Olson's Paradox Revisited: An Empirical Analysis of File-sharing Behavior in P2P File-sharing Communities.

Sylvain Dejean[1]
Corresponding author
CREM, Université de Rennes1, M@rsouin
sylvain.dejean@telecom-bretagne.eu +33 (0) 229 001 445
GIS-M@rsouin, Technopole Brest Iroise
CS 83818 - 29238 BREST Cedex 3
Thierry Penard
CREM, Université de Rennes1, M@rsouin
thierry.penard@univ-rennes1.fr +33 (0) 223 233 520
Faculté des Sciences Economiques, University of Rennes 1, 7 place Hoche
CS 86514, F-35065 Rennes cedex
Raphaël Suire
CREM, Université de Rennes1, M@rsouin
raphael.suire@univ-rennes1.fr +33 (0) 223 233 336
Faculté des Sciences Economiques, University of Rennes 1, 7 place Hoche
CS 86514, F-35065 Rennes cedex

Septembre 2009

Abstract:

This article aims to examine how the size of P2P file-sharing communities affects their functioning and performance (i.e. their capacity to share and distribute content). Olson (1965) argued that small communities are more able to provide collective goods. Using an original database on BitTorrent file-sharing communities, our article finds a positive relation between the size of a community and the amount of collective good provided ; However, the individual propensity to cooperate decreases with community size. These two features seem to indicate that P2P file-sharing communities provide a pure (non rival) public good. We also show that specialized communities are more efficient than general communities to encourage cooperative behavior. Finally, the rules designed by the administrators of these communities play an active role to stimulate voluntary contributions and improve the self-sustainability of file-sharing.

Classification code: H41; L86; K42

Keywords : Olson's paradox, P2P, File-sharing, BitTorrent

[1] This research received financial support from the Brittany Regional Council, as part of the P2Pimage research project. We also thank the participants at IIOC 2009 (Boston) and at the research seminars of Economix, CREM and Telecom Paris, as well as M. Arnold and M.D. Smith for their comments and suggestions.

Introduction

Olson (1965) developed a theory of collective action to explain why the existence of a common interest among a group of persons is not sufficient to act cooperatively. The outcome of collective action has the features of a public good because it benefits all people of the group regardless of their contribution (whether they have contributed a lot or have been free-rider). Olson argued that large groups are less able to promote their common interest than small ones because the incentives to contribute should diminish as group size increases. Olson's theory has influenced a large body of research in economics and politics. In particular, the presumed negative relationship between group size and the ability to provide collective good has been debated. Chamberlin (1974) considered that this depends on the nature of collective goods produced by the community. Rival goods are more likely to be provided by small groups, whereas inclusive or non rival goods are more efficiently produced in large groups. In the former case, the portion of collective good appropriated by each member decreases as the size of the group rises. Hence, a large group reduces individual incentives to contribute and is less likely to succeed in providing collective goods. With inclusive goods, each additional member does not reduce the share of collective goods consumed by existing members. They will slightly reduce their contribution, but this will be largely compensated by the additional contributions of new members. This implies that the amount of voluntary contributions to a non rival good should increase with group size, contrary to the Olson conjecture (Mc Guire, 1974).

More recently, Esteban and Ray (2001) revisited Olson's "group size paradox". They showed that with increasing marginal costs on collective action, large groups are more efficient than small groups to provide collective goods, even in the case of rival goods. But, Pecorino and Temimi (2008) found that the Olson conjecture is satisfied when group members bear a fixed cost of participation and the collective good exhibits a high degree of rivalry (see also Pecorino, 1999; Bergstrom et al., 1986; Gaube, 2001).

The impact of group size on voluntary contributions and free-riding has also been examined in several experimental studies. Except for the article of Isaac, Walker and Williams (1994), most of them found a negative impact of group size on voluntary contributions in the context of public good experiments (Chamberlin, 1978; Isaac and Walker, 1988, Marvell and Ames, 1979).

In this article, we want to revisit Olson's paradox in the context of P2P communities. These virtual communities have some specific characteristics that are particularly interesting to test Olson's conjecture. First, these communities are more exposed to free-riding than physical communities because they gather anonymous and distant users (Adar and Huberman, 2000; DangNguyen and Penard, 2007; Krishnan et al., 2007). Secondly, these communities can be extremely volatile because the cost of entry and exit is low. They can attract thousands of new members in a few days, but their size can also rapidly decrease (Krishnan et al., 2003). Thirdly, data from P2P communities can be easily collected and it is possible to permanently keep track of the file-sharing activity in these communities.

Our article aims to empirically examine whether the size of P2P communities affects the propensity of users to contribute voluntary to the collective good provided by these communities. In other terms, does the reach of a file-sharing community (measured by the number of active members) threaten or strengthen its existence, by reducing or stimulating voluntary contributions? In P2P file-sharing communities, voluntary contributions can take two different forms. First, members can feed the community with new content or files; i.e. they can upload and share a new file that will expand the catalog of content offered. Secondly members can share content that they have downloaded from other peers; i.e. after having downloaded a file, they can let this file available or accessible to the rest of the community (instead of removing it from the hard drive of their computer). In this case, she provides an additional source to download this file, and improves the speed and robustness of file-sharing.

This article is related to Asvanund et al. (2004) who analyzed how the size of music file-sharing communities may affect the availability and downloading quality of music files. The authors collected data on several public P2P networks (OpenNap) and found evidence of both negative and positive network effects in file-sharing communities. They estimated that the marginal benefit from an additional member decreases and the marginal cost increases with the size of the community. This implies that the optimal size for an OpenNap community is bounded.

Our article seeks to extend Asvanund et al. (2004) analysis to P2P communities that share any kind of content (not only music) and use a supposedly more efficient protocol (the BitTorrent protocol). Contrary to the Napster protocol (studied in Asvanund et al., 2004), the BitTorrent protocol prevents congestion by forcing users to share files during the time they are downloading them. But this protocol cannot force peers to voluntary contribute as long as they are not downloading, and it is an important question to know whether these communities have a bounded size or can preserve their efficiency as their size increases.

The administrators of BitTorrent communities have also the possibility to design specific organization rules, like rules for screening new members, monitoring behavior, or filtering content. How does the design of file-sharing communities affect their performance and sustainability?

This article uses data collected on 42 private BitTorrent communities during two months (Dec. 2007-Feb. 2008). These communities require members to be registered and sometimes to be co-opted by a member of the community. For each community and twice a day, we gathered information on the number of members, the number of files available, the number of sharers and downloaders.

Our findings show a positive relationship between the total amount of public good provided (measured by the number of files available and the number of sharers per file) and the size of a P2P community. But, the number of members has a negative impact on the individual propensity to share content (measured by the percentage of file sharers in the communities). According to Chamberlin (1974), these results suggest that the outcome of P2P communities is a pure public good (inclusive or non rival).

The rules designed by the administrators of these communities have also a significant impact on voluntary contributions. Our results highlight the fact that these communities are very innovative to regulate and manage indirect social interaction between anonymous and distant peers. Moreover, communities that share specialized content are more efficient in the provision of collective goods

The article is organized as follows. In the next section, we describe the Bit Torrent file sharing protocol and the costs and benefits to contribute in Bit Torrent communities. The dataset is described in the section 3. Section 4 presents the econometric model and comments the results. Section 5 concludes.

2. Bit Torrent file sharing system

Bit Torrent is now the most popular P2P file-sharing protocol in the world. Originally, in 2001, Bram Cohen designed this protocol to improve file-sharing for large size file. Bit Torrent is a “non pure” peer-to-peer system in which a central server, called the “tracker” collects information on the resources peers want to share (meta-data on the size, name and description of the shared files) and coordinates the transfer of files among users.

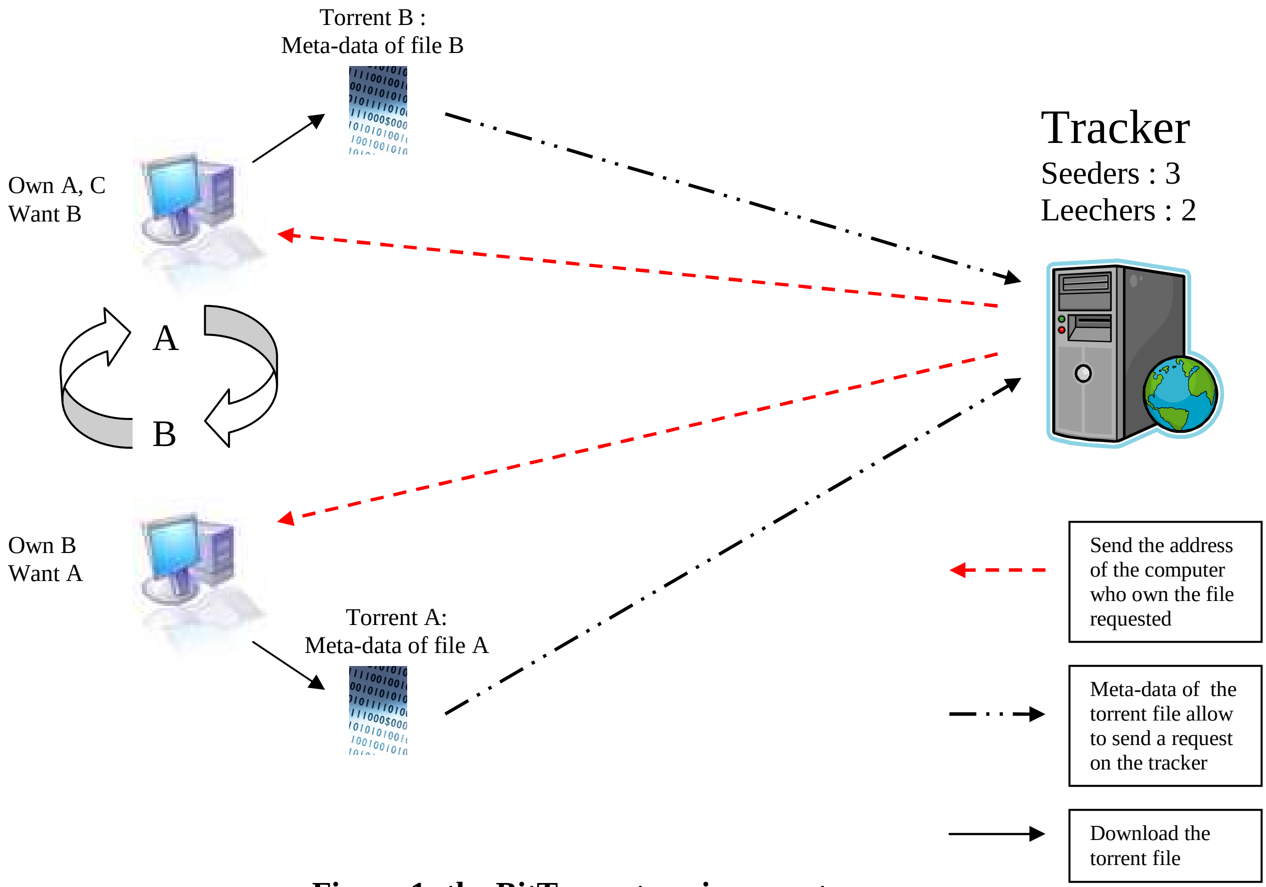


Figure 1: the BitTorrent environment

To download a file, the user has first to install a Bit torrent client (Azureus, Bit comet, µtorrent). Then, the user has to connect to a tracker that will send the address of the torrent that contains the desired file. To optimize the bandwidth allocation, files are divided among identically sized pieces called "schunk" and can be reconstituted only with "hashing information" contained in the torrent file. Once connected to the tracker, the sharers and downloaders of a file are automatically in contact with each other via their bit Torrent client and are exchanging (uploading and downloading) pieces of files (cf. figure 1). By helping users to find each other, the tracker also gathers statistical information about downloads and uploads. For each file indexed in the tracker, the users who are downloading it are called the "leechers" and the users who let the entire file available for other users are called the "seeders". The sum of leechers and seeders corresponds to the number of "peers" (a peer refers to a user who owns at least a piece of this file)[2].

Opportunistic behavior and congestion were strong issues in the first generation of P2P file sharing systems (Adar and Huberman, 2000; Krishnan et al., 2007). The Bit Torrent

[2] In the terminology of bitTorrent protocol, peers who share the same torrent constitute a swarm.

protocol was designed by Bram Cohen, to overcome this issue. It is based on tit-for-tat mechanism of file-sharing that imposes a minimum of cooperation (Cohen, 2003). Each peer is modeled as an intelligent automaton that maximizes its own interest (i.e. the downloading rate), rewarding peers who cooperate and punishing those who do not share. The more pieces of files a leecher is uploading towards another peer, the more pieces of files he can download from that peer.

For the user, the BitTorent protocol is transparent and the process described in the figure 1 is automated by the P2P client software. Unlike Napster or Gnutella, the user is automatically sharing the pieces of files that she is currently downloading. So she can never be a pure free rider and this reduces the risk of congestion (the higher the number of peers downloading the same file, the larger the number of sources to download pieces of this file in the meantime). However, forced sharing (while downloading) is not sufficient to guarantee the long-term viability of a community. Voluntary contributions are also important to feed the community with new content or files (to expand the quantity and diversity of the catalog) and to preserve the existing catalog. Sharing a complete file contributes to increase the choice of files available for downloaders and the speed of downloading.

What are the costs and the benefits for a member to voluntary contribute (i.e. to upload new files or to keep a downloaded file accessible to other members)? The benefit of sharing is to increase one's ratio of uploading to downloading. A better ratio can provide some privileges or priority in many BitTorrent communities. Interestingly this benefit may increase with the number of members in the community, because more members mean more potential downloaders for the files that are shared by the sharer. A contributor can increase more rapidly her individual ratio of uploading/downloading in a large community than in a small one, and can be more rapidly eligible to the privileges reserved to the active contributors. But with a larger community, a sharer may compete with more sharers who offer the same files. Consequently, the benefits of sharing a file could decrease as the community size rises, if the number of sharers for this file increases more rapidly than the number of downloaders.

The cost of contributing in a file-sharing community depends on the nature of the voluntary contribution. Uploading a new file in a P2P community requires some skills. The sharer must check that the file is not already available and that the quality fits with the standards of the community. Then, after having converted the file in the appropriate format and uploaded it on the server, the submission has to be approved by the community moderators before being available for downloading. The cost of sharing an existing file is probably lower, but includes the bandwidth or the hard disk space used to the storage of the shared files and the perceived risk of being sued and fined for illegal file-sharing. This cost seems to be independent of the size of the community.

Consequently, the individual amount of voluntary contributions, but also the aggregate amount of voluntary contributions (number of file sharers), could increase (decrease) with the size of a P2P community if the benefit of contributing tends to increase (decrease) with the number of potential downloaders in the community (as the cost is presumed to be independent of the community size).

3. Data

Description

Our sample is composed of 42 P2P file-sharing communities that can be either general or specialized in a type of content (music, movies, sport, adult, video games, and e-learning)[3]. All of them are "private" and "semi-private" trackers which contrary to "public trackers" (or open P2P communities) require every user to be registered. Between December 17, 2007, and February 17, 2008, and twice per day (at 10 am and 10 pm GMT[4]), we collected the number of unique files available[5], the number of users registered as well as the number of active file sharers and downloaded[6] in each community.

We define the proportion of sharers in a community as the number of file sharers over the number of peers (i.e. file sharers/(file sharers + file downloaders)). As a user can share or download simultaneously several files, this ratio is a proxy for the propensity of active users to contribute voluntary (i.e. to share new files or replicate existing files in the community). The panel gathers 5,097 observations (42 communities observed during 125 periods with 153 missing values).

Table 1 shows that the mean of active file sharers is 28,600 and the mean of active downloaders 12,967. Significant size differences in terms of file sharers, downloaders and registered users exist among the 42 Peer-to-Peer communities (from 556 members for the smallest community to more than 1.8 millions of members for the largest, with a median of 10,496 members). Free-riding seems to be limited in these 42 private communities. The average proportion of file sharers is equal to 81%, with a minimum ratio of 31% and a median ratio equal to 86%.

[3] A description of the 42 trackers is given in Annex 1 (location, category).
[4] We collected data at 11 am and 11 pm in France that is one hour ahead of Greenwich Mean Time.
[5] Files are called the "torrents" in the BitTorrent terminology.
[6] In the BitTorrent terminology, sharers and downloaders are respectively called "seeders" and "leechers". Note that a same user can be both leecher and seeder at the same time.

Table 1: summary statistics of variables collected

	Observation	Mean	Standard Deviation	Minimum	Maximum	Quartiles		
						25%	50%	75%
File sharers	5097	28600	72967	20	406838	1587	4933	19092
File downloaders	5097	12967	42621	1	337372	121	673	3535
Prop. of sharers	5097	0.81	0.15	0.31	0.99	0.78	0.86	0.91
Unique files available.	5097	6299	13310	33	74635	562	1652	4626.5
Registered members	5097	101721	343722	556	1804581	4635	10496	31789

The heterogeneity in our sample of P2P communities seems to be related to the nature of shared content. Some communities are "general" and provide various contents, like movies, TV series, music, video games or software. Others are specialized in a category of content and only accept the sharing of files belonging to this category.

Table 2 displays the features of P2P communities per type of content shared. Seven categories of communities have been considered (Generalist, Music, Adult, Movies, Video Game, E-learning, Sport). The Kruskal-Wallis test shows that these groups of communities are significantly different in terms of size and behavior. The comparison of the proportion of sharers suggests that free-riding is more widespread in our adult content communities than music or e-learning communities. The sample of adult content communities is also characterized by a larger number of registered users.

Table 2: Descriptive statistics (mean) by category of communities

	Generalist	Music	Adult	Movies	Video Game	E-learning	Sport	Kruskal-Wallis test
# communities	25	8	4	3	3	1	5	
File sharers	21520.3	12063	164712	12105.5	8464.5	3491.8	6169	***
Downloaders	16645.72	1622	55970	2536	1291.6	42.46	1154.8	***
Prop. of sharers	8.82	30.81	2.37	5.59	8.61	103.75	6.51	***
Unique files available	5453.3	5985	21488	7337	2076.8	2089.72	1282.96	***
Registered members	22950.6	31919.5	833116	19739.9	17876.8	11643.1	14310.4	***

Some evidence on the relationship between the size of a community and the amount of voluntary contributions

Olson’s conjecture presumes a negative relation between the size of a group and the amount of collective good provided. In the case of P2P file-sharing communities, the amount of collective good offered can be measured by the number of unique file available (the size of the catalog) and by the average number of sharers per unique file . The latter measure is a proxy for the availability of the catalog of files within the community. As the number of sharers per file increases, each downloader has more sources available to get the different pieces of each file.

Figure 2: Relationship between the number of unique files and the number of file sharers per unique files

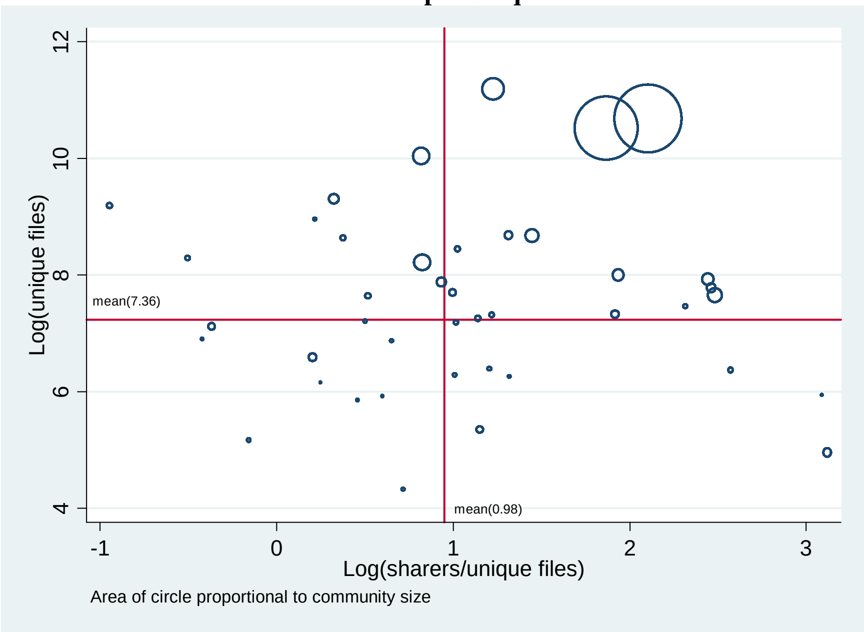
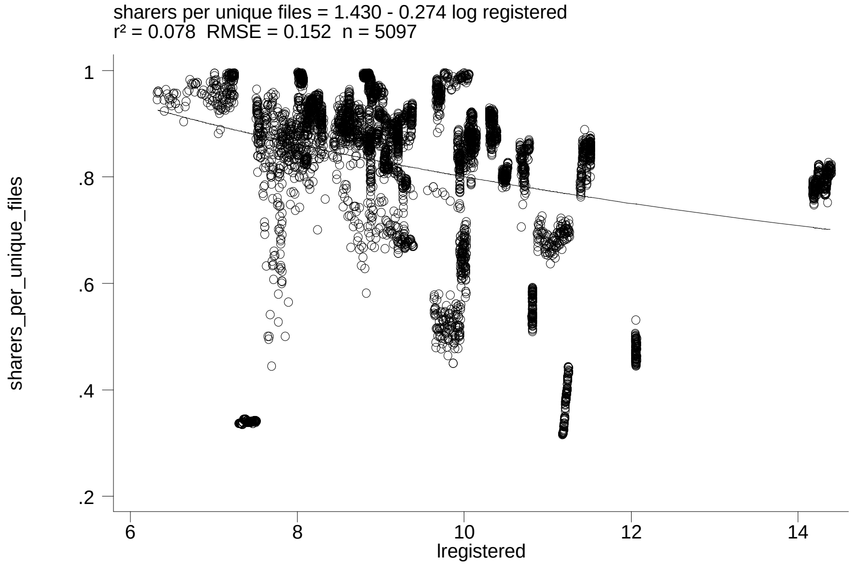


Figure 2 illustrates the relationship between the size of the catalog within a community and the number of file sharers per file. Moreover, each circle is proportional to the size of the community. The largest circles are concentrated in the north-east quarter. The biggest communities tend to provide larger catalogs and to have a high number of sharers per file (larger availability for their catalog).

Figure 3 displays the relationship between the proportion of file sharers in a community (sharers/peers) that is a proxy for the individual willingness to contribute within this community, and the number of registered users. Figure 3 suggests that the propensity to contribute tends to decrease as the size of the community rises.

Figure 3: Relationship between the proportion of file sharers and the size of a community



Some evidence on the relationship between institutional design of P2P communities and the amount of voluntary contributions

Our sample is composed of non-public BitTorrent communities. These communities require their users to be registered before having access to the catalog of files. Tracker administrators can also set other rules to constrain or control members' behavior. Our 42 communities present some differences in terms of organizational rules. These distinct features are taken into consideration through several variables.

First, we distinguish between private and semi-private trackers. A tracker is **"private"** when new users must be invited by a member of the community. This screening device should encourage cooperative behavior or reciprocity between members, and the amount of voluntary contribution should be higher in private communities than in semi-private communities.

We also create a dummy called **"control**" when the administrators of the tracker enforce a "sharing ratio" rule. It means that the members that do not achieve a given ratio of uploading to downloading, cannot download more files or can be excluded from the community. The enforced sharing ratio varies across trackers but is usually around 1 (the members must share at least as much as they download). This coercitive rule should prevent individual voluntary contributions from shrinking whatever the size of the community, enhancing the stability of large communities. But, this rule provides external incentives that could crowd-out intrinsic motivations to contribute (Benabou and Tirole, 2003) and undermine the quality of content shared by the peers.

We also control for the nature of content exchanged. The community is "**specialized**" (versus generalist) when file sharing is restricted to a specific category of content (for example, video games, music or adult video).

Finally, we measure the visibility of our sample of communities in the BitTorrent universe by searching each tracker's name on mininova.org[7]. If the search engine replies by listing several files that belongs to this tracker, we consider that this tracker is "**advertised**". For the administrators of a community, the interest of promoting their tracker on public search engine like mininova is explained by CurlyFries the founder of *TorrentFries*[8]: "*Dump sites are great promotional methods. Sites such as MiniNova and Demonoid allow you to upload torrents tracked elsewhere, so configure your new tracker to accept unregistered IP addresses (temporarily if you intend to go private) and upload your torrents to a bunch of dump sites like that. In the torrents' descriptions, include a comment such as "find more great torrents like this at www.example.com". You can even throw a text file inside the torrent to the same effect. You'd be amazed by how well it works*". By enhancing the visibility of their community in BitTorrent meta-search engine, the administrators can attract new members that will be used to disseminate the catalog of files within the community.

We perform Mann-Witney tests on the average number of sharers, downloaders, members, and files available between the different categories of communities (private, control, advertised, specialized). Table 3 suggests that communities with stricter rules for admission (private) and for downloading (control), or that advertised their content on public search engine, have less registered members and a smaller catalog, but exhibit a higher proportion of file sharers. Specialized communities are also characterized by a higher ratio of sharers, as well as more unique files and membership.

Table 3: Mann-Whitney test for identity of distribution

	# communities	File sharers	File downloaders	Prop. of sharers	Unique files available	Registered
Private	4	8973	1029	0.87	2780	9581
Semi private	38	30552	14154	0.81	6650	110883
Mann-Whitney test		**** (-4.64)	ns (1.59)	*** (-3.88)	*** (-4.6)	*** (7.57)
control	8	21225	11117	0.83	5147	67356

[7] Mininova is the largest torrent search engine with more than 3 billions of visitors per day

[8] Torrent Fries is one of a rare site dedicated to the running of a tracker.

No control	34	59234	20651	0.72	11088	560994
Mann-Whitney test		*** (16.2)	*** (19.5)	*** (-12.7)	*** (9.9)	*** (7.3)
Specialized	24	35775	10695	0.83	7108	159401
Generalist	18	18646	16118	0.79	5178	21698
Mann-Whitney test		*** (-14.7)	*** (-4.6)	** (0.01)	*** (-14.3)	*** (-19.9)
Advertised	19	8112	1803	0.84	3237	17399
Non advertised	23	46435	22684	0.78	8966	175120
Mann-Whitney test		*** (25.2)	*** (21.2)	*** (-12.8)	*** (15.8)	*** (21.5)

Note: ***, **, * mean significant at the level of 1%, 5%, and 10% respectively

4. Econometric models and results

In this section, we present the specifications and the results of the econometric models used to analyze the impact of community size on the amount of collective good provided (both in terms of catalog size and sharers per unique file) and on the individual propensity to contribute (measured by the proportion of file sharers).

Econometric models

We estimate three models (M1, M2, M3) using the number of unique files as dependent variable in M1, the number of sharers per file in M2 and the proportion of sharers in M3. For each model, we consider three different specifications for the number of registered members in the community: a linear specification using the number of registered users, a quadratic specification using the registered users and users-squared and a log-log specification using the log of registered users. For the log-log specification, the three models are given by:

$$\log(\textit{unique files}_{it}) = \beta_0 + \beta_1 \log(\textit{registered}_{it}) + \beta_2(\textit{control}_i) + \beta_3(\textit{private}_i) + \beta_4(\textit{advertised}_i) + \beta_5(\textit{specialized}_i) + \varepsilon_{it} \qquad (M1)$$

$$\log(\textit{sharers per unique file}_{it}) = \beta_0 + \beta_1 \log(\textit{registered}_{it}) + \beta_2(\textit{control}_i) + \beta_3(\textit{private}_i) + \beta_4(\textit{advertised}_i) + \beta_5(\textit{specialized}_i) + \varepsilon_{it} \qquad (M2)$$

$$\log(\textit{prop.of files sharers}_{it}) = \beta_0 + \beta_1 \log(\textit{registered}_{it}) + \beta_2(\textit{control}_i) + \beta_3(\textit{private}_i) + \beta_4(\textit{advertised}_i) + \beta_5(\textit{specialized}_i) + \varepsilon_{it} \qquad (M3)$$

The other explanatory variables are time invariant dummies that control the features and rules of the communities. (M1) and (M2) aim to estimate how the size and the design of a P2P community influence its efficiency or quality, measured by the number of original files uploaded and shared in this community i at time t ($unique\ files_{it}$ M1) and by the average number of peers that share the same file ($sharers\ per\ file_{it}$ M2). In other terms, these two models indicate the quantity of files available for downloading and the quality of downloading. (M3) estimates how the features of a P2P community (number of members and organizational rules) affect cooperative behavior within this community. The propensity to cooperate is measured by the ratio of sharers to peers ($prop.of\ \ file\,sharers_{it}$).

As we have to deal with time-series cross-sectional (TSCS) data with a number of periods superior to the number of communities, potential problems in the error structure have to be addressed. First, the Breush-Pagan/Cook-Weisberg test for constant variance fell within the confidence interval of 10 percent for (M1) (M2) and (M3). Secondly, the strong heterogeneity in the size of our 42 communities is likely to cause a problem of groupwise heteroscedasticity. The modified Wald test for groupwise heteroscedasticity confirmed that the variance of error process differs across units for (M1) (M2) and (M3). Because our data exhibits a large temporal dimension and that observations at 10 am are correlated with observations at 10 pm, we suspected the presence of residuals serial correlation. This was confirmed by a test for autocorrelation in panel-data (Woodridge, 2002)[9], but only for equation (M1). For all these reasons, the feasible general least square (FGLS) is the most appropriate estimator in presence of panel-level heteroscedasticity and autocorrelation. The FGLS is similar to generalized least squares except that it uses an estimated variance-covariance matrix since the true matrix is not known directly. The covariance matrix is estimated by iteration, using the OLS estimators in the first step.

Using the Panel Corrected Standard Errors (PCSE) estimators proposed by Beck and Katz (1995) would have been a possibility. However it is less efficient for panel data when temporal dimension exceeds individual dimension (Chen et al. , 2006).

Table 4 displays the estimates for the three specifications of the models (M1). The columns (1a with log specification), (2a with linear specification) and (3a with quadratic specification) report the baseline OLS (ordinary Least Square) estimates for equation (M1), the robust standard errors in brackets are calculated using the Hubber and White sandwich estimator. The columns (1b), (2b) and (3b) display the FGLS estimates for equation M1,

[9] Drukker (2003) provides a simple program to perform this test in Stata.

controlling for heteroscedasticity and serial autocorrelation with a first order auto regressive coefficient. The *modus operandi* is repeated for the model (M2) and (M3) respectively in the table 5 and 6: the columns (1a) (2a) and (3a) report the OLS estimates while (1b) (2b) and (3b) show the results of the FGLS regression.

[Insert tables 4, 5 and 6]

The impact of community size

Our results suggest that community size (registered users) has a positive impact on the amount and the quality of collective good provided by the file-sharing community. However, the estimated effect is lower with the FGLS method than the OLS approach and supports our choices to control for serial correlation.

Non-linear specifications provide better estimates than the linear specification.. The log-log specification models suggest that the quantity and quality of sharing tend to increase with the number of members, but at a decreasing rate, as the coefficient of log(registered) is positive but significantly below one.[10] It indicates that when the size of a community doubles, the size of the catalog only increases by 5% and the number of sharers per unique files by 16%. Two reasons could explain this weak effect. It is possible that the late members in a BitTorrent community are contributing less than the early members. The entry of new members can also impair cooperative behavior among the early members. However, the total amount of voluntary contributions is not diminishing with community enlargement. The BitTorrent communities seem to adapt efficiently to the entry of new members and apparently are not size-bounded.

The results of the last model (M3) are displayed in Table 6. We find a negative relationship between the size of the community and the individual propensity to voluntary contribute. The results suggest that the individual incentives to share files decrease with the number of members, but at a decreasing rate (the coefficient of log registered is negative but between -1 and 0). According to the log-log specification, in response to a 100% increase of the community size, the proportion of sharers decreases by 3%. This finding supports the idea that the proportion of file sharers would never tend to zero even in large communities. This is not surprising because many of these communities require a minimum ratio of uploading to downloading from their members and are quite proficient to reduce free-riding. People tend to

[10] This result is also consistent with the estimates obtained in the quadratic specification (i.e. for registered, the coefficient is positive and for registered-squared, the coefficient is negative).

be less cooperative in larger community, but there is always a core of contributors who preserve the stability and quality of the file-sharing community. The quadratic specification also suggests that the number of sharers per unique files decreases until it reaches a minimum, and then may increase, as the size of the community rises. This is consistent with the idea that core contributors of a file sharing network can increase their voluntary contributions when congestion effects become too important (Krishnan et al., 2004; DangNguyen and Penard, 2007).

To summarize, even if the individual incentives to contribute voluntary tend to decrease as community size rises, the aggregate collective contributions increase. In other terms, even if each member is sharing less content, the size of the catalog as well as the quality of downloading increase with the number of members among the community. This seems to indicate that the provision of a BitTorrent community is an inclusive or non rival good according to Chamberlin (1974).

The role played by the rules designed by communities' administrators

Olson (1965) stated that large groups could overcome free-riding by providing private incentives or exclusive services to the active members. Some of the dummies used to control for the features of our P2P communities can be analyzed as private incentives (private, control). In the log-log specification, we find that "private" communities provide a higher quantity and quality of collective good. It can be explained by the fact that private communities are more selective and can handpick their members based on their ability to contribute to the collective good. Entry regulation enables not only to prevent opportunistic behavior, but also to better segment users' needs and interests. However, this econometric result is very sensitive to the specification used.

Enforcing a minimum ratio of data uploading over downloading is supposed to rule out opportunistic behavior (free-riding). Table 6 confirms that monitoring behavior increases the proportion of files sharers. However, communities that enforce a minimum ratio of uploading ("control") provide a smaller catalog and offer less available sources of downloading for each file. This result shows that such rules may have counterproductive effects by reducing cooperative behavior based on voluntary reciprocity. Coercitive rules can crowd-out intrinsic motivations to contribute and reduce the global amount of collective good (Benhamou and Tirole, 2003). This kind of rule can also induce strategic behavior that is detrimental to the size and variety of the catalog of files shared in the community, because members know that the easiest way to rapidly reach the required sharing ratio is to share popular files that are frequently requested. Consequently, sharers are concentrated on few files to the detriment of increasing variety.

Specialized communities seem to encourage voluntary contribution. Probably, members of specialized communities are more strongly involved and incited to cooperate with each other (Asvanund *et al.*, 2006). Table 4 and 6 confirm the idea that the propensity to contribute as well as the size of the catalog is higher in a topic-oriented community. Finally, a community that relies on public search engine to promote its catalog (an *advertised* community) has a higher proportion of file sharers, but doesn't provide more amount of collective good.

These findings highlight the fact that, in order to provide a high quantity and quality of collective good, a community must design and implement efficient organizational rules, to prevent free riding behavior and promote a variety of content that matches the preferences of the members. The decrease of searching cost as well as the enhancement of individual capabilities to share is not a sufficient condition to provide a sustainable model of file sharing.

5. Conclusion

This paper has investigated the relationship between the size of file-sharing communities and their ability to provide collective goods (measured by the quantity and availability of content in the community). During two months between December 2007 and February 2008, we collected data on the activity of 42 private and semi-private bitTorrent communities. Our results suggest that the collective provision in these communities can be analyzed as a pure public good. The amount of collective good increases with the number of registered users whereas the individual propensity to contribute decreases with community enlargement. We also show that the rules designed by the administrators of these communities have a significant impact on the performance of the community and on their sustainable size. We find that stricter monitoring schemes have a negative impact on the quantity and quality of files shared. The provision of a large online catalog (or a long tail) of contents that match individual preferences cannot be disconnected from the management and the organizational design of these virtual communities. This challenging question deserves further investigation.

It would be interesting to compare the centralized model of online merchants and the decentralized model of P2P community to manage and promote the long tail. Which model is the more efficient and sustainable to connect the supply and demand of rare content? How do you articulate market and non market incentives, external and intrinsic motivations to provide and distribute rare and popular content that match perfectly the needs of consumers?

References:

Adar, E., Huberman, B., 2000. Free Riding on Gnutella. First Monday **5**.

Asvanund, A., Clay, K., Krishnan, R., Smith, M.D., 2004. An Empirical Analysis of Network Externalities in Peer-to-Peer Music-Sharing Networks. Information Systems Research **15**, 155-174.

Asvanund A., Krishnan, R., Smith M.D., Telang, R., 2006. Interest-Based Self-Organizing Peer-to-Peer Networks: A Club Economic Approach. Working paper available at SSRN http://ssrn.com/abstract=585345.

Beck, N., Katz, J.N., 1995. What to Do and Not to Do With Time-Series Cross Section Data. The American Political Science 89, 634-647.

Benabou, R., Tirole, J., 2003. Intrinsic and Extrinsic Motivation. Review of Economic Studies 70(3), 489-520.

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

Bharambe, A.R., Herley, C., Padmanabhan, V.N., 2005. Analyzing and Improving BitTorrent. Technical Report MSR-TR-2005-03, Microsoft Research.

Chamberlin, J., 1974. Provision of Collective Goods As a Function of Group Size. The Behavioural Science 26, 441-45.

Chamberlin, J., 1978. The Logic of Collective Action: Some Experimental Results. The American Political Science Review 68, 707-716.

Chen, X., Lin, S., Reed, W.R., 2006. A Monte Carlo Estimation of the Efficiency of the PCSE Estimator. Working paper, College of Business and Economics University of Canterbury.

Cohen, B., 2003. Incentives Build Robustness in BitTorrent. Proceedings of International Workshop on peer-to-peer Systems.

Dang-Nguyen, G., Pénard, T., 2007. Network Cooperation and Incentives Inside Online Communities. In : Brousseau, E., Curien, N., (Eds.), Internet and Digital Economy, Cambridge University Press.

Drukker, D.M., 2003. Testing For Serial-Correlation in Linear Panel-data Models. The Stata Journal 3, 168-177.

Esteban, J., Ray, D., 2001. Collective Action and the Group Size Paradox. The American Political Science Review 95, 663-672.

Gaube, T., 2001. Group size and free riding when private and public goods are gross substitutes. Economics Letters 70, 127 –132.

Huang, K., Wang, L., Zhang, D., Liu, Y., 2007. A Dynamic Quota-Based Peers Selection in Bit Torrent. Proceedings of the Sixth International Conference on Grid and Cooperative Computing.

Isaac, R.M., Walker, J.M., 1988. Group Size Effect in Public Good Provision: The Voluntary Contributions Mechanism. The Quaterly Journal of Economics 103, 179-99.

Isaac, R.M., Walker, J.M., Williams, A.W., 1994. Group Size and the Voluntary Provision of Public Goods Experimental Evidence Utilizing Large Group. Journal of Public Economics 54, 1-36.

Krishnan, R., Smith, M.D., Tang, Z., Telang, R., 2004. The Virtual Commons: Why Free Riding Can Be Tolerate in File Sharing Network. Working paper, available at SSRN http://papers.ssrn.com/sol3/papers.cfm?abstract_id=450241.

Krishnan, R., Smith, M.D., Tang, Z., Telang, R., 2007. Digital Business Models for Peer-to-Peer Networks: Analysis and Economic Issue. Review of Network Economics, 6, 194-213.

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

Legout A., Urvoy-Keller G., Michiardi, P., 2005. Rare First and Chock Algorithms are Enough. Proceding of the 6th ACM SIGCOMM conference on Internet measurement.

McGuire, M.C., 1974. Group size, group homogeneity, and the aggregate provision of a pure public good under Cournot behavior. Public Choice 18, 107–126.

Marwell, G., Ames, R.E, 1979. Experiments on the Provision of Public Goods. I. Resources, Interest, Group Size and the Free-Rider Problem. The American Journal of Sociology 84, 1335-1360.

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

Pecorino, P., 1999. The Effect of Group Size on Public Good Provision in a Repeated Game Setting. Journal of Public Economics 72, 121-134.

Pecorino, P., Temimi, A., 2008. The Group Size Paradox Revisited. Journal of Public Economic Theory 10, 785-799.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. Cambridge, MIT press, Cambridge

Yue Y., Lin C., Tan, Z., 2006. Analysing the Performance and fairness of Bit-Torrent Like Protocol Using a general Fluid Model. Computer Communication, Vol. 29, 3946-3956.

Zhang, M., Zhu, F., 2007. Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia. Working Paper#07-22, NET Institute.

Annex 1

Site of the tracker	N° tracker	content	private	specialized	control	advertised
http://www.captain-tracker.fr/index.php	1	general	0	0	1	0
http://www.sharing-torrents.com/	2	general	0	0	1	1
http://leparrain.mine.nu/torrents.php	3	general	0	0	0	0
http://www.unlimited-tracker.net/	4	general	0	0	1	1
http://www.nhltorrents.co.uk/	5	sport	0	1	1	1
http://xtremewrestlingtorrents.net/static.php	6	sport	0	1	1	0
http://www.dimeadozen.org/index.php	7	music	0	1	0	1
http://www.indietorrents.com/index.php	8	music	1	1	1	0
http://shnflac.net/index.php	9	music	0	1	0	0
http://jamtothis.com/	10	music	0	1	1	0
http://www.browntracker.net/browse.php	11	music	0	1	0	0
http://anvilofsound.com/	12	music	0	1	0	1
http://mixes.dfx.at/index.php	13	music	0	1	1	0
http://asiandvdclub.org/	14	cinéma	0	1	1	0
http://alt.bitworld.to/browse.php	15	general	0	0	1	1
http://www.araditracker.com/	16	general	0	0	1	0
http://www.titaniumtorrents.net/	17	general	1	0	1	1
http://dididave.com/	18	general	0	0	1	0
http://www.quebectorrent.com/	19	general	0	0	1	1
http://cinemageddon.org/	20	cinéma	0	1	1	1
http://www.blades-heaven.com/index.php	21	general	0	0	1	1
http://www.puretna.com/	22	adult	0	1	0	0
http://www.kingdomxxx.com/	23	adult	0	1	0	0
http://www.empornium.us/	24	adult	0	1	1	0
http://www.pornevo.com/	25	adult	0	1	1	0
http://www.underground-gamer.com/	27	vidéo game	0	1	1	1
http://www.pleasuredome.org.uk/	28	vidéo game	0	1	1	1
http://my-gamebox.com/	29	vidéo game	0	1	1	1
http://thepeerhub.com/	30	general	0	0	1	1
http://bitnation.com/index.php	31	general	1	0	1	1
http://p2pworld.ulmb.com/	32	general	0	0	1	1
http://torrent-hackers.co.uk/	33	general	0	0	1	1
http://www.sport-scene.net/	35	sport	0	1	1	0
http://www.sportbit.org/	36	sport	0	1	1	0
http://www.prosporttorrents.net	37	sport	0	1	1	1
http://www.mamietracker.com/index.	38	general	0	0	0	1

php						
http://zombtracker.the-zomb.com/	39	music	0	1	1	0
http://cinematik.net/	40	cinéma	1	1	1	0
http://www.zinebytes.org/	41	e-learning	0	1	1	0
http://www.mytracker.ru/index.php	42	general	0	0	1	0
http://linuxmafia.net/	43	general	0	0	1	1
http://zerotracker.com/index.php	44	general	0	0	1	0

Table 4: Estimation of the equation (M1) using OLS and FGLS estimators

	Dep. Var= unique files					
	OLS (1a)	FGLS (1b)	0LS (2a)	FGLS (2b)	OLS (3a)	FGLS (3b)
Log(registered)			0.67	0.05		
			(73.51)***	(8.66)***		
Registered	2.43e-02	1.06e-02			1.70e-01	3.72e-04
	(132.27)***	(7.66)***			(18.61)***	(0.77)
Registered^2					-8.88e-08	1.82e-08
					(16.09)***	(3.17)***
Control	-1772.58	-2294.24	-0.37	-0.56	-3762.61	-8653.65
	(5.81)***	(10.65)***	(7.13)***	(11.46)***	(14.75)***	(5.39)***
Private	-1158.42	744.00	0.72	0.62	1458.68	-13096.13
	(5.59)***	(3.04)***	(22.97)***	(11.62)***	(10.44)***	(2.11)**
Advertised	-2589.04	-1202.97	-0.13	-0.28	-812.10	2918.08
	(6.43)***	(10.12)***	(3.44)***	(6.50)***	(3.87)***	(1.65)*
Specialized	-2528.10	-1116.86	-0.25	0.27	-2849.88	6081.48
	(5.71)***	(9.51)***	(6.97)***	(6.51)***	(8.56)***	(3.58)***
Constant	8035.01	4933.10	1.47	7.59	5386.79	-12813.23
	(16.11)***	(21.59)***	(13.22)***	(93.15)***	(14.72)***	(7.27)***
Observations	5097	5097	5097	5097	5097	5097
# of communities		42		42		42
R2	0.43		0.51		0.6	

Table 5: Estimation of the equation (M2) using OLS and FGLS estimators

Dep. Var= file sharers/unique files						
	OLS (1a)	FGLS (1b)	0LS (2a)	FGLS (2b)	OLS (3a)	FGLS (3b)
Log(registered)			0.18	0.16		
			(25.03)***	(94.67)***		
Registered	6.29e-07	1.74e-06			6.37e-06	9.04e-06
	(5.89)***	(26.26)***			(6.31)***	(18.09)***
Registered^2					-3.50614e-12	-4.67296e-12
					(5.82)***	(14.56)***
Control	-3.66078	-1.56912	-0.31976	-0.15571	-3.74	-2.83
	(12.77)***	(13.41)***	(7.24)***	(11.70)***	(13.05)***	(22.51)***
Private	0.19	-0.58	0.21	0.07	0.30	-0.62
	(1.05)	(6.05)***	(5.51)***	(3.37)***	(1.61)	(6.30)***
Advertised	-2.03	-1.22	-0.45	-0.46	-1.96	-1.12
	(10.82)***	(23.20)***	(14.52)***	(52.55)***	(10.22)***	(21.67)***
Specialized	-1.11	-0.55	-0.19	-0.17	-1.12	-0.65
	(5.47)***	(11.88)***	(6.72)***	(18.37)***	(5.53)***	(14.20)***
Constant	9.23	5.91	-0.07	-0.12	9.12	7.04
	(23.50)***	(44.97)***	(0.65)	(4.90)***	(22.98)***	(55.96)***
Observations	5097	5097	5097	5097	5097	5097
# of communities		42		42		42
R2	0.11		0.18		0.11	

Table 6: Estimation of the equation (M3) using OLS and FGLS estimators

	Dep. var= proportion of file sharers					
	OLS (1a)	FGLS (1b)	OLS (2a)	FGLS (2b)	OLS (3a)	FGLS (3b)
Log(registered)			-0.03 (11.58)***	-0.03 (33.42)***		
Registered	-7.46e-10 (-0.18)	-1.34e-08 (5.14)***			-1.44e-06 (29.55)***	-1.51e-06 (43.52)***
Registered^2					8.79e-13 (29.43)***	9.20e-13 (43.87)***
Control	1.13e-01 (16.90)***	0.101 (35.29)***	0.189 (15.73)***	0.132 (26.99)***	0.13 (18.82)***	0.15 (47.42)***
Private	0.04092 (8.86)***	-0.00 (0.76)	0.04 (8.57)***	0.00 (0.51)	0.01 (3.44)***	-0.02 (8.19)***
Advertised	0.08 (18.88)***	0.07 (45.32)***	0.11 (15.08)***	0.06 (27.63)***	0.06 (14.52)***	0.04 (19.40)***
Specialized	0.08 (16.90)***	0.04 (19.52)***	0.15 (20.10)***	0.07 (22.63)***	0.08 (18.91)***	0.03 (16.58)***
Constant	0.64 (73.88)***	0.69 (216.95)***	-0.26 (8.04)***	-0.08 (7.57)***	0.67 (74.38)***	0.70 (204.69)***
Observations	5097	5097	5097	5097	5097	5097
# of communities		42		42		42
R2	0.16		0.22		0.28	