

DRAFT
COMMENTS WELCOME

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

Atip Asvanund, Karen Clay, Ramayya Krishnan, Michael D. Smith

{atip, kclay, rk2x, mds}@andrew.cmu.edu

H. John Heinz III School of Public Policy and Management
Carnegie Mellon University, Pittsburgh, PA 15213

This Draft: March 2004

<http://ssrn.com/abstract=433780>

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

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

Abstract

Peer-to-peer file sharing networks are becoming an important medium for the distribution of information goods. However, there is little academic research into the optimal design of these networks under real-world conditions. Our research represents an initial effort to analyze the impact of positive and negative network externalities on the optimal size of these P2P networks. Our analysis uses a unique dataset collected from the six most popular OpenNap peer-to-peer networks between December 19, 2000 and April 22, 2001. We find that users contribute value to the network in terms of additional content and additional replicas of content at a diminishing rate, while they impose costs on the network in terms of congestion on shared resources at an increasing rate. Together these results suggest that the optimal size of these centralized peer-to-peer networks is bounded — at some point the costs a marginal user imposes on the network will exceed the value they provide.

1. Introduction

At their core, peer-to-peer (P2P) network architectures enable resource sharing directly between autonomous individual network users also known as peers. These resources are most commonly files containing digitized information content such as music, movies, pictures, software, or text, but can also include storage capacity, 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 from the traditional client-server architecture where all network resources are contained in and managed by a central server (Parameswaran, Susarla, Whinston 2001).

The recent popularity of P2P networking 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. Presently, the most popular such network is Kazaa, which according to Download.com has been downloaded over 330 million times since its introduction in July 2000.

More recently, entrepreneurs and programmers, recognizing the potential of P2P architectures to facilitate resource sharing by autonomous peers, have 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 archival (e.g., Publius, FreeHaven).

Despite their 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, a defining characteristic of these P2P networks, on their design and real-world operation. Systematic research to address these questions is important for a variety of constituencies including engineers designing protocols to support P2P networks, entrepreneurs developing P2P-based businesses, and intellectual property holders seeking to develop their own networks and to minimize the use of non-complying 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. In P2P networks, these externalities arise in two forms. First, users who choose to share their content bring positive externalities to the network in the form of new content and replicas of existing content. Viewed in isolation, these positive externalities mean that larger networks will provide more value to users than smaller networks do (e.g., Strahilevitz 2002). At the same time, users have the potential to impose negative externalities (e.g., congestion) on other network users through their consumption of scarce network resources, and this factor has received far less discussion. Furthermore, users may not internalize the positive and negative externalities their choices impose on other network users, potentially resulting in inefficiently low levels of content provision or inefficiently high levels of content consumption.

This study seeks to measure how these positive and negative externalities vary in P2P networks as a function of network size. We do this by gathering a unique dataset from the six most popular OpenNap networks from December 19, 2000 to April 22, 2001. 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 value an additional user provides to the network in terms of additional content or replicas of existing content decreases in network size while the marginal cost they impose on the network in terms of congestion on shared resources increases in network size. This suggests that the optimal size of P2P networks 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 benefit of increasing network capacity in reducing congestion. This model 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 imply decreasing benefits to 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 (see for example Saloner and Spence 2002, p. 54), ours is one of the first papers to measure the specific 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 and identifies areas for future research.

2. Background and Literature

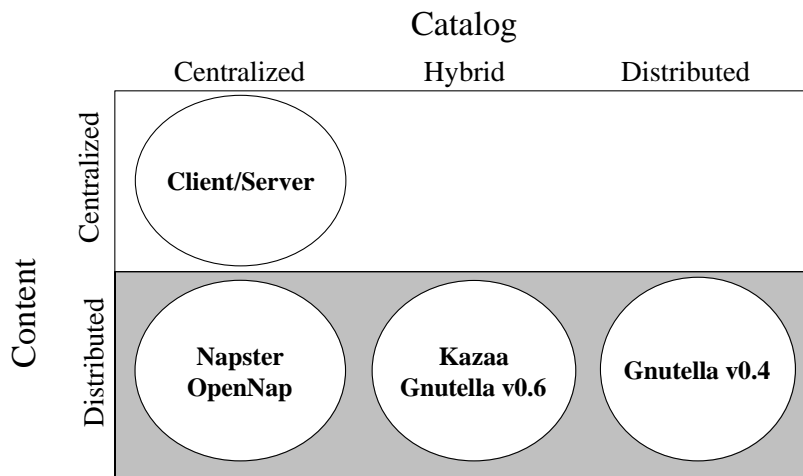
2.1. Architecture

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 determines 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 determines 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).

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.¹ 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 and 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.

Figure 1: Taxonomy of Content Distribution Architectures



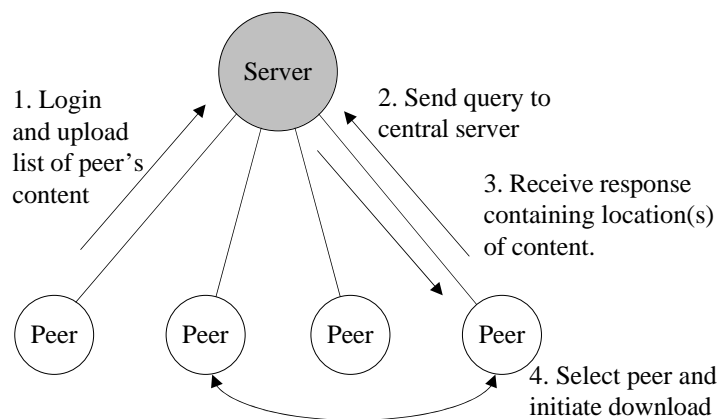
Centralized P2P Architectures: The two most popular and well-known centralized P2P architectures are Napster and OpenNap. As Napster grew in popularity after its introduction in May 1999, a group of programmers reverse engineered the Napster protocol and implemented an

¹ 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.

Open Source version of the protocol. Individuals established their own “Napster”-like networks by installing the OpenNap implementation. Many individuals did so and created hundreds of networks competing alongside Napster. These networks were distinct and separate from each other and from Napster’s closed network. During our study period, a user searching one OpenNap network would not see users logged into the other networks and most importantly actions taken by Napster, for example efforts to remove copyrighted materials late in our study, did not impact the availability of content on the OpenNap networks we study.

As noted above, these networks operated with a particular hierarchy. 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., listed on Napigator.com). The peer maintains a stateful connection to this central server for the duration of the user’s presence on the network. Because these servers have limited capacity for simultaneous stateful connections with peers, this creates a potential source of congestion.

Figure 2: Napster / OpenNap Network Operation



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 a ping request to each provider and sorts the list in ascending order by the amount of time it took to receive a pong message (a response) 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 3 other peers. Since 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 7. Each peer who receives a query message decrements the TTL value and only forwards queries when the TTL is greater than 0. The TTL effectively limits the size of the network each node can reach to on the order of 10,000 nodes.²

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, there is an explicit limit on the number of leaf nodes that can be reached from any particular leaf node. We discuss the implications of these designs on our methodology and findings in more detail below.

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 is increasing in the number of users that the individual can talk to. An indirect network externality arises when the utility of a product is increasing in 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).

In this section, we begin by reviewing prior empirical work on direct and indirect network externalities. We then discuss three other literatures that our work is related to: i) the literature on the motivation of participants in online forums, ii) the literature on how to explicitly create incentives for P2P participants to behave in certain ways, namely contribute content, and iii) the literature on the optimal design of P2P networks at the protocol level to address current problems.

While the role of positive direct network externalities in the context of telecommunications is well-known and widely discussed, there have been very few papers that have been able to

² See the Gnutella 0.4 protocol at <http://rfc-gnutella.sourceforge.net/> for more details on the protocol.

measure these effects directly. This is because the direct measurement of network externalities is complicated by limitations of available data, which are typically either time-series or cross-sectional (Varian 2003). If the data are time series it can be difficult to disentangle network effects from other unrelated changes (e.g., falling equipment prices). If the data are cross sectional it can be difficult to determine whether regional differences are attributable to peer group effects (differences in preferences) or network externalities. Thus, while the role of positive direct network externalities in the context of telecommunications is well known and widely discussed, there have been very few papers that have been able to measure these effects directly.

Among the papers that have empirically analyzed network effects, positive networks effects have been found to 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 1995) and video cassette recorders (Park 2003). Positive network effects have also been shown to influence the speed of adoption of technologies in the context of telecommunications equipment (Economides and Himmelberg 1995; Augereau, Greenstein, and Rysman 2003), banking networks (Saloner and Shepard 1995; Kauffman, McAndrews and Wang 2000) and consumer electronics equipment (Park 2003, Dranove and Gandal 2003; Goolsbee and Klenow 2003; Gandal, Kende, and Rob 2000). Finally, positive network effects have been shown to create switching costs for networking equipment (Forman and Chen 2003).

Our work differs from prior empirical work 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 paper to empirically analyze the impact of both positive and negative network externalities. Second, we are able to collect panel data (incorporating both time series

and cross-sectional observations) where the cross-sectional observations are made across networks using exactly the same server software and protocol design, limiting uncontrolled differences across networks. Thus, the nature of this data solves some of the data problems mentioned above that have limited the ability of researchers to empirically measure network effects using only time series of cross-sectional data alone. Third, we analyze network externalities in the context of P2P networks. P2P networks 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.

As noted above, our work is related to three additional literatures. The first literature concerns the motivation of participants in online forums. This literature 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, Sproull, and Kiesler 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 related to Butler (2001) who develops a resource-based model of social interaction in online communities and applies this model to data

from listserv communities on the Internet. He finds that larger networks have both advantages and disadvantages with respect to attracting and retaining members. Advantages in Bulter's setting derive from increased potential interactions with other participants and through an increased audience, and disadvantages derive from fewer opportunities to participate, reduced opportunities to form personal relationships with other network members, and lower levels of contribution by members. He concludes that online communities face many of the same size constraints as offline communities do.

We differ 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 compare empirically 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 literature is the literature on how to explicitly create incentives for P2P participants to behave in certain ways. This literature builds on the economics literature on public and club goods (Samuelson 1954, Buchanan 1965). The services provided over P2P networks have some of the characteristics of public goods. Membership is typically not controlled and once a user gains access to the network they can access all the resources provided by the network, so the networks are non-excludable. In the absence of free riding, song replication should scale with network size, creating a situation where the consumption of network resources is non-rivalrous. If users do free ride, however, rivalry can emerge among users for scarce network content and bandwidth. P2P networks can impose some degree of excludability through membership rules or place limits on access to network services, making them more similar to clubs.

There are, however, notable differences between the content offered in P2P networks and typical public and club goods. In a typical public or club goods settings, non-contribution is the “default” choice, whereas typical P2P client programs are designed to contribute (share files) as the default. Which means that turning off contribution requires the P2P user to take action.

Further, in typical public and club goods environments, contribution is in the form of monetary outlay and is separate from consumption. In P2P environments, consumers become contributors by default — downloading a good from the network makes a person a contributor of that good if they do not free ride. This unique characteristic of P2P goods means that in an ideal case the heaviest consumers of network resources can also be among the most valuable contributors.

Drawing on the suboptimality results from the public and club goods literatures and indications of suboptimality in practice (e.g., Adar and Huberman 2000), there has been a recent emergence of papers from computer science to address this problem. The focus has been on providing participants with explicit incentives to provide content and other resources. Researchers have investigated network pricing (Cole, Dodis, and Roughgarden 2003); micro-payment systems (Golle, Leyton-Brown and Mironov 2001); reputation systems (Lai et al. 2003); autonomous club formation (Asvanund et al. 2003); and admission control systems (Kung and Wu 2003).

The third literature is the literature on the optimal design of P2P networks at the protocol level to address current problems and improve performance. For example, researchers have investigated enhancing network performance 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, Maggs, and Zhang 2001). Here again, our

research differs from this literature by focusing on the impact of user behavior on network performance.

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 they are interested in (F) and the (dis)utility of a vector of congestion effects they face (C):

$$U(F(N), C(N)) = U_F(F(N)) + U_C(C(N)) \tag{1}$$

Modeling utility as separable in consumption and congestion appears to be 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:

$$\partial U / \partial f > 0 \tag{2}$$

$$\partial U / \partial c < 0 \tag{3}$$

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:

$$\partial f / \partial N \geq 0 \tag{4}$$

Likewise, with regard to congestion, new users will (weakly) increase congestion measures, which in our setting include login time, query time, and download times:³

$$\partial c / \partial N \geq 0 \tag{5}$$

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

³ 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” file 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

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

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

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.”⁴

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:⁵

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

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.

$$\partial^2 c / \partial N^2 > 0 \tag{11}$$

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 users bring to the network, the marginal increase in value decreases in network size.

Hypothesis 2: For all measures of cost users impose on other members of the network, the marginal increase in cost increases in network size.

where (10) corresponds to the statement of Hypothesis 1 and (11) corresponds to our statement of Hypothesis 2.

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 available on the network. Replication is a particularly important measure of network behavior. As noted above, a default property of the Napster software is that consumers of a song also become providers of the song, 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 in order for download performance to scale

⁴ We thank an anonymous referee for this insight.

⁵ In fact only one of the inequalities in (10) and (11) needs to strict for $\partial^2 U / \partial N^2$ to be strictly negative.

well. However, this will not be the case to the extent that users choose to free ride by consuming network resources while disabling sharing. We discuss this in more detail below.

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 2000 to April 2001 and include user count, server count, login congestion, query congestion, and song availability, song replication, and broadband song replication (Table 1).

Login congestion measures the difficulty of logging on to the network. This requires the client to establish a stateful connection with the catalog sever. As noted in Section 2.1, the server has a fixed capacity for simultaneous stateful connections with clients. We expect login congestion to be low initially and to quickly rise as network size approaches server capacity. Query congestion measures the delay in waiting for a search query result. When users perform search queries for a file, they place traffic demand on the centralized servers that perform database lookups, 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. 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.

Table 1: Summary Statistics

| <i>Variable</i> | <i>Obs.</i> | <i>Mean</i> | <i>St. Dev.</i> | <i>Min</i> | <i>Max</i> |
|--|-------------|-------------|-----------------|------------|------------|
| Login Congestion, Query Congestion, and Song Availability | | | | | |
| User Count | 323 | 3,118 | 2,283 | 68 | 8,618 |
| Server Count | 323 | 7 | 3.40 | 1 | 15 |
| Song Availability | 83,640 | 0.54 | 0.50 | 0 | 1 |
| Song Availability (Broadband Connection) | 83,640 | 0.45 | 0.50 | 0 | 1 |
| Replication (Number of copies of a song) | 83,640 | 11 | 30 | 0 | 555 |
| Replication (Broadband Connections) | 83,640 | 6 | 21 | 0 | 460 |
| Login Congestion (Seconds) | 323 | 3 | 8 | 0 | 71 |
| Query Time (Seconds) | 323 | 10 | 17 | 0.13 | 90 |
| Download Attempts and Speed | | | | | |
| User Count | 13 | 2,620 | 687 | 1,458 | 3,588 |
| Download Attempts | 582 | 2.85 | 4.37 | 1 | 45 |
| Download Speed (kbps) | 582 | 32 | 33 | 0 | 200 |

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.⁶

We also developed control variables to mark significant announcements made by Napster during our study period. In our regressions, Time Period I, refers to the start of the data collection until 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 between January 29, 2001 and March 2, 2001 when Napster started filtering copyrighted tracks from its service. Time Period III refers to the period of March 2, 2001 until the end of our data collection period. It is important to note that these 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 allow us to control for such changes.

To further explore how congestion varied with network size, we collected an additional dataset on download congestion and speed in March 28, 2001 to April 19, 2001 (Table 1). This dataset includes information on the size of the network and two measures of the congestion a user would

⁶ 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 5 songs that were contained in

face when trying to download a song. The first measure, download attempts, is the number of download attempts our agent had to make (starting with the listing with the lowest ping time) before finding a peer that did not queue the download request. As noted in Section 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. The second measure is the download speed (in kbps) our agent observed when downloading the song.

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 availability are presented in the first three columns of Table 2. Availability is measured as a binary value, where 0 indicates that the song is not available and 1 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.

albums that moved off the Billboard album charts during our sample period. Eliminating these songs from our

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.⁷ 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 indicators variables for time periods II and III. $Dgenre_i$ and $Dnetwork_j$ are dummy variables for the 17 song genres and the six networks respectively.

The replication equations are similar to the availability equations, except that as replication is a count of the number of copies, the equations are estimated using ordinary least squares.

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

analysis did not change any of our results.

⁷ $user_count'_{jt} = user_count_{jt} - user_count_{jt}$.

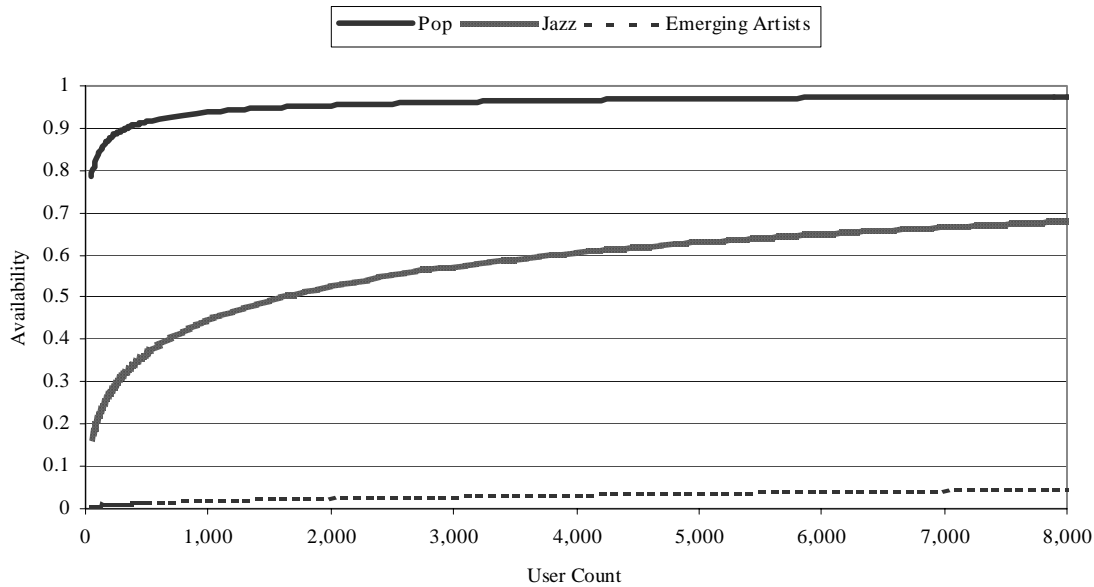
Table 2: Regression Results for Positive Network Externalities

| | 1 | 2 | 3 | 4 | 5 | 6 |
|--------------------------|------------------------|---------------------|-------------------------|------------------------|------------------------|-------------------------|
| <i>Regression Model</i> | <i>Logit-Linear</i> | <i>Logit- Log</i> | <i>Logit-Polynomial</i> | <i>OLS-Linear</i> | <i>OLS-Natural Log</i> | <i>OLS-Polynomial</i> |
| 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 (.0117) | -0.486 (0.0118) | -0.485 (.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=.05).

The concavity of availability in network size is consistent with expectations and suggests that that the probability 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 propensity to free ride selecting the more popular networks, or both. We discuss this interpretation in more detail below.

Figure 3: Availability Regression Result*



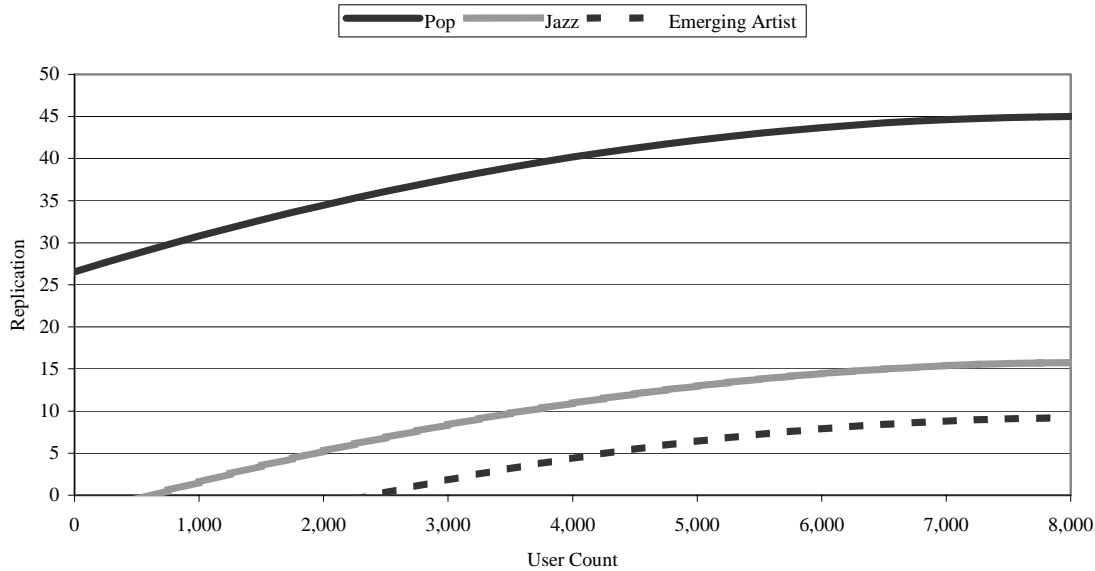
* In each case, we used the coefficients for the specification that provided the best fit. User count was allowed to vary from 0 to 8,000 users and all other variables were set to their respective sample averages.

5.2. *Negative Network Externalities*

Negative network externalities may be reflected in four measures: i) an increase in the number of login retries necessary to access the network, ii) longer query times, iii) an increase in the number of queued download attempts, and iv) 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 (Table 1). Obviously it would have been better to collect these data at the same time as the other data. That said, 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 gotten higher or lower coefficients had we had data on download congestion and download time for the earlier periods, our core results almost certainly would still hold.

Figure 4: Replication Regression Result



Both i) login retries and iii) 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, there are two processes generating the login retries data. One that determines the number of retries when the system is not at capacity and one that determines how many retries are necessary given that 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: $Probability\ of\ no\ login\ congestion_{jt} = user_count_{jt} + server_count_{jt} + Dtime_II_t + Dtime_III_t + Dnetwork_j + \varepsilon_{jt}$

ZIP-Poisson: $Number\ of\ login\ retries_{jt} = user_count_{jt} + server_count_{jt} + Dtime_II_t + Dtime_III_t + Dnetwork_j + \varepsilon_{jt}$

Poisson: $Number\ of\ download\ attempts_{ijt} = user_count_{jt} + Dnetwork_j + Dgenre_i + \varepsilon_{ijt}$

Both ii) query times and iv) 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 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 ten 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, DSL, T1, and T3 or greater). Our specifications are as follows:

OLS — Log: $Log(Query\ Time)_{jt} = user_count_{jt} + server_count_{jt} + Dtime_II_t + Dtime_III_t + Dnetwork_j + \varepsilon_{jt}$

OLS — Linear: $Log(Download\ time)_{ijt} = user_count_{jt} + Dgenre_i + Dconnection_{ijt} + \varepsilon_{ijt}$

Table 3 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. We assume each login retry and download attempt to take 12 and 15 seconds respectively. We estimate download speed for downloading a

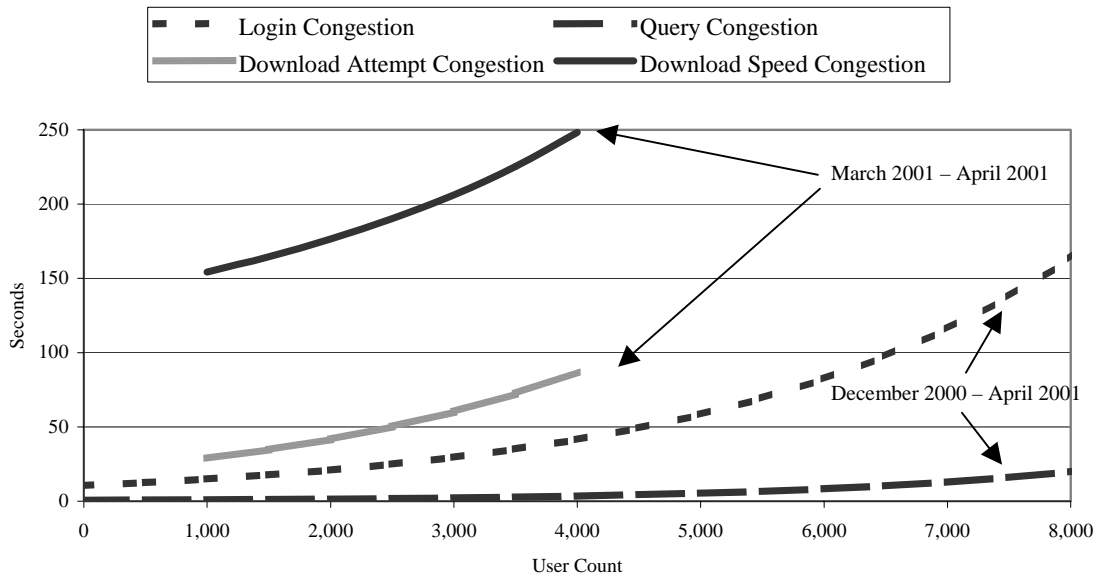
5MB file from a cable modem. These assumptions reflect the average values in our empirical analysis.

Table 3: Regression Results for Negative Network Externalities

| | 1a | 1a | 2 | 3 | 4 |
|---------------------------|-----------------------------------|---------------------------|------------------------|-------------------------------|---------------------------|
| <i>Regression</i> | <i>Login Inflated</i> | <i>Login Poisson</i> | <i>Query Time</i> | <i>Download Attempts</i> | <i>Download Speed</i> |
| <i>Method</i> | <i>Zero-Inflated Poisson</i> | | <i>OLS</i> | <i>Poisson</i> | <i>OLS</i> |
| <i>Data Period</i> | <i>December 2000 - April 2001</i> | | | <i>March - April 2001</i> | |
| <i>Dependent Variable</i> | <i>Prob. No Congestion</i> | <i># of Login Retries</i> | <i>Log(Query Time)</i> | <i># of Download Attempts</i> | <i>Log(Download Time)</i> |
| user_count | -7.1e-04 (1.7e-04) | 2.8e-04 (4.9e-05) | 4.7e-04 (-7.3e-05) | 4.17e-04 (-6.9e-05) | -4.1 (1.98) |
| server_count | -0.056 (0.064) | 0.0085 (0.148) | -0.079 (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=.05).

Figure 5: Congestion Summary



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.

$$P(\text{login congestion}) = \frac{\rho^c}{c!} \bigg/ \sum_{i=0}^c \frac{\rho^i}{i!} \quad (12)$$

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 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 stateful 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, Mendenhall, and Scheaffer 2001):⁸

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

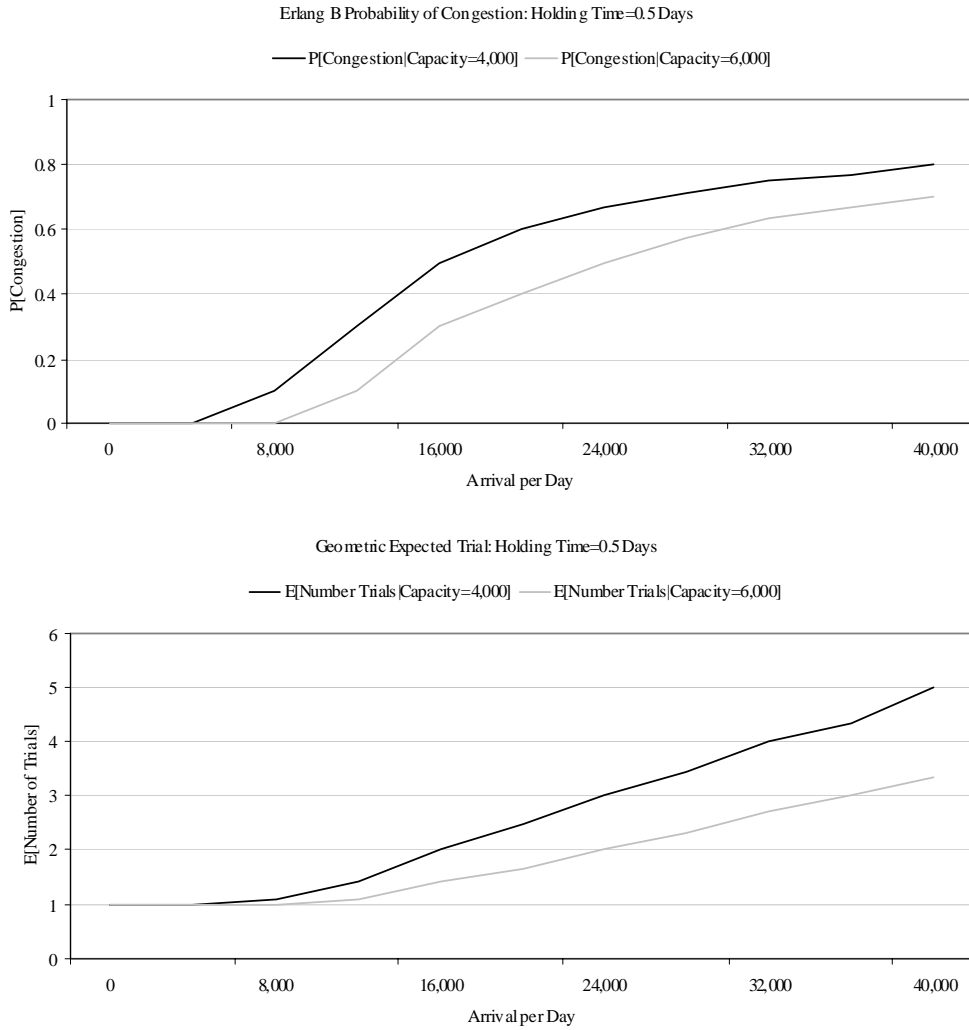
Figure 6 illustrates the results of this analysis. For any given arrival rate, it is clear that as capacity increases, both measures of congestion decrease. However, as the arrival rate increases, the same levels of congestion recur in the higher capacity network. Further, as noted in Section 5.1, the 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 would bring to the community may not justify the cost of the additional capacity, particularly given that 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 since additional users will add value to the network at a diminishing rate. 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 content database across

⁸ Note that in the data collected, the agent repeatedly tried to login 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.

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.

Figure 6: Illustration of Increasing Capacity on Login Congestion



5.4. Limitations and Generalizability

There are several limitations of our study that deserve further discussion. First, there are a variety of external factors that could have affected our results suggesting that free riding increased in larger networks. It is possible that the RIAA's legal actions against Napster during our time period caused an increase over time in the proportion of OpenNap users who free ride. It is also

possible that Napster users are more likely to free ride than OpenNap users and an influx of Napster users in our later time period is reflective of this. 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.

To the extent that any of these explanations hold it would weaken our conclusions with regard to increasing free riding in larger networks and our findings with regard to the relationship between replication, download attempts, and download time. It should, however, have no effect on the findings that availability is concave in network size, and login time and query time are convex in network size. Thus, our core finding that positive and negative network externalities limit the optimal size of P2P networks should still hold.

Further, we note that all of our results still hold if the regressions are run within each individual time period. Since 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.

It may seem surprising that the RIAA's legal actions against Napster would not significantly increase free riding among OpenNap users. However, it is important to note that the RIAA at this time was only targeting network operators with legal action, not network users. Additionally, P2P users during this time had strong reasons to believe that they would not be targeted.

Available evidence suggests that at the time the RIAA believed that targeting users would be a strategic mistake. For example, Kiron (2001, p. 1) notes that "the music industry, however, was not interested in suing Napster users, for they were also the music industry's customers."

The RIAA's decision to target individual users occurred in the summer of 2002, long after our data collection period. Even these efforts have been greeted with a great deal of skepticism. For example, Warner (2002, p. 115) writes that targeting individual users is "guaranteed to infuriate and alienate music fans, and it underscores the awful bind record labels are in."

Thus, individual OpenNap users during our study period seem to have every reason to believe that sharing files on OpenNap was a "safe" activity. OpenNap users had never been targeted, the RIAA seemed unwilling to target individual users (Napster or OpenNap). The only incident of individuals being targeted during this time frame, occurred in March 2000 (before our data collection started) when Metallica used the DMCA's "notice and take down" provision to have selected Napster user names removed from the system for illegally sharing the band's files. This action was widely viewed as a failure because users simply logged in with new user names and the action generated significant negative backlash against the band. To the extent that this action on the Napster network changed OpenNap user behavior, its effect would have been constant throughout our study period and would not induce spurious correlation with growth in the networks. However, we suspect that the fact that it was ineffective, not repeated by other bands, and not directed at OpenNap users, suggests that it likely had an insignificant impact on OpenNap user behavior. (We also note that no Metallica songs were included in the 170 random songs we tracked.)

Second, our results should be interpreted as applying in the context of centralized architectures. However, while analyzing positive and negative network externalities in decentralized and hybrid architectures would be a useful area for future research, the design of each of these architectures is consistent with our findings. As noted in Section 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 inter-peer or inter-ultrapeer 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 were searching for. The most 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 (i.e. a larger effective network size) does not justify the marginal cost their query would impose on the network. This interpretation, if correct, 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. P2P networks are increasingly being used in corporate settings (e.g., Deloitte and Touche's use of NextPage for knowledge management) and in non-music sharing setting (e.g., Virgin Record's use of Blue Falcon for streaming media), and the positive and negative network externalities in these settings may differ from our environment.

6. Discussion and Areas for Future Research

Using data collected from the six most popular OpenNap networks from December 2000 to April 2001 we find that additional users generated positive network externalities based on the quantity and uniqueness of the files they provide and negative network externalities in terms of login, query, download attempt and download speed congestion. Further, the marginal value of an additional user declined with the number of existing users while the marginal cost of congestion

imposed by users increased with network size. These findings imply that optimal network size is bounded for these centralized P2P networks.

We also find that the number of replicas of content per user decreases with network size. This is consistent with findings in the public economics literature that free riding worsens with group size and suggests that increased free riding in larger P2P network may be driven by the collective action of users in the private provision of public goods. While it is impossible to isolate this explanation from a decrease in replicas due to unobserved customer heterogeneity, an increase in free riding with increasing network size would imply that in larger networks autoreplication may fail to scale to meet demand and that over-consumption in these large networks can lead to congestion.

A business implication of this study is that the value of the P2P network does not scale in the way that traditional networks, such as telecommunication networks, do. The value present in telecommunications networks is a function of the number of users where marginal value is increasing in network size. In contrast, utility in P2P networks is based on collective content and congestion, and the marginal value of collective content is decreasing in the number of users while the marginal disutility from congestion is increasing in the number of users. Because of this, P2P networks are likely to have explicit limits on network size or implicit limits on network reach (by limiting the number of users each peer can access). These limits may in turn dampen the “winner-take-all” nature of competition — allowing multiple similar networks to exist side-by-side. This also suggests that network operators may wish to adopt niche strategies based on features or content to maximize the value provided to their share of network users.

A policy implication of our findings is that P2P networks, in their current stage, follow the economic theory of private provision of public goods. Free riding is prevalent in these networks and unless appropriate private incentives are implemented through managerial rules or pricing policies, the degree of free riding will limit the scalability and performance of P2P networks (Krishnan et al. 2003). From a technological perspective, these observations stress the importance of protocol designs that align private user incentives with the goals of the collective network.

This research can be extended in a variety of ways. First, our data suggest that free riding behavior is more pronounced in larger networks. Because free riding reduces the scalability and performance of P2P networks, it will be important for future research to analyze incentives structures necessary to increase resource provisioning in these networks. This analysis could build on Ba, Stallaert, and Whinston's (2001) recent findings regarding the importance of incentive alignment for Information Technology design. Second, while the design of decentralized and hybrid P2P architectures is consistent with our results regarding a limit on the optimal size of these networks due to positive and negative network externalities, future research should extend our analysis into other content domains such as streaming media or intra-enterprise knowledge management. , Finally, not all parties wish to see improved scalability in P2P networks. The RIAA, for example, is engaged in a variety of efforts to reduce the scalability and performance of these networks including raising the cost of sharing by filing lawsuits against individual users, and flooding the networks with "fake" content. Future research could analyze the effectiveness of these efforts in reducing network performance and scalability.

References

- Adar, E., B. A. Huberman. 2000. Free Riding on Gnutella. *First Monday: Peer-Reviewed Journal on the Internet*. 5(10). Retrieved August 16, 2001 from www.firstmonday.dk/issues/issue5_10/adar/index.html.
- 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, Angelique, Shane Greenstein, Marc Rysman. 2003. Coordination Versus Differentiation in a Standards War: The Adoption of 56K Modems. Working Paper, Boston University, Boston, MA.
- Avery, Simon. 2002. Hollywood Takes Aim At Online Film-Swappers. Associated Press, July 20.
- Ba, S., J. Stallaert, A.B. Whinston. 2001. Optimal Investment in Knowledge Within a Firm Using a Market Mechanism. *Management Science* 47(9) 1203-1219.
- Bhattacharjee, B., S. Chawathe, V. Gopalakrishnan, P. Keleher, B. Silaghi. 2003. "Efficient Peer-To-Peer Searches Using Result-Caching. Proceedings of ITPS2003.
- Brynjolfsson, Erik, Chris Kemerer. 1998. Network externalities in microcomputer software: An econometric analysis of the spreadsheet market. *Management Science* 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*. Information Systems Research (12:4), 2001, pp. 346-365.
- Cole, R., Y. Dodis, T. Roughgarden. 2003. Pricing Networks with Selfish Routing. Working Paper, New York University, New York, NY.
- Conner, Kathleen R. 1995. Obtaining strategic advantage from being imitated: When can encouraging "clones" pay? *Management Science* 41(2) 209-225.
- Constant, D., L. Sproull, S. Kiesler. 1996. The kindness of strangers: The usefulness of electronic weak ties for technical advice. *Organization Science* 7 119-135.
- Dranove, D., N. Gandal. 2003. The DVD vs. DIVX Standard War: Network Effects and Empirical Evidence of Preannouncements Effects. *Journal of Economics and Management Strategy* Forthcoming.
- Economides, Nicholas, Charles Himmelberg. 1995. Critical mass and network size with application to the U.S. fax market, 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, Gabel ed. *Product Standardization as a Tool of Competitive Strategy* North-Holland, New York, 1-22.
- Forman, Chris, Pei-Yu 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, Neil. 1994. Hedonic price indexes for spreadsheets and an empirical test for network externalities. *The RAND Journal of Economics* **25**(1) 160-170.
- Gandal, Neil, Michael Kende, Rafael Rob. 2000. The Dynamics of Technological Adoption in Hardware/Software Systems: The Case of Compact Disc Players, *RAND Journal of Economics* **31**: 43-61.
- Golle, P., K. Leyton-Brown, I. Mironov. 2001. Incentive for Sharing in Peer-to-Peer Networks. Working Paper, Computer Science Department, Stanford University, Palo Alto, CA.
- Goolsbee, Austan, Peter Klenow. 2003. Evidence on Learning and Network Externalities in the Diffusion of Home Computers. *Journal of Law and Economics* **40**(2) 317-344.
- Gu, Bin, Sirkka Jarvenpaa. 2003. Are Contributions to P2P Technical Forums Private or Public Goods. Working Paper, University of Texas at Austin, Austin, TX.
- Katz, M., C. Shapiro. 1986. Technology Adoption in the Presence of Network Externalities. *Journal of Political Economy* **94**(4) 822-841.
- Katz, M., C. Shapiro. 1994. Systems Competition and Network Effects. *Journal of Economic Perspectives* **8**(2) 93-115.
- Kauffman, Robert J., James McAndrews, Yu-Ming Wang. 2000. Opening the “black box” of network adoption. *Information Systems Research* **11**(1) 61-82.
- Kirk, P. 2003. Gnutella Protocol Development — An Overview of 0.6. Protocol Specification, available at <http://rfc-gnutella.sourceforge.net/rfcGnutella/index.html> (July 5, 2003).
- Kiron, David. 2001. Napster. Harvard Business School Case Number 9-801-219.
- Kraut, R. E., P. Attewell. 1993. Electronic mail and organizational knowledge. Working paper, Carnegie Mellon University, Pittsburgh, PA.
- Krishnan, Ramayya, Michael D. Smith, Zhulei Tang, Rahul Telang. 2003. The Virtual Commons: Why Free riding Can Be Tolerated in File Sharing Networks.

- 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.
- Parameswaran, Manoj, Anjana Susarla, Andrew B. Whinston. 2001. P2P Networking: An Information Sharing Alternative. *IEEE Computer*, July, 1-8.
- Park, Sangin. 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, Garth, Andrea Shepard. 1995. Adoption of Technologies with Network Effects: An Empirical Examination of the Adoption of Automated Teller Machines. *RAND Journal of Economics* **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 Review of Economics and Statistics* **36**(4) 387-389.
- Sripanidkulchai, K., B. Maggs, H. Zhang. 2001. Efficient Content Location and Retrieval in Peer-to-Peer Systems by Exploiting Locality in Interests. Proceedings of SIGCOMM'01.
- Stoica, I., R. Morris, D. Karger, M.F. Kaashoek, H. Balakrishnan. 2001. Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications. Proceedings of SIGCOMM'01.
- Strahilevitz, L.J. 2002. Charismatic Code, Social Norms, and the Emergence of Cooperation on the File-Swapping Networks. John M. Olin Law and Economics Working Paper, Number 162, University of Chicago, Chicago, IL, September.
- 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, Hal. 2003. Economics of Information Technology. Working Paper, University of California at Berkeley, Berkeley, CA.
- Wackerly, Dennis, William Mendenhall, Richard Scheaffer. 2001. *Mathematical Statistics with Applications*. 6th Ed. Duxbury Press, Florence, KY.
- Warner, Melanie. 2002. The New Napsters. *Fortune*. August 12.

Wasko, Molly M., Samer Faraj. 2000. Is it what one does: Why people participate and help others in electronic communities of practice. *Journal of Strategic Information Systems* **9** 155-173.