and 4 show these results graphically for the Pop, Jazz, and Emerging Artist genres. Consistent with Hypothesis 1, the marginal value a user provides to the network in terms of availability and replication decline with network size. The concavity of replication with network size is particularly interesting given that, as noted above, in the absence of free-riding we would expect replication to scale linearly with network size. The results shown are consistent with an increase in free-riding caused by larger group size, users with a greater propen914sity to free-ride selecting the more popular networks, or both.

5.2. Negative Network Externalities

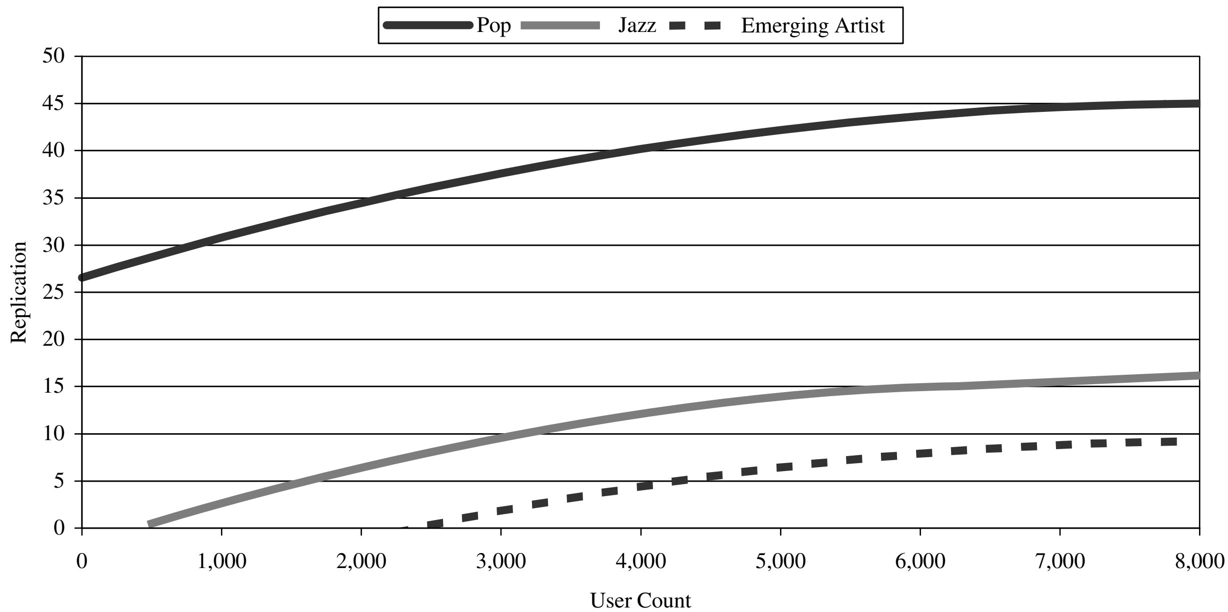
Negative network externalities may be reflected in four measures: (1) an increase in the number of login retries necessary to access the network, (2) longer query times, (3) an increase in the number of queued download attempts, and (4) longer download times. We use four separate regressions to analyze how these measures change with the number of users on a network. The first two measures of congestion were collected at the same time we collected the data on availability. As noted above, our data on download congestion and speed were collected between March 28, 2001, and April 19, 2001 (Tables 1 and 2). Obviously it would have been better to collect these data at the same time as the other data. That said,

Figure 3 Availability Regression Result*



*In each case, we used the coefficients for the specification that provided the best fit. User count was allowed to vary from 0 to 8,000 users and all other variables were set to their respective sample averages.

Figure 4 **Replication Regression Result**



we conducted tests using the longer data set to determine whether the relationship between availability, replications, congestion, and user count changed dramatically across the three time periods. In all cases, our results for the individual time periods were consistent with the full data set. This suggests that while we might have achieved higher or lower coefficients if we had data on download congestion and download time for the earlier periods, our core results almost certainly would still hold.

Both (1) login retries and (3) download attempts are count data. Count data can be estimated by ordinary least squares, but are typically estimated using the Poisson model, which accounts for the discrete nature of the data. We use the Poisson model to estimate download attempts data. In the case of login retries, we use a zero-inflated Poisson regression model to control for the fact that below network capacity, no retries are necessary. In essence two processes generate the login retries data. One determines the number of retries when the system is not at capacity, and the other determines how many retries are necessary when the system is at capacity. The zero-inflated Poisson regression is appropriate for this type of data because it estimates the two generating processes separately: the probability that there is no congestion (zero inflation) and, conditional on congestion, the number of retries needed (Poisson) (Lambert 1992).

Our specifications for these regressions are as follows:

ZIP-Inflated: $\textit{Probability of no login congestion}_{jt} = \textit{user\_count}_{jt} + \textit{server\_count}_{jt} + \textit{Dtime\_II}_t + \textit{Dtime\_III}_t + \textit{Dnetwork}_j + \varepsilon_{jt}$.

ZIP-Poisson: $\textit{Number of login retries}_{jt} = \textit{user\_count}_{jt} + \textit{server\_count}_{jt} + \textit{Dtime\_II}_t + \textit{Dtime\_III}_t + \textit{Dnetwork}_j + \varepsilon_{jt}$.

Poisson: $\textit{Number of download attempts}_{ijt} = \textit{user\_count}_{jt} + \textit{Dnetwork}_j + \textit{Dgenre}_i + \varepsilon_{ijt}$.

Both (2) query times and (4) download times are estimated using ordinary least squares regressions, where the dependent variable is the log of query time or the log of download time. We estimated each of these relationships by using the same three functional forms for user count that we used in the positive externalities case: linear, log, and polynomial. In the interest of space, we report the regression with the best fit. The variables in the regressions are the same as in the previous set of regressions, with one exception. In the case of download speed, we created 10, 0-1 indicator variables corresponding to the 10 self-reported connection speeds displayed in OpenNap query results (i.e., 14.4 kbps, 28.8 kbps, 33.6 kbps, 56.7 kbps, 64K ISDN, 128K ISDN, Cable,

Table 4 Regression Results for Negative Network Externalities

	1a	1a	2	3	4
Regression	Login inflated	Login Poisson	Query time	Download attempts	Download speed
Method	Zero-inflated Poisson		OLS	Poisson	OLS
Data period	December 2000–April 2001			March–April 2001	
Dependent variable	Prob. no congestion	# of login retries	Log(query time)	# of download attempts	Log(download time)
user_count	−7.1e-04	2.8e-04	4.7e-04	4.17e-04	−4.1
	(1.7e-04)	(4.9e-05)	(−7.3e-05)	(−6.9e-05)	(1.98)
server_count	*−0.056*	*0.0085*	−0.079		
	(*0.064*)	(*0.148*)	(0.029)		
Time period [2]	Yes	Yes	Yes		
Network [5]	Yes	Yes	Yes	Yes	
Song genre [16]				Yes	Yes
Connection speed [10]					Yes
Observations	323	323	323	582	582
(Pseudo) R^2			0.45	0.12	0.18

Notes. Standard errors are in parentheses. Values in brackets denote the number of fixed effect variables. Fixed effects are suppressed for simplicity. Italicized coefficients are insignificant ($P = 0.05$). "Yes" means control variables are included for the specified dependent variables.

DSL, T1, and T3 or greater). Our specifications are as follows:

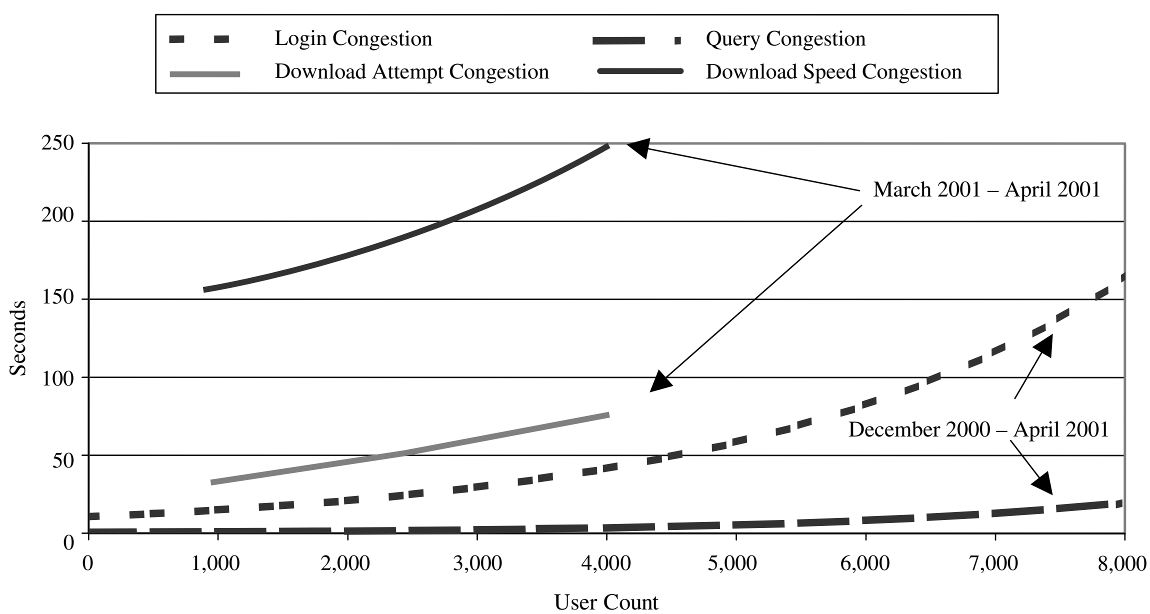
OLS-Log: $\text{Log(Query Time)}_{jt} = user\_count_{jt} + server\_count_{jt} + Dtime\_II_t + Dtime\_III_t + Dnetwork_j + \varepsilon_{jt}$.

OLS-Linear: $\text{Log(Download time)}_{ijt} = user\_count_{jt} + Dgenre_i + Dconnection_{ijt} + \varepsilon_{ijt}$.

Table 4 presents the results of the regressions. Consistent with Hypothesis 2, in all of the cases, congestion is increasing in network size at an increasing rate.

The relationship between user count and congestion is shown graphically in Figure 5, which projects our results in terms of length in seconds. To display each measure on a common figure (with the y-axis

Figure 5 Congestion Summary



measuring seconds), we assume that each login retry and download attempt takes 12 and 15 seconds, respectively, and convert the observed download speed (kbps) into the amount of time necessary to download a 5MB file over a cable modem. These assumptions reflect the average values in our empirical analysis and are made for graphical simplicity only.

5.3. The Impact of Increasing Server Capacity

Our empirical results are consistent with Hypotheses 1 and 2, which in turn suggest that network utility is concave in the number of users and that the optimal network size is bounded in the number of users. One question that may arise from this analysis is how these bounds will change as capacity is added to the network.

$$P(login\ congestion) = \frac{\rho^C/c!}{\sum_{i=0}^{c}(\rho^i/i!)}. \tag{12}$$

The Erlang-B equation models the probability of congestion in a central switch with call handling capacity c and $\rho = \lambda/\mu$, where λ and μ are Poisson random variables denoting the average number of users who arrive at a network each day (λ) and the service rate for each connection (μ). In our setting, the call handling capacity c corresponds to the servers' capacity to maintain multiple connections. The service rate μ is the duration of time that peers stay on the network, and λ is the rate at which peers arrive at the network.

We calibrate the model parameters as follows. Our empirical data indicate that on average users hold a connection for 12 hours, therefore $\mu = 2$ connections per day. We use two capacity sizes: $c = 4{,}000$ (approximately the mean network size in our data), and $c = 6{,}000$ (a larger network in our data). To model increases in arrival rate resulting from increasing capacities, we allow λ to vary between 0 and 40,000 users per day.

Given the probability of congestion from the Erlang-B model, we model the number of retries before a successful login as a geometric random variable. A geometric random variable corresponds to the expected number of Bernoulli trials before success. It is directly applicable to our environment where users face a blocking probability, derived by our Erlang-B model. The average number of retries is given by the mean of the geometric random variable (Wackerly et al. 2001):[12]

$$E(login\ retries) = \frac{1}{1 - P(login\ congestion)}. \tag{13}$$

Figure 6 illustrates the results of this analysis. For any given arrival rate, it is clear that as capacity increases, both measures of congestion decrease. However, as the arrival rate increases, the same levels of congestion recur in the higher capacity network. Further, as noted in §5.1, the additional users attracted by the additional capacity provide value in terms of availability and replication at a diminishing rate.

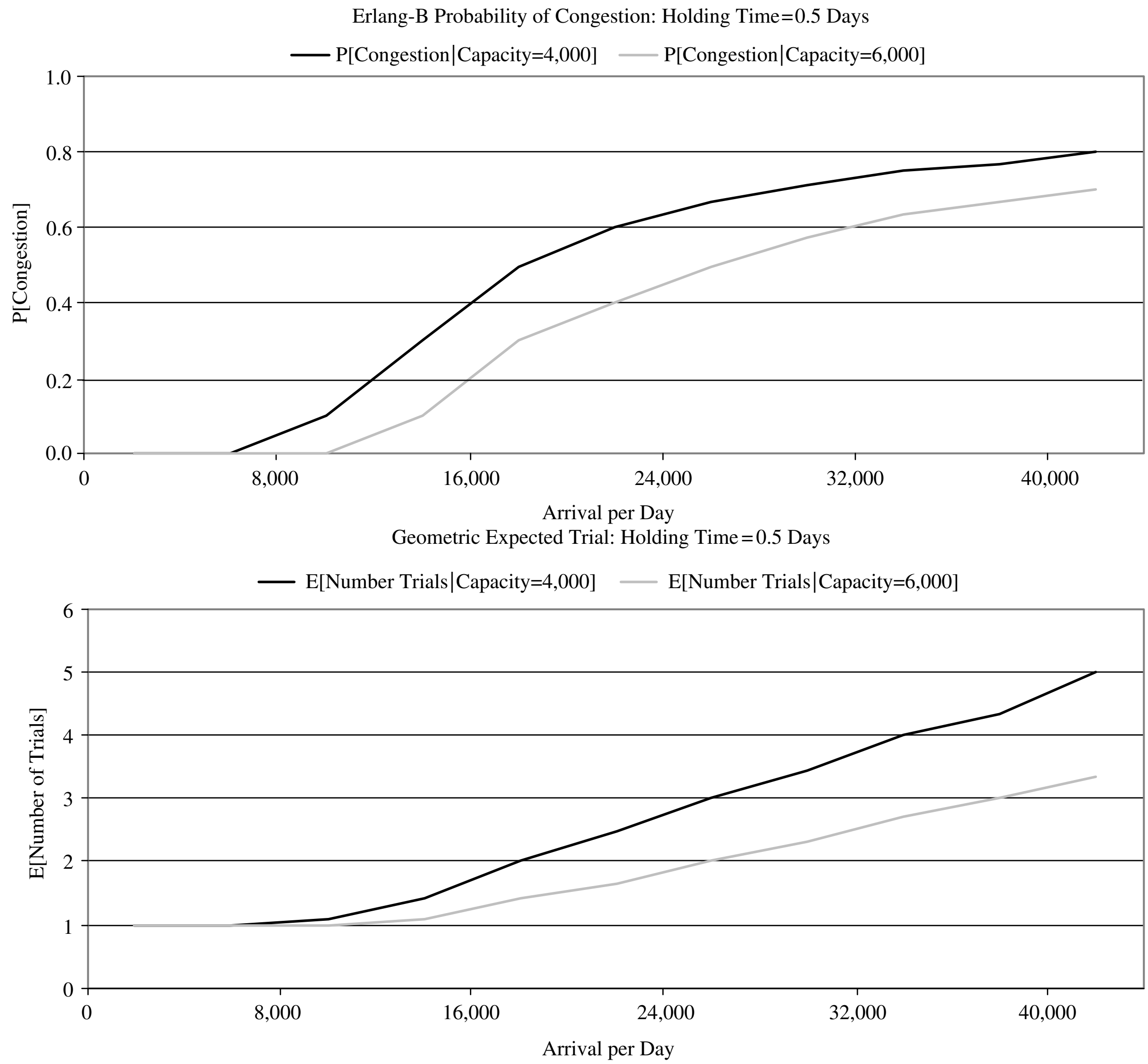
Together this analysis suggests that for the centralized P2P architecture, while additional server capacity will allow more users to join the network, for a sufficiently large network the marginal benefit these additional users will bring to the community may not justify the cost of the additional capacity. This is particularly true because additional users add value to the network at a diminishing rate. Thus, with the arrival of new users, congestion on this larger network will eventually rise to the same levels as before, with potentially little gain in value from replication or availability. Further, additional capacity does not solve the primary user-level problems demonstrated above: increasing free-riding, increasing download attempt congestion, and decreasing download speed with larger networks. Additional server capacity also raises new sources of congestion, such as the overhead cost necessary to maintain mirrored copies of the central database across multiple servers. In sum, while increased capacity may increase the optimal size of a network, it is unlikely to eliminate the upper bound on optimal network size altogether.

5.4. Limitations and Generalizability

Several limitations of our study deserve further discussion. First, three external factors could affect our conclusion that users of larger networks were also more likely to free-ride. It is possible that the RIAA's legal actions against Napster during our time period caused an increase over time in the proportion of

[12] Note that in the data collected, the agent repeatedly tried to log in without any wait time. Thus, the retry attempts may not be independent, as assumed by the geometric model. Nevertheless, the model should yield generally consistent results when compared to our data.

Figure 6 Illustration of Increasing Capacity on Login Congestion



OpenNap users who free-ride. It is also possible that Napster users are more likely to free-ride than OpenNap users and the influx of Napster users in our later time periods led to an increase in free-riding. Finally, it is possible that users who are more likely to free-ride are also more likely to search out the largest OpenNap network, or alternately that smaller networks are more likely to identify and remove free-riding users.

However, we have reason to believe that these potential limitations are not driving our result that the propensity to free-ride increases in larger networks. First, all of our results still hold if the regressions are run within each individual time period. Because the time periods mark important announcements in the RIAA's legal case, and in Napster's subsequent restructuring, the fact that our results hold within time periods (in addition to across time periods) suggests that the RIAA's legal actions and the influx of former Napster users are not driving our results. Second, it is important to note that throughout our data collection, the RIAA was only targeting network operators with legal action, not individual peers—and that most of their energies were focused on the Napster network, not OpenNap. Thus, individual peers had little reason during this timeframe to turn off sharing for legal reasons (Kiron 2001). Finally, our data collection software agent, which did not share files, was never blocked from an OpenNap network, suggesting that the networks in our sample, whether large or small, did not police users to remove free-riders.

A second limitation of our study is that our results should be interpreted as applying to centralized P2P architectures. Analyzing positive and negative network externalities in decentralized and hybrid architectures would be a useful area for future research. However, the design of each of these architectures is consistent with our findings. As noted in §2.1, the designers of the major decentralized and hybrid networks have limited the *effective* size of the network each peer can reach by limiting the time-to-live parameter on interpeer and interultrapeer queries. Relaxing this time-to-live parameter would increase the number of other peers a particular user could reach and therefore increase each peer's likelihood of finding the content they sought. A natural explanation for limiting the time-to-live parameter is that at some point the marginal benefit of increased network reach to a particular user does not justify the marginal cost that their query would impose on the network. This interpretation is entirely consistent with our finding that the optimal network size of a centralized P2P network architecture is bounded because the marginal value additional users provide to the network decreases in network size, while the marginal cost they impose on the network increases in network size.

Finally, our results should be interpreted as applying to *consumer* P2P file sharing networks. These networks are increasingly being used in corporate settings (e.g., Deloitte and Touche's use of NextPage for knowledge management, [Fontana 2002]) and in nonmusic sharing settings (e.g., Virgin Record's use of Blue Falcon for streaming media, [Schonfeld 2002]), and the positive and negative network externalities in these settings may differ from our environment.

6. Discussion

P2P networks are an important architecture for the distribution of information goods and show promise to become an important medium for resource sharing in other domains. However, in spite of their potential importance, there has been little empirical research into the impact of user behavior and network design on the operation of these networks.

Our research seeks to bridge this gap by measuring the impact of positive and negative network externalities on the scalability of OpenNap P2P networks. Much of the early discussion of P2P network scalability has focused exclusively on positive network externalities, concluding that larger P2P networks will always provide better performance. Using data gathered from the six most popular OpenNap networks from December 2000 to April 2001, we find that this is not necessarily the case. In our data, the marginal value that an additional user brings to the network declines in larger networks, and the marginal cost that an additional user imposes on the network increases in larger networks. Together, these results suggest that the optimal size of this centralized P2P network is bounded—at some point the marginal cost an additional user will impose on other users will be larger than the benefit they provide.

It is important to note that in P2P environments where the optimal network size is bounded, network designers retain a variety of options to improve network scalability and performance. First, in centralized P2P networks, increased server capacity has the ability to relax congestion constraints, increase (though not without bound) the optimal size of the network, and thereby improve network performance to individual users. Network operators can use our methods to measure how the value and costs that additional users impose on other network members change with additional capacity, and they can use these calculations to determine when the expense of additional capacity would be justified by the resulting increase in network performance.

Similarly, in hybrid and decentralized networks, network operators can choose the reach of the network to optimize performance. As noted in §2.1, while many popular hybrid and decentralized networks appear to function as a single network, they are in fact designed such that the number of other users an individual peer can reach is limited by the time-to-live packet on their queries (typically set to seven). By limiting network reach in this way, network operators are able to balance the benefits the individual peer receives from being able to query more users against the costs that this increased reach imposes on other nodes who have to process these queries. The optimal size of parameters determining the reach of the network may change over time. The value of increased reach to an individual peer will increase with diversity of content shared on the network and user preferences for this content. Likewise, the cost imposed on other members of the network in processing extra queries will decrease as the bandwidth and processing power available to these peers increase. Network designers can consider these factors when initially designing the network and should periodically revisit their choices to ensure that their networks are operating efficiently.

An effective limit on the size of local networks also suggests that network operators should ensure that users are clustered with other users sharing similar content interests. In many popular P2P networks (e.g., Kazaa and Gnutella), users are randomly assigned to a position in the network. Recent research suggests that query performance can be improved by enabling users to identify and join subnetworks consisting of other users who have similar interests (Asvanund et al. 2003).

Finally, our results suggest that the increased propensity of users to free-ride in larger networks is an important factor limiting network scalability. This finding, while initially surprising, is consistent with findings in the public economics literature that free-riding worsens with group size for the private provision of public goods in more general environments. For P2P networks, free-riding limits scalability because it damages the auto-replication characteristic of P2P networks discussed in §2.1. Auto-replication supports scalability because content on the network is available in proportion to its popularity and because the consumption of network resources through downloads is balanced by the provision of network resources through sharing. Because free-riding limits auto-replication and network scalability, network operators should consider designing user incentives to reduce the propensity of free-riding. Some potential incentives include decreased search, login, or download performance for free-riders (Krishnan et al. 2003) or, where appropriate, monetary payments in exchange for consumption of network resources (Cole et al. 2003, Golle et al. 2001).

A related implication of our results is that copyright holders seeking to limit unauthorized file sharing may wish to adopt strategies that increase free-riding on P2P networks. Recently, copyright holders appear to be moving away from strategies that focus solely on shutting down the networks themselves. The problem with focusing solely on shutting down networks is that, because of the nature of network externalities, the number of networks (or Kazaa and Gnutella local subnetworks) is quite large. Furthermore, shutting down individual networks does little to change user behavior: Individual users simply look for new networks where they can trade files.

The more recent strategy adopted by copyright holders of bringing legal action against violators may be more successful even though the proportion of users who are targeted is a small fraction of the total number of users. The success of this strategy depends on raising the implicit cost of sharing for users by raising their legal risks. Increased sharing costs will then raise their propensity to free-ride and may ultimately reduce the utility offered by "illicit" file trading over P2P networks enough to make the legitimate purchase of the music an attractive option for users.

Acknowledgments
The authors thank participants at the Carnegie Mellon University Graduate School of Industrial Administration; the University of California, Berkeley SIMS and Haas Schools; the University of Maryland Laboratory for Telecommunications Science; the 2002 International Conference on Information Systems (ICIS); and the 2001 Telecommunications Policy Research Conference (TPRC). The authors also thank the editors of *ISR* and three anonymous referees for valuable comments on this research. Financial support was provided by the National Science Foundation through Grant IIS-0118767.

References

Adar, E., B. A. Huberman. 2000. Free riding on Gnutella. *First Monday: Peer-Reviewed J. Internet* **5**(10).

Asvanund, A., R. Krishnan, M. D. Smith, R. Telang. 2003. Intelligent club management in peer-to-peer networks. Working paper, Carnegie Mellon University, Pittsburgh, PA.

Augereau, A., S. Greenstein, M. Rysman. 2003. Coordination versus differentiation in a standards war: The adoption of 56K modems. Working paper, Boston University, Boston, MA.

Avery, S. 2002. Hollywood takes aim at online film-swappers. *Columbus Dispatch* (July 20) 1F.

Ba, S., J. Stallaert, A. B. Whinston. 2001. Optimal investment in knowledge within a firm using a market mechanism. *Management Sci.* **47**(9) 1203–1219.

Bhattacharjee, B., S. Chawathe, V. Gopalakrishnan, P. Keleher, B. Silaghi. 2003. Efficient peer-to-peer searches using result-caching. *Proc. 2nd Internat. Workshop Peer-to-Peer Systems*, Berkeley, CA.

Brynjolfsson, E., C. Kemerer. 1996. Network externalities in microcomputer software: An econometric analysis of the spreadsheet market. *Management Sci.* **42**(12) 1627–1647.

Buchanan, J. M. 1965. An economic theory of clubs. *Economica* **32**(125) 1–14.

Butler, B. S. 2001. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Inform. Systems Res.* **12**(4) 346–365.

Cole, R., Y. Dodis, T. Roughgarden. 2003. Pricing network edges for heterogeneous selfish users. *Proc. 35th Annual ACM Sympos. Theory Comput.*, 521–530.

Conner, K. R. 1995. Obtaining strategic advantage from being imitated: When can encouraging "clones" pay? *Management Sci.* **41**(2) 209–225.

Constant, D., L. Sproull, S. Kiesler. 1996. The kindness of strangers: The usefulness of electronic weak ties for technical advice. *Organ. Sci.* **7** 119–135.

Dranove, D., N. Gandal. 2003. The DVD vs. DIVX standard war: Network effects and empirical evidence of preannouncements effects. *J. Econom. Management Strategy* **12** 363–386.

Economides, N., C. Himmelberg. 1995. Critical mass and network size with application to the U.S. fax market. G. Brock, ed. *Toward a Competitive Telecommunications Industry: Selected Papers from the 1994 Telecommunications Policy Research Conference*. Lawrence Erlbaum Associates, New York.

Farrell, J., G. Saloner. 1987. Competition, compatibility and standards: The economics of horses, penguins and lemmings. H. L. Gabel, ed. *Product Standardization as a Tool of Competitive Strategy.* North-Holland, New York, 1–22.

Fontana, J. 2002. NextPage taking P2P content sharing offline. *NetworkWorldFusion* (April 12, 2004).

Forman, C., P.-Y. Chen. 2003. Network effects and switching costs in the market for routers and switches. Working paper, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA.

Gandal, N. 1994. Hedonic price indexes for spreadsheets and an empirical test for network externalities. *The RAND J. Econom.* **25**(1) 160–170.

Gandal, N., M. Kende, R. Rob. 2000. The dynamics of technological adoption in hardware/software systems: The case of compact disc players. *RAND J. Econom.* **31** 43–61.

Golle, P., K. Leyton-Brown, I. Mironov. 2001. Incentive for sharing in peer-to-peer networks. *Proc. 3rd ACM Conf. Electronic Commerce*, Tampa, FL, 264–267.

Goolsbee, A., P. Klenow. 2003. Evidence on learning and network externalities in the diffusion of home computers. *J. Law Econom.* **40**(2) 317–344.

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

Hibbard, J. 2001. Peer pressure: Forget the music business. In the latest attempt to commercialize peer-to-peer technology, startups want Akamai's market. *Red Herring* **103**(September 1) 62.

Katz, M., C. Shapiro. 1986. Technology adoption in the presence of network externalities. *J. Political Econom.* **94**(4) 822–841.

Katz, M., C. Shaprio. 1994. Systems competition and network effects. *J. Econom. Perspectives* **8**(2) 93–115.

Kauffman, R. J., J. McAndrews, Y.-M. Wang. 2000. Opening the "black box" of network adoption. *Inform. Systems Res.* **11**(1) 61–82.

Kirk, P. 2003. Gnutella protocol development—An overview of 0.6. protocol specification (July 5, 2003).

Kiron, D. 2001. Napster. Harvard Business School Case Number 9-801-219 (March 29).

Kraut, R. E., P. Attewell. 1993. Electronic mail and organizational knowledge. Working paper, Carnegie Mellon University, Pittsburgh, PA.

Krishnan, R., M. D. Smith, Z. Tang, R. Telang. 2003. The virtual commons: Why free riding can be tolerated in file sharing networks. Working paper, Carnegie Mellon University, Pittsburgh, PA.

Kung, H. T., C. Wu. 2003. Differentiated admission for peer-to-peer systems: Incentivizing peers to contribute their resources. Working paper, Harvard University, Cambridge, MA.

Lai, K., M. Feldman, I. Stoica, J. Chuang. 2003. Incentives for cooperation in peer-to-peer networks. Working paper, University of California at Berkeley, Berkeley, CA.

Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34** 1–14.

Minar, N., M. Hedlund. 2001. A network of peers: Peer-to-peer models through the history of the Internet. A. Oram, ed. *Peer-to-Peer: Harnessing the Power of Disruptive Technologies.* O'Reilly Publishing, Cambridge, MA, 3–20.

Parameswaran, M., A. Susarla, A. B. Whinston. 2001. P2P networking: An information sharing alternative. *IEEE Comput.* **34**(7) 31–38.

Park, S. 2003. Quantitative analysis of network externalities in competing technologies: The VCR case. Working paper, State University of New York Stony Brook, Stony Brook, NY.

Saloner, G., A. Shepard. 1995. Adoption of technologies with network effects: An empirical examination of the adoption of automated teller machines. *RAND J. Econom.* **26**(Autumn) 479–501.

Saloner, G., A. M. Spence. 2002. *Creating Value and Capturing Value: Perspectives and Cases on Electronic Commerce*. John Wiley and Sons, New York.

Samuelson, P. 1954. The pure theory of public goods. *The Rev. Econom. Statist.* **36**(4) 387–389.

Schonfeld, E. 2002. Peer-to-peer technology goes legit. *Business 2.0.* (April 12, 2004).

Sripanidkulchai, K., B. Maggs, H. Zhang. 2002. Enabling efficient content location and retrieval in peer-to-peer systems by exploiting locality in interests. *ACM SIGCOMM Comput. Communication Rev.* **32**(1) 80.

Stoica, I., R. Morris, D. Karger, M. F. Kaashoek, H. Balakrishnan. 2001. Chord: A scalable peer-to-peer lookup service for Internet applications. *Proc. ACM Special Interest Group Signal Comm.*, San Diego, CA.

Strahilevitz, L. J. 2002. Charismatic code, social norms, and the emergence of cooperation on the file-swapping networks. *Virginia Law Rev.* **89**(May) 505–596.

Subramani, M. R., N. Peddibhotla. 2002. Contributing to document repositories—An examination of prosocial behavior. Working paper, Carlson School of Management, University of Minnesota, Minneapolis, MN.

Varian, H. 2003. Economics of information technology. Working paper, University of California at Berkeley, Berkley, CA.

Wackerly, D., W. Mendenhall, R. Scheaffer. 2001. *Mathematical Statistics with Applications*, 6th ed. Duxbury Press, Florence, KY.

Warner, M. 2002. The new Napsters. *Fortune* **146**(3) 110–112.

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