NET Institute*

www.NETinst.org

Working Paper #06-23

October 2006

Bandwidth Allocation in Peer-to-Peer File Sharing Networks

Albert Creus Mir, Universitat Politecnica de Catalunya
Ramon Casadesus-Masanell, Harvard Business School
Andres Hervas-Drane, Universitat Autonoma de Barcelona

* The Networks, Electronic Commerce, and Telecommunications ("NET") Institute, http://www.NETinst.org, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, "virtual networks" comprised of computers that share the same technical standard or operating system, and on network issues in general.

Bandwidth Allocation in Peer-to-Peer File Sharing Networks*

Albert Creus Mir† Ramon Casadesus-Masanell‡ Andres Hervas-Drane§

October 2, 2006

Abstract

We present a model of bandwidth allocation in a stylized peer-to-peer file sharing network. Given an arbitrary population of peers composed of sharers and freeriders, where all peers interconnect to maximize their allocated bandwidth, we derive the expected bandwidth obtained by sharers and freeriders. We show that sharers are always better off than freeriders and that the difference decreases as the size of the network grows. This paper constitutes a first step towards providing a general analytical foundation for resource allocation in peer-to-peer networks.

Keywords: Peer-to-Peer, Network formation, Resource allocation, Congestion

1 Introduction

Paired with the widespread adoption of peer-to-peer (p2p) networks, a growing academic literature has emerged in recent years. Research has mainly focused on file sharing networks, the most popular and disruptive application of peer-to-peer technology. The advent of file sharing has not only had a strong impact on the content industry, forcing a change to new business models, but also on the data networks of telecommunications operators worldwide. The copious amount of data traffic generated by file sharing has shaped an increasing demand for bandwidth. It is currently estimated that file sharing accounts for over 50 per cent of total Internet traffic.

Peer-to-peer applications configure overlay networks over the physical network layer. The structure and bottlenecks of these overlay networks depend critically on the activity of participating peers. Several empirical studies have examined their topology, employing different techniques to gather information. Network crawlers are deployed in [5], [14] and [15]. While in [10] and [16], p2p traffic is recovered from a trace performed at the physical network's backbone. Together, both approaches have helped characterize the main properties of these networks with respect to peer availability and activity patterns, data traffic paths, and bandwidth bottlenecks.

A parallel literature has focused on constructing theoretical models of p2p file sharing. Contributions to this literature have emerged from both computer science ([3], [8], [9]) and economics

*Casadesus-Masanell and Hervas-Drane are grateful for financial support from the HBS Division of Research, the *NET Institute* (http://www.NETinst.org), Fundación BBVA and the FPU program of the Spanish Ministry of Education and Science.

†Universitat Politècnica de Catalunya. (acreus@lsi.upc.edu)

‡Harvard Business School. (casadesus@gmail.com)

§Universitat Autònoma de Barcelona. (andresonline@gmail.com)

([1], [6], [13]). This work attempts to explain stylized properties of file sharing networks by studying the incentives of peers to contribute resources. The literature has focussed on the effects of freeriding: peers that participate in p2p networks to enjoy resources contributed by others without contributing in return. Several papers have explored mechanisms to reduce freeriding and drive higher resource contribution levels.

This theoretical literature, however, has so far failed to incorporate insights deriving from the structure of the network. With just a few exceptions (see, for instance, [12]), the structure of p2p networks is generally not modeled, resulting in resource allocation rules lacking a foundation. In our view, p2p research can benefit from explicitly considering the properties emerging from the underlying network structure. One main difficulty in pursuing this approach is the lack of analytical tractability of network games. In this paper we present a simple approximation of bandwidth allocation in file sharing networks which can be readily applied. While our model is stylized, we believe that it captures the essential features of real p2p networks and provides a foundation for theoretical models of file sharing.

We consider an arbitrary p2p file sharing network. We take the number of sharers and freeriders as given. We assume that peers have homogeneous bandwidth capacities, that content offered by sharers is equally valuable, and that peers derive positive utility from bandwidth. Thus, the only reason why peers enjoy different utility levels in our model is heterogeneous access to bandwidth contributed to the network.[1]

Our stylized model of p2p assumes that:

1. Each sharer provides one unit of upload bandwidth;
2. All peers have at least one unit of download bandwidth capacity;
3. Every peer connects to one sharer only;
4. A sharer may *not* connect to herself;
5. Bandwidth obtained from a sharer is allocated equably amongst all peers connected to her.

We refer to a set of links connecting peers to sharers as an *allocation*. A *stable* allocation is one where no peer can be made strictly better off by connecting to a different sharer. Otherwise, if given the choice, at least one peer would prefer to update her link and connect to a different sharer. We assume all stable allocations arise in the network with equal probability. In any stable allocation, every sharer will have approximately the same number of peers connected to her. By 'approximately' we refer to the cases in which the number of peers is not divisible by the number of sharers; the number of peers connected to each sharer will then differ by one peer. If the total bandwidth provision in the network could be aggregated and then shared by all peers, then each peer would obtain the exact same bandwidth (#sharers/#peers). However, because bandwidth has to be allocated discretely (sharer-by-sharer), when the number of peers is not divisible by the number of sharers there are two possible amounts of bandwidth that each peer may end up obtaining in a stable allocation.

[1] We do not explicitly model the impact of overhead traffic originated by the p2p protocol and lower level communication protocols. Thus, we implicitly assume that these protocols have and equivalent impact across the population of peers. See [7] for a technical discussion on this issue. Although sharing in the network certainly generates more overhead traffic than freeriding, the magnitude of this overhead is generally small in relation to the volume of data traffic.

We derive an exact formula for the expected bandwidth obtained by sharers and freeriders, where expectations are taken across *all* stable allocations. We show that, somewhat counterintuitively, the expected bandwidth obtained by sharers is always larger than that available to freeriders. Sharers can be allocated to fewer sources as they face the constraint of not connecting to themselves. The constraint captures the fact that sharers in file sharing networks obtain no additional utility from accessing the content they share (their bandwidth) as they already own it.[2] Freeriders, on the other hand, have access to more sources as they can connect to every sharer. As it turns out, the constraint faced by sharers ends up favoring them (in expected terms).[3]

Furthermore, we show that #sharers/#peers is a good approximation to the expected bandwidth obtained by both sharers and freeriders. Sharers (freeriders) always obtain expected bandwidth larger than (smaller than) #sharers/#peers. And as the size of the network grows, the difference between expected bandwidth and #sharers/#peers quickly decreases. Already in a network of size 10, the expected bandwidth obtained by sharers and freeriders differs from #sharers/#peers by, at most, 10^{-4}. And when network size is 100, the difference is always less than 10^{-6}.

Because the approximation has an exceedingly simple form, it provides a foundation for applied theoretical work on p2p. The exact formula, on the other hand, is discouragingly complex and it is unusable for all practical purposes. Casadesus-Masanell and Hervas-Drane [4], for example, use the approximation to construct a model of a p2p file sharing network with endogenous sharing that competes against a for-profit firm that offers content on a client-server architecture at positive prices. By incorporating the costs associated with file sharing together with the foundation for bandwidth allocation presented here, the authors show that such a model can explain important stylized facts identified in the literature. While the present paper is devoted to the study of bandwidth allocation in p2p file sharing networks, we believe our approach is quite general and the results can be extended to other scarce rival resources in p2p networks.

The remainder of the paper is organized as follows. Section 2 states the problem formally. In Section 3 we introduce two reductions and proceed to count all the stable allocations. We present the formula in Section 4 and discuss some properties. Section 5 provides graphical representations. We present intuition and discuss limitations in Section 6.

2 Setup

Let N, S, and F, be the set of all peers, the set of sharers (peers that contribute bandwidth), and the set of freeriders (peers that do not contribute) respectively. Let n, s, and f, be the cardinalities of these sets. Every peer is either a sharer or a freerider. That is, $F = N \backslash S$ and $f = n - s$.

[2]This effect is similar to the contribution of other rival resources in p2p applications beside file sharing, such as storage space or computational power. As any rival resources contributed cannot be put to an alternative use, they are of no direct utility to the sharer.

[3]The difference that we obtain is small. Nevertheless, the result may have implications for dynamic systems with unstable equilibria where initial, small differences may be amplified over time. The result may also be exploited in the design of networks if our assumptions can be artificially implemented in the system.

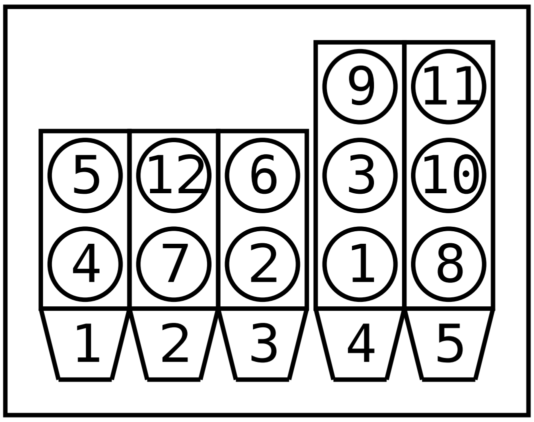


Figure 1. We can represent an allocation graphically in terms of urns and balls. Every urn represents a sharer, every ball is a peer. Urns and balls are numbered, but the order of balls in urns is unimportant.

Recall that a *stable* allocation is one where no peer can be made strictly better off by re-allocating her to another sharer. The figure illustrates such an allocation. Obviously, given N and S there are multiple stable allocations. And whenever the number of peers is not divisible by the number of sharers, the bandwidth obtained from the network will differ across peers. Let $b := s - (n \bmod s)$, and $a := n \operatorname{div} s$. In every stable allocation we must have a peers allocated to b sharers and the rest $(a + 1)$ allocated to the remaining $s - b$ sharers. Therefore, $n = ba + (s - b)(a + 1)$. The following figure illustrates these zones.

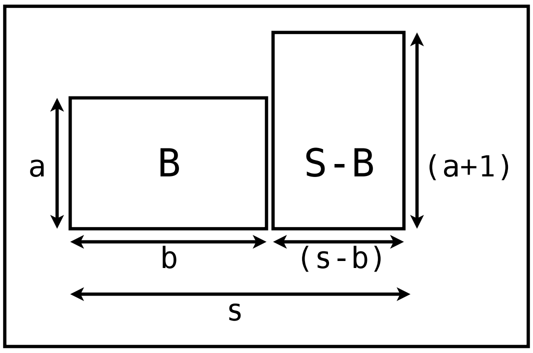


Figure 2. Relevant zones.

To compute the *expected* bandwidth obtained by freeriders and sharers, we begin by computing the bandwidth obtained by each peer in every stable allocation. Let x be a peer, and let $G(x)$ be the bandwidth obtained by the peer in any given allocation. $G(x)$ is a random variable that can take two values only, $1/(a + 1)$ and $1/a$. For any given stable allocation, we define the set B of 'fortunate' peers as $B = \{x|G(x) = 1/a\}$ and the set of 'unfortunate' peers as $S - B = \{x|G(x) = 1/(a + 1)\}$. Notice that the specific peers in B and $S - B$ depend on the particular allocation under consideration.

The *expected* bandwidth obtained across all peers, both sharers and freeriders, is $E := E(G(x)) = s/n$. Obviously, $1/(a + 1) < s/n \leq 1/a$. We may also consider the conditional expectations $E_S := E(G(x)|x \in S)$ and $E_F := E(G(x)|x \in F)$ (these are the expected bandwidths obtained by sharers and freeriders, respectively). If sharers were to connect to themselves the calculation would be trivial. In this case, symmetry implies that $E_S = E_F = E = s/n$.

In peer-to-peer networks, however, sharers do not connect to themselves. In this case, the conditional expectations E_S and E_F turn out to be different. Nevertheless, the total bandwidth available in the network, s, must still be equal to the number of sharers times E_S plus the number of freeriders times E_F. That is, $s/n \cdot E_S + (n-s)/n \cdot E_F = s/n$. Therefore, if we compute E_S, we can immediately obtain E_F.

Because every stable allocation is assumed to be equiprobable, to compute E_S we need to count all of them and compute the average of $G(x)_{x \in S}$ for each. Let H be the total number of stable allocations. Let h_i be the number of these allocations with i sharers in B. Notice that for all these allocations, the average of $G(x)_{x \in S}$ is exactly the same. Therefore, we can consider h_i as the histogram of this value. Then,

$$E_S = \sum_{i=0}^{s} \left(\frac{1}{a}\frac{i}{s} + \frac{1}{a+1}\frac{s-i}{s} \right) \frac{h_i}{H}$$

or

$$E_S = \frac{a + \sum_{i=0}^{s} \frac{i}{s}\frac{h_i}{H}}{a(a+1)}.$$

3 Count

In this section we compute h_i. That is, we orderly enumerate and count all the stable allocations that have i sharers in B. Due to the magnitude of this task, we reduce and decompose the problem as much as possible.

3.1 First reduction

Given n and s, consider the set of all stable allocations. We can divide this set into subsets depending on which sharers support B (having only a peers allocated to them). These subsets are disjoint, and they add up to the entire set of stable allocations, so they are classes. Specifically, there are $\binom{s}{b}$ classes.

Each of these classes has the same number of allocations. To see this, notice that there are bijections between the classes, obtained by changing the names of the elements. Consider class r and let h_i^r be the number of allocations with i sharers in B for class r. Notice that for all r, the value of h_i^r (for every i) is the same. Therefore, by studying one single allocation we can obtain the total $h_i = \binom{s}{b} h_i^r$. Thus, to simplify the analysis we, will study the class in which the sharers supporting B are the first ones. The next figure illustrates the approach.

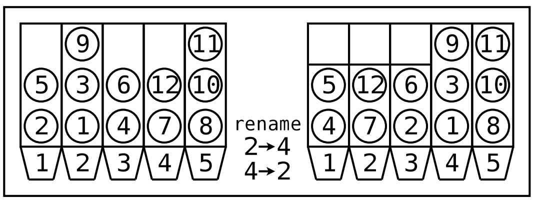


Figure 3. The first reduction is to rename sharers to get a clearer view.

3.2 Second reduction

Given n and s, consider the set of all stable allocations where B is supported by the first sharers. We now compute how many ways we can assign the n peers between the two zones B and $S - B$.

3.2.1 Counting reduction 2

In general, there are $\binom{n}{ba}$ possibilities. There are three exceptions:

1. When $b = s$ there is 1 possibility.
2. When $b = 1$ there are $\binom{n-1}{ba}$ possibilities.
3. When $b = s - 1$ there are $\binom{n-1}{ba-1}$ possibilities.

These three cases will always be exceptions to the general formula we derive. They are illustrated in the following figure. The last two cases are special because they require that a given sharer be assigned to 'the other' zone.

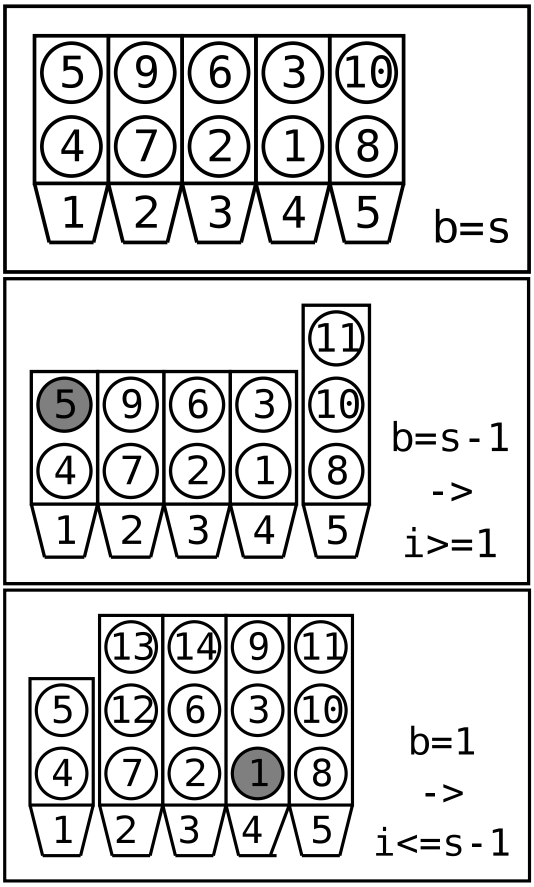


Figure 4. The three exceptions.

In the general case, $1 < b < s - 1$, there are $\binom{n-s}{ba-i}\binom{s}{i}$ possibilities that have exactly i sharers in B. Using known formulae, adding up we obtain

$$\binom{n}{ba} = \sum_{i=\max\{0,s+ba-n\}}^{\min\{s,ba\}} \binom{n-s}{ba-i}\binom{s}{i},$$

where the limits come from observing the problem and ensuring the combinatorial numbers are well defined.

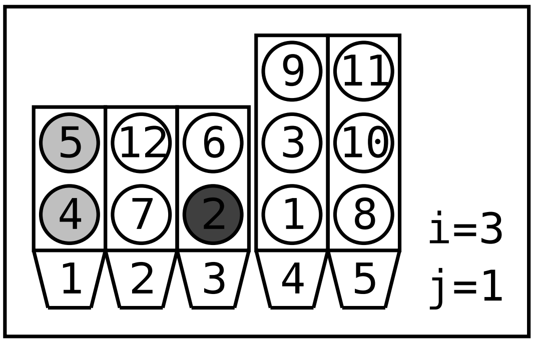


Figure 5. In zone B, freeriders are plotted in white, easy sharers in grey and complicated sharers in black.

As illustrated in the figure above, j sharers out of i may be 'complicated.' A complicated sharer is one that supports B and happens to be in B. We have to take this into account in order to not assign a complicated sharer to herself. An allocation that has a sharer assigned to herself is called a 'coincidence.' There are $\binom{n-s}{ba-i}\binom{s-b}{i-j}\binom{b}{j}$ possibilities that contain exactly j 'complicated' sharers. Again using known formulae, we add up and obtain

$$\binom{n}{ba} = \sum_{i=\max\{0,s+ba-n\}}^{\min\{s,ba\}} \binom{n-s}{ba-i} \sum_{i=\max\{0,b+i-s\}}^{\min\{b,i\}} \binom{s-b}{i-j}\binom{b}{j}.$$

3.2.2 Relationship with the original problem

For every possibility of the reduced problem, we now count how many possibilities exist in the original problem. We proceed by assigning the peers in B to the b sharers that support them.

If there were no 'complicated' sharers, this would be a simple multinomial problem with formula $\frac{(ab)!}{(a!)^b}$. However, because there are j complicated sharers, we perform the same count but need to subtract all those possibilities with coincidences. For example, if $j = 1$,

$$\begin{aligned}\frac{(ab)!}{(a!)^b} - \frac{(ab-1)!}{(a-1)!(a!)^{b-1}} &= \frac{(ab)!}{(a!)^b} - \frac{a}{ab}\frac{(ab)!}{(a!)^b} \\ &= \frac{(ab)!}{(a!)^b}\left(1 - \frac{a}{ab}\right).\end{aligned}$$

If $j = 2$, we subtract twice the cases with one coincidence, but add once the cases with two coincidences,

$$\frac{(ab)!}{(a!)^b} - 2\frac{(ab-1)!}{(a-1)^1!(a!)^{b-1}} + 1\frac{(ab-2)!}{(a-1)!^2(a!)^{b-2}} = \frac{(ab)!}{(a!)^b}\left(1 - 2\frac{a^1}{(ab)} + 1\frac{a^2}{(ab)(ab-1)}\right)$$
$$= \frac{(ab)!}{(a!)^b}\binom{2}{0}\frac{a^0(ab-0)!}{(ab)!} - \frac{(ab)!}{(a!)^b}\binom{2}{1}\frac{a^1(ab-1)!}{(ab)!} + \frac{(ab)!}{(a!)^b}\binom{2}{2}\frac{a^2(ab-2)!}{(ab)!}.$$

In this process, we are applying the principle of inclusion-exclusion. The problem is similar to a *derangement*. By analogy, we define the *Multinomial Derangement number*,

$$MD(b,a) = \frac{(ab)!}{(a!)^b}\sum_{k=0}^{b}\binom{b}{k}(-1)^k a^k \frac{(ab-k)!}{(ab)!},$$

and the *Generalized Multinomial Derangement number*,

$$GMD(b,a,j) = \frac{(ab)!}{(a!)^b}\sum_{k=0}^{j}\binom{j}{k}(-1)^k a^k \frac{(ab-k)!}{(ab)!}.$$

Notice that $MD(b,a) = GMD(b,a,b)$, and $MD(b,1) = Derangements(b) = !b$.

As we have to take into account both zones, B and $S-B$, for each possibility of the reduced problem there are $GMD(b,a,j)GMD(s-b,a+1,(s-b)-(i-j))$ possibilities of the original problem.

4 Formula

Consider the original problem described in Section 2. Let n be the number of peers, s the number of sharers, $a = n \operatorname{div} s$ and $b = s - (n \bmod s)$.

0) For $b = s$, we have that $E_S = 1/a = s/n$. Otherwise, if $H := \sum_i h_i$,

$$E_S = \frac{a + \sum_{i=0}^{s} \frac{i}{s}\frac{h_i}{H}}{a(a+1)}.$$

1) For $b = 1$, $i \in [\max\{0, s+ba-n\}, \min\{s-1, ba\}]$.
2) For $b = s-1$, $i \in [\max\{1, s+ba-n\}, \min\{s, ba\}]$.
3) Otherwise, $1 < b < s-1$, $i \in [\max\{0, s+ba-n\}, \min\{s, ba\}]$.

Where

$$h_i = \binom{n-s}{ba-i}\sum_{j=\max\{0,b+i-s\}}^{\min\{b,i\}}\binom{s-b}{i-j}\binom{b}{j}GMD(b,a,j)\cdot GMD(s-b,a+1,(s-b)-(i-j)),$$

and

$$GMD(b,a,j) = \frac{(ab)!}{(a!)^b}\sum_{k=0}^{j}\binom{j}{k}(-1)^k a^k \frac{(ab-k)!}{(ab)!}.$$

4.1 Properties

If sharers can connect to themselves we obtain $E_S = s/n$. To see this, notice that in this case the $GDM()$ factors disappear and

$$\sum_i \frac{i}{s}\frac{h_i}{H} = \frac{\sum_i \frac{i}{s}\binom{n-s}{ba-i}\binom{s}{i}}{\binom{n}{ba}} = \frac{\sum_i \binom{(n-1)-(s-1)}{(ba-1)-(i-1)}\binom{(s-1)}{(i-1)}}{\binom{n}{ba}} = \frac{\binom{n-1}{ba-1}}{\binom{n}{ba}} = \frac{ab}{n}.$$

Therefore, the difference is given by the $GMD()$ factors.

Also note that $(ab)!/(a!)^b$ is not relevant within $GDM()$. Because this expression does not depend on i or j, it also appears in $H := \sum_i h_i$. Therefore it cancels out.

Two efficient ways to compute $GMD()$ recursively are as follows. Let

$$d_k := \binom{j}{k}(-1)^k a^k \frac{(ab-k)!}{(ab)!}$$
$$= \{\frac{(j-0)(j-1)...(j-k+1)}{12...k}\}\{(-1)(-1)...(-1)\}\{aa...a\}\{\frac{11...1}{(ba-0)(ba-1)...(ba-k+1)}\},$$

$$f_k := \frac{(-1)(j-k+1)(a)}{(k)(ba-k+1)},\ f_0 := 1,$$

$$d_k := f_k d_{k-1},\ d_0 = 1,$$

and

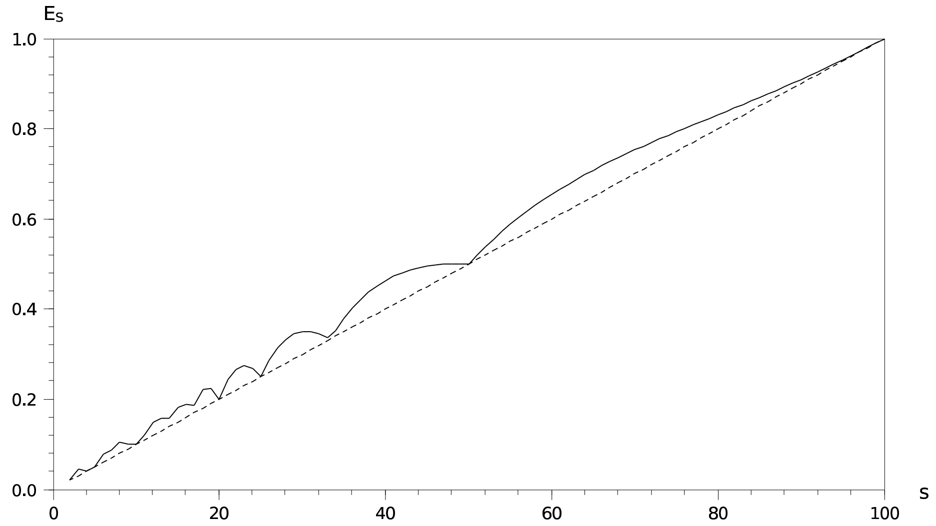
$$g_k := f_{j-k}(1+g_{k-1}),\ g_0 := f_j.$$

Then,

$$GMD(b,a,j) = \frac{(ab)!}{(a!)^b}\sum_{k=0}^{j} d_k = \frac{(ab)!}{(a!)^b} g_j.$$

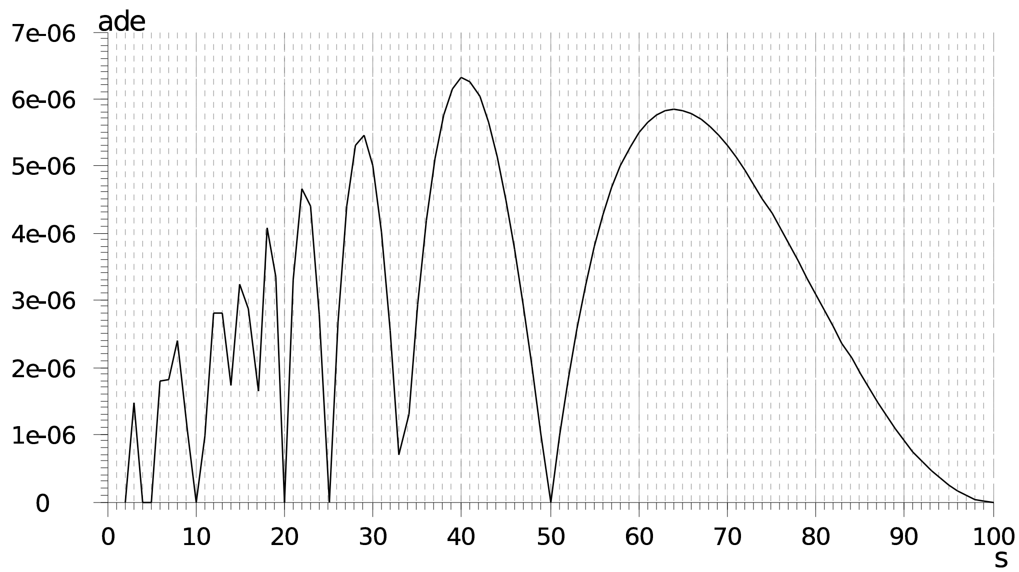
5 Plots

In this section we present a few plots to illustrate the properties of E_S, the expected bandwidth obtained by sharers. We initially fix the number of peers to $n = 100$, increase the number of sharers from $s = 2$ to 100, and plot E_S. The plot reveals that $E_S \geq s/n$ (the diagonal line). E_S is a curve with several peaks always above s/n. Of course, $E_S > s/n$ implies that $E_F < s/n$. This follows from the fact that $s/n \cdot E_S + (n-s)/n \cdot E_F = s/n$. The difference between E_S and s/n is small. In the following plot we have augmented 10,000 times the difference to make it visible.



Plot 1. $E_S \geq s/n$

We now plot the absolute difference, $ade = E_S - s/n$. We should note that, in order to generate a statistical plot of ade by the Montecarlo method, one needs to be especially careful not to end up with a biased generator. A full 3D plot is included in the appendix. Although a plot of the relative difference in percentage terms between E_S and s/n could be more informative, such a curve has peaks that vary wildly in size and cannot be drawn well.

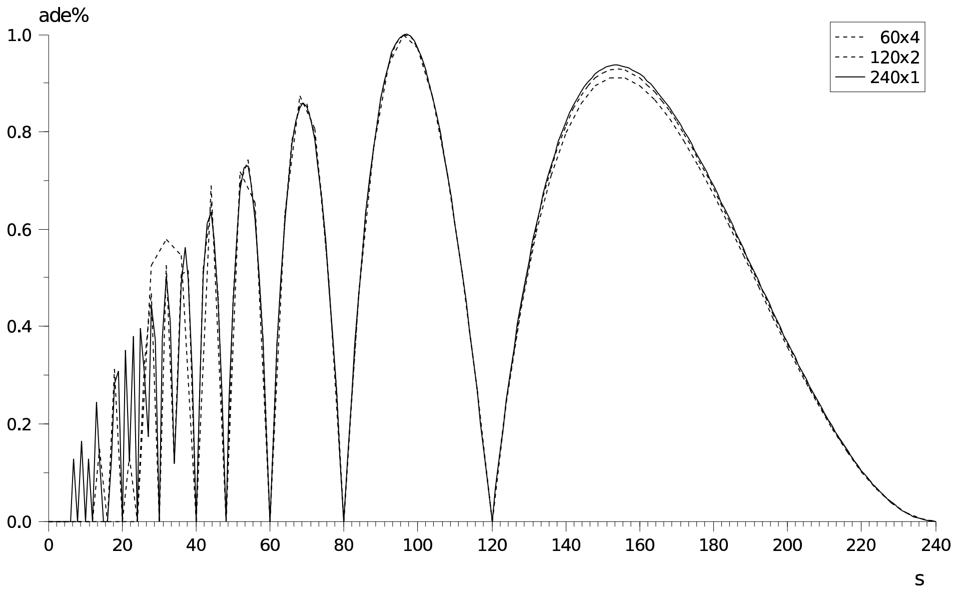


Plot 2. A reference plot of ade.

As we increase n, the differences between E_S and s/n maintains a similar pattern, only the magnitudes differ. The next plot shows ade for different values of n ($n = 60$, $n = 120$, and

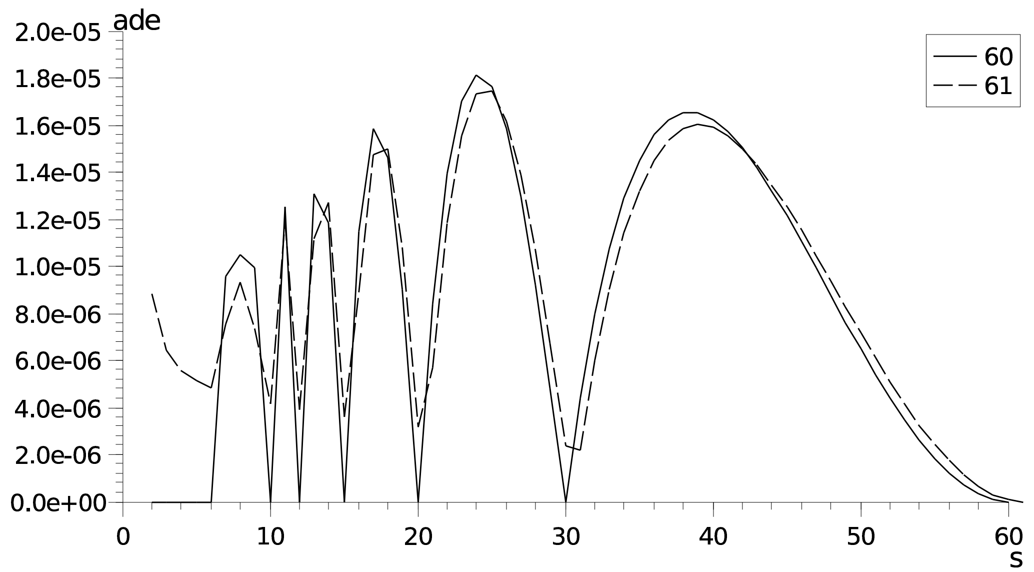
$n = 240$), normalized in order for the largest peak to reach 1. We have also stretched the horizontal axis. For example, the plot for $n = 60$ has its horizontal axis stretched 4 times. And the plot for $n = 120$ has its horizontal axis stretched twice. The similarity between the three plots is remarkable.

.



Plot 3. The *ade* curve has almost the same shape everywere.

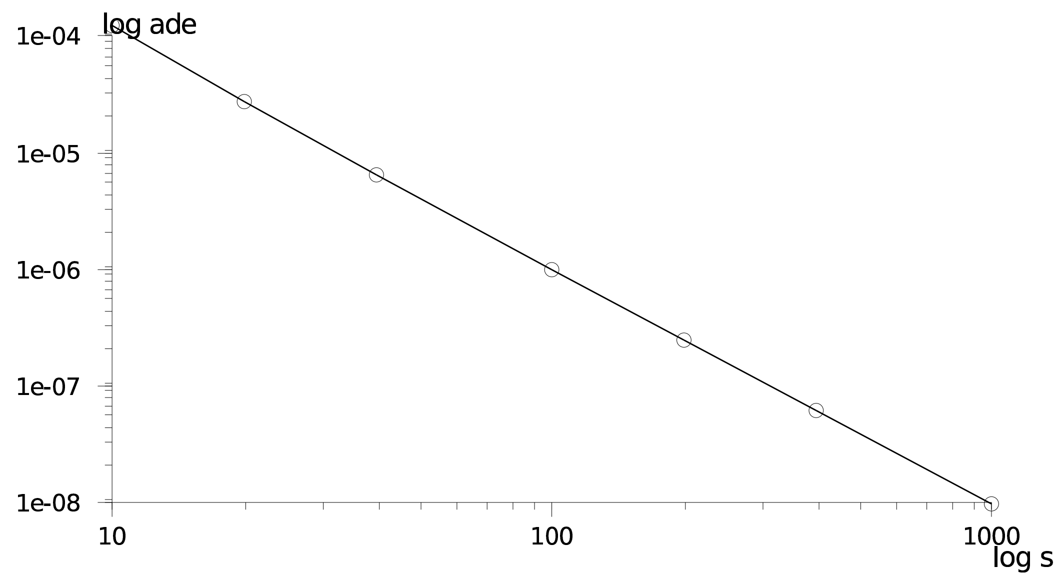
As n changes, the most pronounced differences in the shape of *ade* occur at the values of s where *ade* approaches zero ($n/2$, $n/3$, $n/4$,...). If n is divisible by s, then *ade* is equal to zero. Otherwise, $ade > 0$. The following plot shows *ade* for $n = 60$ and $n = 61$. 60 is divisible by 2, 3, 4, 5, 6, 10, 12, 15, 20, and 30. At all these points *ade* is zero. 61, however, is prime. In this case, *ade* never reaches zero.



Plot 4. The biggest differences are on the zeros.

The fact that *ade* is similar for all n allows us to tabulate *ade* for a given n and then extrapolate its value for other n. In this way, we can obtain approximate values of *ade* without costly calculations.

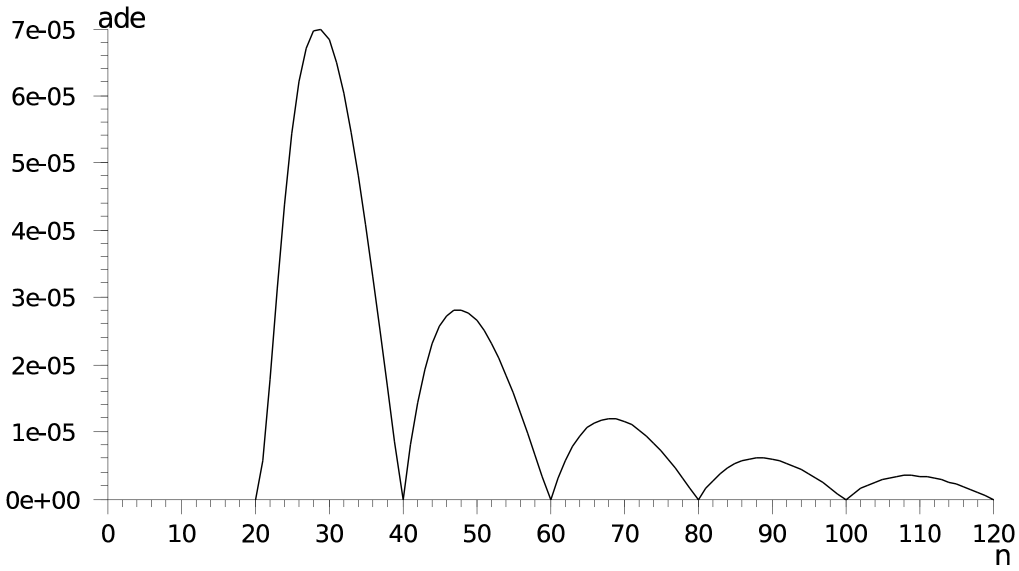
It is of interest to evaluate how *ade* varies as we increase n; or how the bandwidth difference between sharers and freeriders evolves as the size of the network increases. We next plot the values of *ade* at the last peak as we increase both n and s, thus maintaining the proportion of sharers present.



Plot 5. The *ade* curve's last peak decreases at an exponential rate of -2 as both new sharers and freeriders arrive.

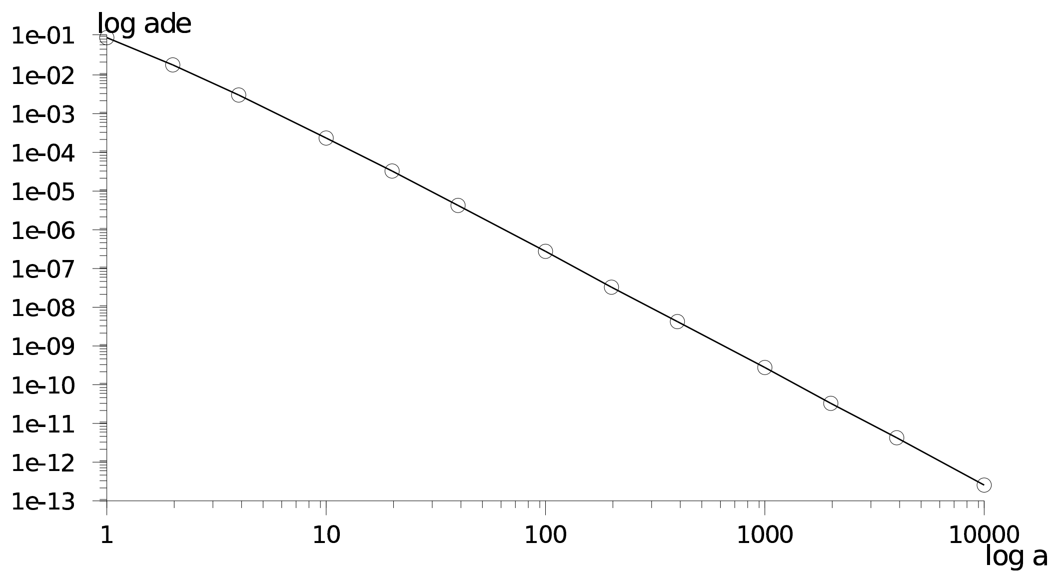
The plot shows that *ade* decreases at a quadratic rate. This suggests a quick method to approximate the value of *ade* given s and n: look up $s_{100} = s/n\,100$ in the second plot, and apply the formula $ade = ade_{100}(100/n)^2$.

It is also interesting to see how *ade* evolves for a given s as n increases. In the following plot, we set $s = 20$ and let n vary from 20 to 120.



Plot 6. The *ade* curve when only freeriders arrive.

The next plot shows the value of *ade* at the last peak for this case. The decrease rate is also exponential, but even larger with a -3 exponent.



Plot 7. The *ade* curve's last peak decreases at an exponential rate of -3 as new freeriders arrive.

These results suggest that *ade* quickly converges to 0 as n increases.

6 Discussion

6.1 Connectivity constraints

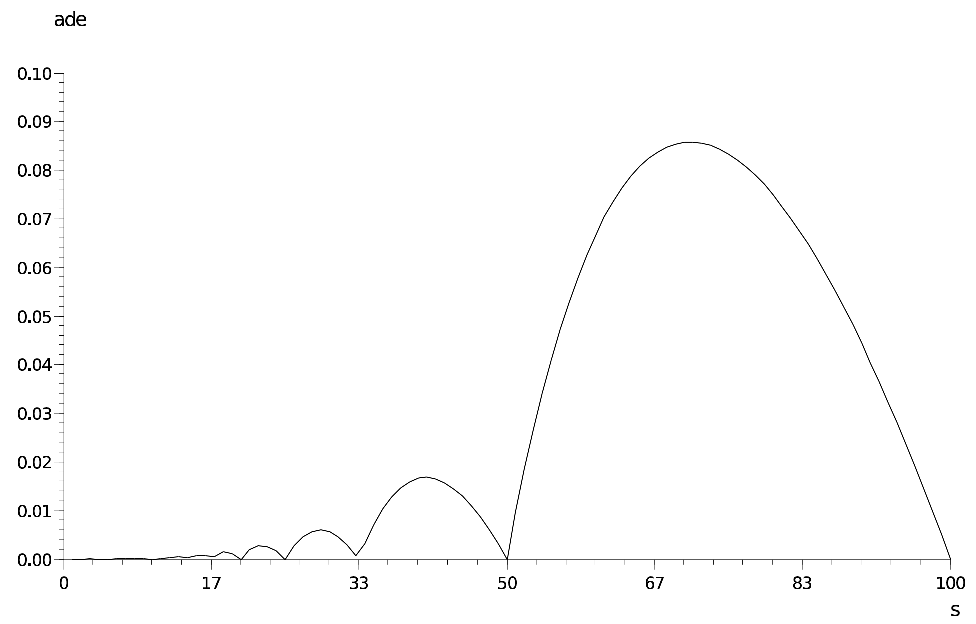
As we have just shown, the difference in expectations E_S and E_F is due to the constraint that a sharer cannot connect to herself. Furthermore, not only $E_S \neq E_F$, but $E_S \geq E_F$. The constraint 'helps' sharers; they end up better off (in expected terms). Counterintuitively, peers with more options available (freeriders) are worse off than those with less options (sharers). Technically, the result suggests that if freeriders could commit not to be allocated to a given sharer, they would do so.

Going into this problem, we expected that individuals who had less options available (sharers) would end up worse off than those with more options available (freeriders). To gain some intuition, consider the following heuristic argument. Suppose that there are s sharers and f freeriders. Let E_S^0 and E_F^0 be the expected amount of good allocated to sharers and freeriders, respectively, when no sharer can be allocated to his own good. (This has been the assumption in most of the paper. In fact, $E_S^0 \equiv E_S$ and $E_F^0 \equiv E_F$.) Suppose now that one of the sharers *can* be allocated to his own good. Let E_F^1 be the expected amount of good allocated to freeriders in this new situation. Clearly, $E_F^1 = (1-\alpha) \cdot E_F^0 + \alpha \cdot E_F^N$, where E_F^N is the expected amount of good allocated to freeriders in the 'new' equilibria (all those equilibria where the 'special' sharer is allocated to his own good) and α is the proportion of 'new' equilibria to all equilibria. It is easy to see that $E_F^N > E_F^0$. The reason is that having a sharer allocated to himself (the new equilibria) means that (on average) there is less competition for all the goods offered by the other sharers, although there is slightly more competition for the one good contributed by the 'special' sharer. As a consequence, we must have $E_F^1 > E_F^0$. Applying this argument s times we see that $E_F^s > E_F^{s-1} > E_F^{s-2} > \ldots > E_F^1 > E_F^0$. Notice finally that when every sharer can be allocated to its own good, sharers and freeriders are symmetric and $E_F^s = E_S^s = s/n$. Of course, this implies that $E_F \equiv E_F^0 < s/n$ and $E_S > s/n$.

Consider the following modified model where each sharer can *only* be connected to herself. Notice that the constraint is now stronger than before; sharers have only one feasible link and connections to other sharers are no longer available. The analysis is immediate,

$$E_S = \frac{1}{s}(b\frac{1}{a} + (s-b)\frac{1}{a+1}) = \frac{1}{s}\frac{sa+b}{a(a+1)}.$$

The following plot illustrates the behavior of *ade* in the modified model.



Plot 8. The *ade* curve when sharers can only connect to themselves.

As shown by the graph, $ade > 0$. And in fact, the difference between E_S and s/n is now even larger. This example shows that the more constrained peers are, the better off they end up.

6.2 Equiprobability

We have assumed throughout that stable allocations are equiprobable. We next motivate this assumption in the context of a model where peers decide with whom to connect to. The nature of p2p applications suggests one-sided link formation, where peers can decide which sharer to connect to without the consent of the sharer.[4] In this setting, a simultaneous one-shot game yields a set of equilibria that coincides by definition with our set of stable allocations. Clearly, this is the set of allocations that are of interest for the analysis.[5]

A one-shot game, however, provides no insight on the relative probability of each outcome. To construct a probability distribution over this set we need to consider a sequential game, where peers decide orderly with whom to connect to. To model such a game, consider a randomized connecting order with myopic peers. That is, peers take the current allocation as given when choosing their connection; no forward induction takes place (which is unfeasible given the size and complexity of p2p networks). In the model described, however, the probability distribution over the set of stable allocations depends on the fine details of the connection process.

Consider the following example. There are six peers, three sharers (S1, S2 and S3) and three freeriders (F1, F2 and F3). The random ordering of peers is S3, S2, F1, F2, F3, and S1. S3

[4] See Jackson [11] for a survey of models of networks in Economics and a discussion of different models of link formation.

[5] We should note that in our model no tradeoff exists between stability and efficiency. An allocation is Pareto efficient if and only if the upload bandwidth provided by *all* sharers is being utilized in the network. That is, if the whole bandwidth provision in the network is enjoyed by peers. Clearly, if an allocation does not satisfy this condition, a peer reassigned to a free sharer can be made better off without worsening the remaining peers. Hence stable allocations are a subset of Pareto allocations.

connects to S2; S2 connects to S3; F1 connects to S1; F2 connects to S2; and F3 connects to S3. At this point, S1 can only connect to S2 or S3. But neither constitutes a stable allocation.

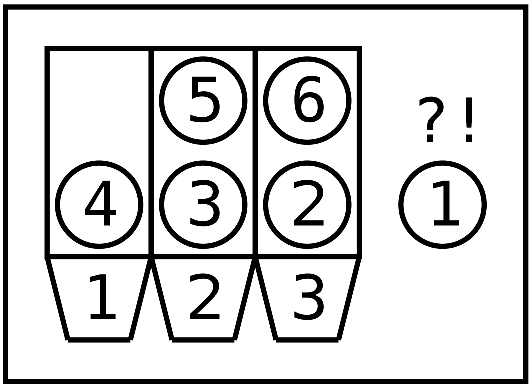


Figure 6. Sequential connecting orderings with myopic peers may yield unstable allocations.

It turns out that constraining sharers not to connect to themselves while demanding equiprobability is somewhat equivalent to allocating them *first*, or at least, not last. This provides further intuition as to why the more constrained they are, the better off they end up.

And as shown by the example, further assumptions are required concerning the mechanism by which peers may update their links if unstable allocations arise. Different mechanisms may favor either sharers or freeriders compared to the equiprobability benchmark. Due to the dependence of the solution on the fine details of the modeling choice, we assume equiprobability.

6.3 Applicability and limitations

We end with a few observations regarding the limitations of the model. First, in real networks peers hold multiple links. However, peers cannot connect to *all* sharers simultaneously; connectivity is generally limited. Our single link assumption attempts to capture this fact. We have considered a generalization of the model with more links per peer, but it increases substantially the complexity of the formula. Similarly, in real networks the upload bandwidth capacity differs between sharers. To consider heterogeneity in sharer capacities implies that certain sharers are able to provide higher utility to the peers that connect to them. As a result, in stable allocations more peers are connected to sharers with higher bandwidth capacity than to those without. In both cases, if the number of freeriders and sharers is large with respect to either the number of links or the asymmetries between sharers, we expect the advantage of sharers ($E_S \geq E_F$) and the approximation σ/n to hold, where σ is the total upload bandwidth offered by sharers.

We have not considered heterogeneity in the number of links, in the download capacity of peers, or the value of the content offered by sharers. Moreover, our model is static, it does not consider how stable allocations are reached nor the evolution of the network over time. In real networks, peers enter and leave. Even with these limitations, we hope to have provided a benchmark on which to construct these (and other) extensions. A clear understanding of congestion in p2p file sharing networks is a necessary first step towards providing a theory of incentives to contribute resources to p2p.[6]

[6]For a recent application using the s/n approximation result, see [4].

References

[1] Antoniadis, P., Courcoubetis, C. and Mason, R. (2004), 'Comparing economic incentives in peer-to-peer networks,' Computer Networks, Vol. 46, pp. 133-146.

[2] Asvanund, A., Clay, K., Krishnan, R. and Smith, M. D. (2004), 'An Empirical Analysis of Network Effects in Peer-to-Peer Music-Sharing Networks,' Information Systems Research, Vol. 15, pp. 155-174.

[3] Buragohain, C., Agrawal, D. and Suri, S. (2003), 'A Game Theoretic Framework for Incentives in P2P Systems,' Proceedings of the 3rd International Conference on Peer-to-Peer Computing.

[4] Casadesus-Masanell, R. and Hervas-Drane, A. (2006), 'Peer-to-Peer File Sharing and the Market for Digital Information Goods,' Mimeo, Harvard Business School.

[5] Chu, J., Labonte, K. and Levine, B. (2002), 'Availability and Popularity Measurements of Peer-to-Peer File Systems,' Proceedings of ITCom: Scalability and Traffic Control in IP Networks.

[6] Cunningham, B. M., Alexander, P. J. and Adilov, N. (2004), 'Peer-to-peer file sharing communities,' Information Economics and Policy, Vol. 16, Issue 2, pp. 197-221.

[7] Feldman, M., Lai, K. Chuang, J. and Stoica, I. (2003), 'Quantifying Disincentives in Peer-to-Peer Networks,' Proceedings of the 1st Workshop on Economics of Peer-to-Peer Systems.

[8] Feldman, M., Papadimitriou, C., Chuang, J. and Stoica, I. (2004), 'Free-riding and whitewash- ing in peer-to-peer systems,' Proceedings of the ACM SIGCOMM 2004 workshop on Practice and Theory of Incentives and Game Theory in Networked Systems (PINS).

[9] Golle, P., Leyton-Brown, K., Mironov, I. and Lillibridge, M. (2001), 'Incentives for Sharing in Peer-to-Peer Networks,' Lecture Notes in Computer Science, Vol. 2232, pp. 75+.

[10] Gummadi, K., Dunn, R., Saroiu, S., Gribble, S., Levy, H. and Zahorjan, J. (2003), 'Measurement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload,' Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP-19).

[11] Jackson, M.O. (2003), 'A Survey of Models of Network Formation: Stability and Efficiency,' Mimeo, California Institute of Technology.

[12] Jian, L. and MacKie-Mason, J. (2006), 'Why Share in Peer-to-Peer Networks?,' Proceedings of the ACM EC 2006 workshop on the Economics of Networked Systems (NetEcon).

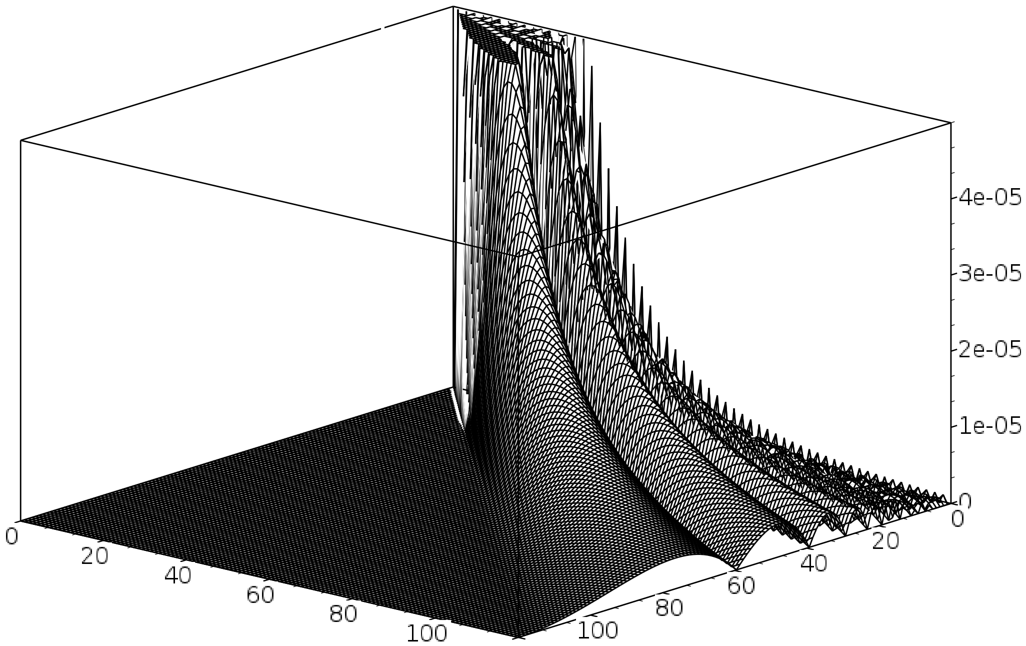
[13] Krishnan, R., Smith, M. D., Tang, Z. and Telang, R. (2004), 'The Virtual Commons: Understanding Content Provision in Peer-to-Peer File Sharing Networks,' Carnegie Mellon Heinz School, Working Paper No. 2004-19.

[14] Ripeanu, M. and Foster, I. (2002), 'Mapping the Gnutella Network: Macroscopic Properties of Large-Scale Peer-to-Peer Systems,' Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS'02).

[15] Saroiu, S., Gummadi, P. and Gribble, S. (2002), 'A Measurement Study of Peer-to-Peer File Sharing Systems,' Proceedings of Multimedia Computing and Networking.

[16] Sen, S. and Wang, J. (2004), 'Analyzing peer-to-peer traffic across large networks,' ACM/IEEE Transactions on Networking, Vol. 12, No. 2.

A Appendix



A three dimensional plot of *ade*.