

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

**Sylvain Dejean<sup>12</sup>, Thierry Penard<sup>3</sup> and Raphaël Suire<sup>4</sup>**

**March 2009**

**Preliminary draft**

## **Abstract:**

This article aims to examine the impact of group size on the provision of collective good provided by P2P file-sharing communities. Olson (1965) argued that small communities are more able to provide collective actions. 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 group 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 managers of a community play an active role to stimulate voluntary contributions and improve the self-sustainability of file-sharing.

---

<sup>1</sup> This research received financial support from the Brittany Regional Council, as part of the P2Pimage research project.

<sup>2</sup> CREM, University of Rennes 1, Marsouin, sylvain.dejean@univ-rennes1.fr

<sup>3</sup> CREM, University of Rennes 1, Marsouin, thierry.penard@univ-rennes1.fr

<sup>4</sup> CREM, University of Rennes 1, Marsouin, Raphael.suire@univ-rennes1.fr

## **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 good public features because it benefits all, even those who have not contributed. 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 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. Exclusive or rival goods that are limited in supply 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 with group size. Hence, a large group reduces the individual incentive 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 overcompensated by the additional contribution of the new member. This implies that the amount of voluntary contributions should increase with group size, contrary to the Olson's 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, Blume and Varian, 1986; Gaube, 2001).

The impact of group size on voluntary contributions and free-riding has been examined in several experimental studies. Except for the article of Isaac, Walker and Williams (1994), most of them have found a negative impact of group size on the voluntary contribution process 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 hypothesis. First, these communities gather anonymous and distant users and are theoretically strongly exposed to free riding as shared files are non-exclusive and non-rival (Adar and Huberman, [2000], DangNguyen and Penard [2007], Krishnan et al., [2007]). Compared to real communities, the size of such communities is also extremely volatile, with most of the time a weak cost of entry and exit (Krishnan et al., [2003]). Third, data from P2P communities can be easily collected and it is possible to permanently keep track of the file-sharing activity in these communities.

We aim to empirically examine whether the size of P2P communities affects the propensity of users to contribute voluntarily to the collective good provided by these communities. In other terms, does the reach of a file-sharing community (measured by its size) threaten or strengthen its existence, by reducing or stimulating voluntary contributions?

In P2P file-sharing communities, voluntary contributions can take two different forms. First, a member can feed the community with new content or files; i.e. she can upload and share a new file that will expand the catalog of files offered. Secondly a member can share content that she has downloaded from another peer; i.e. after having downloaded a file, she can let this file available or accessible to the rest of the community (instead of removing it from her hard drive). In this case, she provides an additional source to download this file, and improves the speed and robustness of file-sharing.

Asvanund et al. (2004) 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 both evidence of 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 these P2P communities should be bounded.

Our article aims to extend Asvanund *et al.* analysis to P2P communities that share any kind of content (not only music) and that use a specific protocol (the BitTorrent protocol). We have collected data 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.

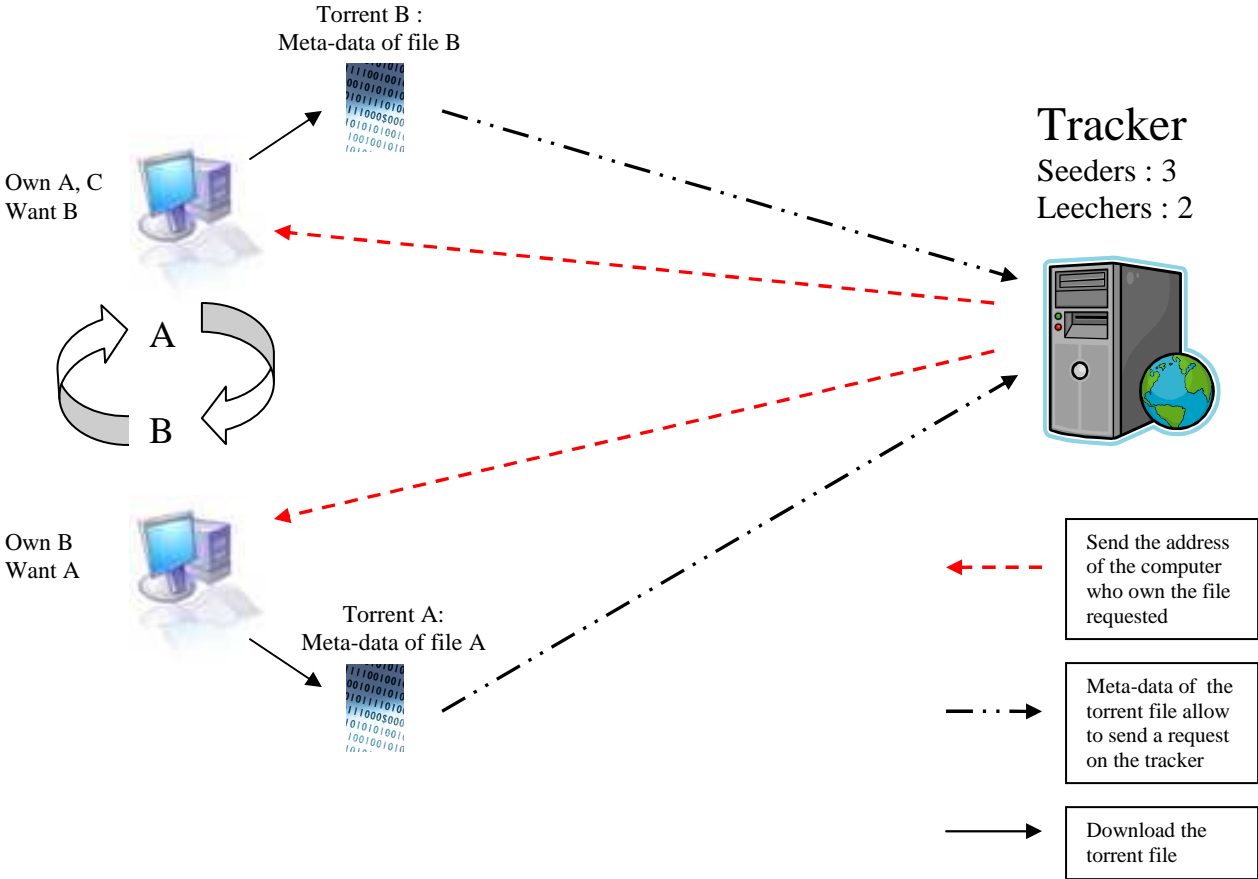
Our findings show a positive relationship between the amount of public good provided (measured by the number of files available and the number of sharers per file) and the size of the community. But, the number of members has a negative impact on the individual

propensity to share content. Both these results suggest that the outcome of P2P communities is a pure public good. We also find a significant impact of the rules designed by the managers of these communities on voluntary contributions.

The article is organized as follows. In the next section, we describe the Bit Torrent file sharing protocol and the cost and benefit 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 file size. 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, she has to connect to a tracker that will send her 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 could 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 files indexed in the tracker, the number of 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).

The excess of opportunistic behavior was a main issue in the first generation of P2P file sharing system (Adar and Huberman, 2000, Krishnan, Smith, Tang and Telang, 2007). The Bit Torrent 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. The protocol also contains original functionalities like “optimistic unchoking” that can initiate cooperation by randomly uploading to a peer regardless of his download rate<sup>5</sup> or “rarest first” policy that makes the rarest “schunk” of a particular file more requested than the more common ones (in order to facilitate the diffusion of files). Even if the robustness and the performance of the protocol have been largely demonstrated (Legout, Urvoy-Keller and Michiardi, 2005, Yue, Lin and Tan, 2006), this system is far from being perfect and several technical improvements have been proposed to the original protocol (Bharambe, Herley and Padmanabhan , 2005, huang, Wang, Zhang and Liu, 2007).

For the user, the BitTorrent 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 problem (The higher the number of peers downloading the same file, the larger the number of sources to download the pieces of this file at this time). However, this forced sharing (while downloading) is not sufficient to

---

<sup>5</sup> According to Bram Cohen (2003) “Optimistic unchokes correspond very strongly to always cooperating on the first move in prisoner’s dilemma”

guarantee the long-term viability of the 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. What are the cost and the benefits for a member to upload a new file or to keep a downloaded file accessible to other members?

The benefit of contributing is a better ratio of uploading to downloading that may provide some privileges or priority in many BitTorrent communities. In a sense, this kind of benefit can rise with the number of members in the community, by increasing the number of leechers per file. The second form of benefit is more indirect. By sharing files, you can induce positive reciprocity and expect contributions from the other members. But this effect is stronger in small communities than in large one. Thus, the relationship between the benefit of contributing and the size of the community is unclear.

The cost of contributing in a file-sharing community depends on the nature of the contribution. Uploading a new file in a closed community is not so easy. The “uploader” must check that the file is not already available and the quality fits with the standard 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.

To sum-up, the global amount of voluntary contributions (number of file sharers), but also the individual amount of voluntary contributions with community size if the benefit of contributing tends to increase with the number of members in the community (as the cost is always constant).

### **3. Data**

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)<sup>6</sup>. 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, we collected twice per day (at 11am and 11pm) the number of unique files

---

<sup>6</sup> The location and categories of the 42 trackers are given in Annex 1.

available<sup>7</sup>, the number of users registered as well as the number of active file sharers and downloaded<sup>8</sup> in each community. We define the ratio of contributors in a community as the number of file sharers over 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 cooperate (i.e. to share files in the community). The panel gathers 5,097 observations (42 communities observed during 125 periods with 153 missing values).

**Table 1: summary statistics of variables collected**

	Observation	Mean	Standard Deviation	Minimum	Maximum
File sharers	5097	28600	72967	20	406838
File downloaders	5097	12967	42621	1	337372
Ratio	5097	0.81	0.15	0.31	0.99
Unique files available.	5097	6299	13310	33	74635
Registered members	5097	101721	343722	556	1804581

Table 1 shows that the mean of active file sharers is 28,600 and the mean of active downloads 12,967. Significant size differences in terms of file sharers, downloaded 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). Free-riding issues seem to be limited in these 42 private communities since the average proportion of file sharers is equal to 81%. The heterogeneity of individual behavior in P2P communities can be

emphasized by comparing the average ratio  $\left( \frac{\sum_{t=1}^T \sum_{i=1}^N files\ shared_{it}}{\sum_{t=1}^T \sum_{i=1}^N peers_{it}} \right)$  equals to

0.68 and the observed ratio  $\left( \sum_{t=1}^T \sum_{i=1}^N (files\ shared_{it} / peers_{it}) \right)$  equals to 0.81. If our 42

communities were homogeneous, the observed ratio should be the same than the average ratio. The heterogeneity can be explained by the nature of shared content. Some communities are “general” and provide 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.

<sup>7</sup> Files are called the “torrents” in the BitTorrent terminology

<sup>8</sup> 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 2: Descriptive statistics 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	***
Ratio	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	***

Table 2 displays the features of P2P communities per type of content shared. The Kruskal-Wallis test shows that the differences between our 7 categories of communities are significant. Hence the comparison of the file sharers/peers ratios per category suggests that free-riding is more widespread in adult content communities than music or e-learning communities. Adult content communities are also characterized by a larger number of registered users.

To test the relevance of Olson conjecture in the case of P2P file-sharing communities, we define two measures of the amount of collective good provided by the community. The first is the number of unique file available that measures the size of the catalog. The second variable is the average number of sharers per unique file that is a proxy for the availability of the files in the community. As the number of sharers per file increases, each downloader has more sources available to get the different pieces of this file.



**Figure 1: Typology of communities regarding the number of unique files, file sharers per unique files and the size**

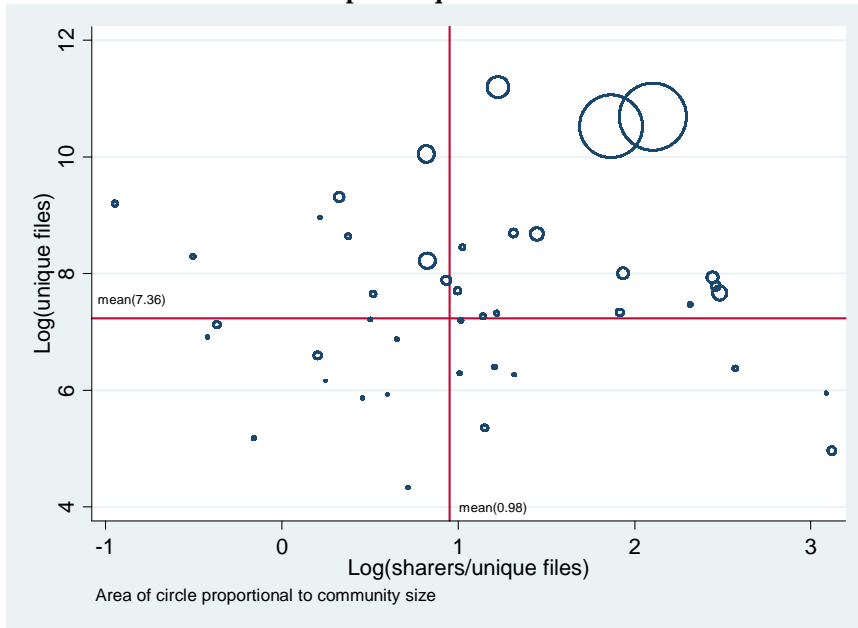


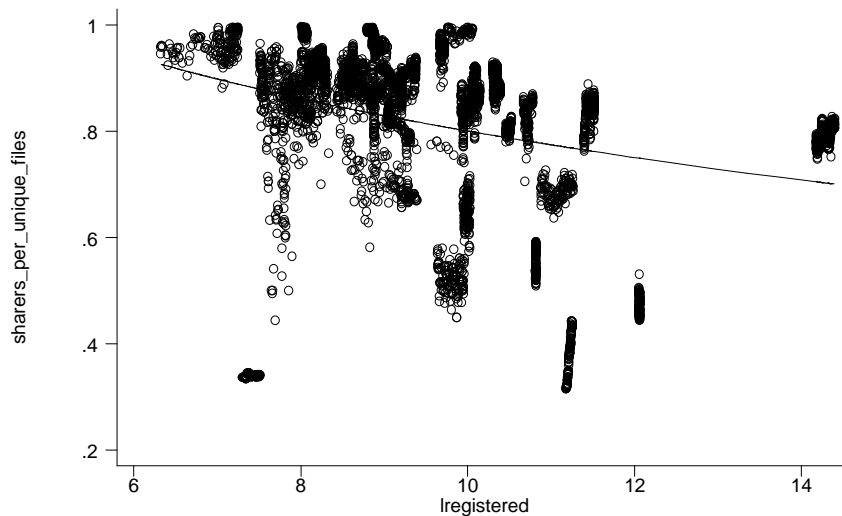
Figure 1 illustrates the relationship between the size of the catalog of a community and the number of file sharers per unique file. Moreover, each circle is proportional to the size of the community. The largest circles are concentrated in the north-east quarter. The bigger community tends to provide a larger catalog and number of sharers per file.

Figure 2 displays the relationship between the proportion of file sharers in a community (sharers/peers) and the number of registered users. Apparently, the individual propensity to contribute tends to decrease as the size of the community rises.

**Figure 2: relationship between proportion of file sharers per peers and communities size**

$$\text{sharers per unique files} = 1.430 - 0.274 \log \text{registered}$$

$$r^2 = 0.078 \quad \text{RMSE} = 0.152 \quad n = 5097$$



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 design and behavioral rules. These differences are controlled through several dummies in our econometric regressions.

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.

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 trackers in the bitTorrent galaxy by searching each tracker's name on mininova.org<sup>9</sup>. 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*<sup>10</sup>: "*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

---

<sup>9</sup> Mininova is the largest torrent search engine with more than 3 billions of visitors per day

<sup>10</sup> Torrent Fries is one of a rare site dedicated to the running of a tracker.

admission (private) and for downloading (control) have less registered members and a smaller catalog, but exhibit a higher proportion of file sharers (ratio). Specialized communities are also characterized by a higher ratio as well as more unique files and membership.

**Table 3: Mann-Whitney test for identity of distribution**

	# communities	File sharers	File downloaders	Ratio	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
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 model and results

In this section, we present the econometric models used to analyze the impact of group size on the amount of collective good provided by the community (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). We will focus on the marginal effect of additional members to determine the efficient size of these communities and their self-sustainability in a dynamic context.

The dependent variables are estimated under three different specifications for the registered members (log, linear and with a quadratic term). The log models are given by:

$$\log(\text{unique files}_{it}) = \beta_0 + \beta_1 \log(\text{registered}_{it}) + \beta_2(\text{control}_i) + \beta_3(\text{private}_i) + \beta_4(\text{saturated}_i) + \beta_5(\text{advertised}_i) + \beta_6(\text{specialized}_i) + \varepsilon_{it} \quad (M1)$$

$$\log(\text{files shared per unique file}_{it}) = \beta_0 + \beta_1 \log(\text{registered}_{it}) + \beta_2(\text{control}_i) + \beta_3(\text{private}_i) + \beta_4(\text{saturated}_i) + \beta_5(\text{advertised}_i) + \beta_6(\text{specialized}_i) + \varepsilon_{it} \quad (M2)$$

$$\log(\text{prop.of files shared}_{it}) = \beta_0 + \beta_1 \log(\text{registered}_{it}) + \beta_2(\text{control}_i) + \beta_3(\text{private}_i) + \beta_4(\text{saturated}_i) + \beta_5(\text{advertised}_i) + \beta_6(\text{specialized}_i) + \varepsilon_{it} \quad (M3)$$

(M1) and (M2) aim to estimate the determinants of the overall amount of voluntary contributions: *unique files<sub>it</sub>* measures the number of original files uploaded and shared in the community *i* at time *t* (the size of the catalog) while *sharers per file<sub>it</sub>* is the average number of peers that share the same file (the availability of the catalog). (M3) estimates the drivers of cooperative behavior in each community. The propensity to cooperate is measured by the ratio of sharers to peers (*prop.of file sharers<sub>it</sub>*). The other variables are time invariant dummies that control the features and rules of the communities.

As we have to deal with a time-series cross-sectional (TSCS) dataset 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 heterogeneity in the size of communities makes more likely the existence 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 11pm were correlated with observations at 11 am, we suspected the presence of residuals serial correlation. This was confirmed by a test for autocorrelation in panel-data (Woodridge, 2002)<sup>11</sup> 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. Using the Panel Corrected Standard Errors (PCSE) estimators proposed (Beck and Katz, 1995) would have been a possibility however it is less efficient for panel data when temporal dimension exceeds individual dimension (Chen, Lin and Reed, 2006).

Table 4, available in annex 2, displays the estimates for the three specifications of the models (M1). The columns (1a with log specification), (2a with linear specification) and (3a

---

<sup>11</sup> Drukker (2003) provides a simple program to perform this test in Stata.

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, 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 (annex2): the columns (1a) (2a) and (3a) report the OLS estimates while (1b) (2b) and (3b) show the results of the FGLS regression.

All specifications show a significantly positive impact of community size (registered users) on the amount of collective good provided by the file-sharing community. However, for the unique files, the coefficient of  $\text{Log}(\text{registered})$  falls to 0.05 with GLS from 0.67 with OLS. This underlines the importance to control for serial correlation in the error term.

Log specifications always provide better estimates for equations (M1) and (M2) than linear and quadratic specifications. The quantity and availability of files tend to increase with the number of members, but a decreasing rate (the coefficient of  $\text{log}(\text{registered})$  is positive but below 1). Two reasons could explain this result. It is possible that the late members in a BitTorrent community are contributing less than the early members. The new members can also impair cooperative behavior among the early members. However, the amount of voluntary contributions does not seem diminish with community enlargement.

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 contribute voluntary. The non-linear specifications (log and quadratic) fit better for the estimated model. The log-likelihood of the quadratic specification is also higher than the one of the log specification suggesting that a U curve better fit with the data. The incentives to share files decreases with the number of members, but a decreasing rate. This finding could suggest that the proportion of file sharers tends to converge to a lower bound higher than 0. This is not surprising because many of these communities incite their members to exceed a minimum ratio of uploading to downloading. We can also interpret this lower bound as the minimum level of cooperative behavior to maintain the community viable or sustainable. According to (3b), this lower bound could be equal to 25% of file sharers (a minimum of 25 sharers for 100 peers).

The findings exhibited in Table 4, 5 and 6 are, to some extent, rather consistent with Asvanund et al. (2004) conclusions. The individual incentives to contribute voluntary tend to decrease with the size of a P2P community. However, the collective contributions tend to increase as the community size rises. In other terms, even if each member is sharing less

content, the size of the catalog as well as the availability of files increase with the number of members among the community. This seems to indicate that the provision of a BitTorrent community is a pure inclusive good (Chamberlin, 1974).

Olson (1965) underlined that large groups can overcome free-riding issues by providing private incentives or exclusive services only to the active members. Some of the dummies in our estimated models can be analyzed as private incentives (control, private). Following (2b), the most reliable estimation of table 4 and 5, only “private” communities seem to provide a higher amount of voluntary contributions whereas “control” communities have the opposite effect. These results are not so surprising, “private” communities are the most closed form of organisation which ensure that users are handpick based on their ability to contribute to the collective good. Laying down a rule of download based on the ratio of data uploaded over data downloaded (the sharing ratio) can involve strategic behaviour harmful for the size of the catalog. Indeed, the easiest way to rapidly increase the share ratio is to share a content which is highly requested. As a consequence, shared sources are concentrated on few files and limit the extend of the catalog. However, table 6 shows that the control of the sharing ratio can have a positive impact on the proportion of files shared and thus “everything equal otherwise” on the proportion of sharer.

The quantity and availability of files are higher when a community regulates more strictly the admission of new members and monitors file-sharing behavior. To select users at entry and/or to control their behaviour once they are in the community it’s necessary, not only to prevent it from the most opportunistic behaviour, but also to ensure a good segmentation of interests and preferences. Specialized communities seem also 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) also increases the proportion of file shared in the community but doesn’t have a positive impact on the amount of collective good provided.

### *Interaction variable*

Until now, regressions have only considered the impact of regulation policies on the level of the dependant variables. Rules and incentives can also be involved in the relationship between voluntary contribution and group size, to this end we implement interaction variable in the most reliable estimation of table 4 (2b) ,5 (2b) and 6 (3b). Estimations are given by:

$$\begin{aligned} \log(\text{unique files}_{it}) = & \beta_0 + \beta_1 \log(\text{registered}_{it}) + \beta_2(\text{control}_i) + \beta_3(\text{control}_i * \log(\text{registered}_{it})) + \beta_4(\text{private}_i) \\ & + \beta_5(\text{private}_i * \log(\text{registered}_{it})) + \beta_6(\text{advertised}_i) + \beta_7(\text{advertised}_i * \log(\text{registered}_{it})) + \beta_8(\text{specialized}_i) \\ & + \beta_9(\text{specialized}_i * \log(\text{registered}_{it})) + \varepsilon_{it} \end{aligned} \quad (M4)$$

$$\begin{aligned} \log(\text{files shared per unique file}_{it}) = & \beta_0 + \beta_1 \log(\text{registered}_{it}) + \beta_2(\text{control}_i) + \beta_3(\text{control}_i * \log(\text{registered}_{it})) \\ & + \beta_4(\text{private}_i) + \beta_5(\text{private}_i * \log(\text{registered}_{it})) + \beta_6(\text{advertised}_i) + \beta_7(\text{advertised}_i * \log(\text{registered}_{it})) \\ & + \beta_8(\text{specialized}_i) + \beta_9(\text{specialized}_i * \log(\text{registered}_{it})) + \varepsilon_{it} \end{aligned} \quad (M5)$$

And considering the quadratic specification when the dependant variable is the proportion of files shared gives

$$\begin{aligned} \log(\text{proportion of files shared}_{it}) = & \beta_0 + \beta_1(\text{registered}_{it}) + \beta_2(\text{registered}_{it}^2) + \beta_3(\text{control}_i) \\ & + \beta_4(\text{control}_i * \text{registered}_{it}) + \beta_5(\text{control}_i * \text{registered}_{it}^2) + \beta_6(\text{private}_i) + \beta_7(\text{private}_i * \text{registered}_{it}) \\ & + \beta_8(\text{private}_i * \text{registered}_{it}^2) + \beta_9(\text{advertised}_i) + \beta_{10}(\text{advertised}_i * \text{registered}_{it}) \\ & + \beta_{11}(\text{advertised}_i * \text{registered}_{it}^2) + \beta_{12}(\text{specialized}_i) + \beta_{13}(\text{specialized}_i * \text{registered}_{it}) \\ & + \beta_{14}(\text{specialized}_i * \text{registered}_{it}^2) + \varepsilon_{it} \end{aligned} \quad (M6)$$

The result of M4, M5 and M6 estimates are displayed in annex 3. The table 7 summarizes the net effect of each dummy on the slope of the regression.

The impact of regulation policies on the relationship between group size and the amount of voluntary contribution is ambiguous. Being “control” “advertised” or “specialized” lower the benefit expected from an additional user in term of contribution to the catalog. This result aims to consider the upload of a new file in the community as a very singular form of contribution. Indeed, to ensure that files are original and of high quality, communities make the upload of new files a very costly process Moreover the more closed and demanding are the communities the highest is the cost of uploading a new file. As a result, we suppose that the production of new files is in the hand of a very narrow group of individuals who try to preserve the privilege of this form of contribution which can explain why, in these communities, additional members provide less new file.

**Table 7: Effect of regulation policies on the slope of the curve estimating the relationship between dependant variables and the size of the community.**

	Dep. Var = Unique files	Dep.var = Files Shared/ unique files	Dep.var = Proportion of. files shared
Control=0 Private=0 Advertised=0 Specialized=0	0,82 . log(registered)	0,02 . Log(registered)	-1,25E-06 . registered -3,85E-12 . registred <sup>2</sup>
Control=1 Private=0 Advertised=0 Specialized=0	0,39 . log(registered)	ns	-1,65E-06 . registered -3,52E-12 . registred <sup>2</sup>
Net effect of “control”	-	ns	-
Private=1 Control=0 Advertised=0 Specialized=0	ns	-0,28 . log(registered)	-2,12E-05 . registered + 7,3E-10 . registred <sup>2</sup>
Net effect of “private”	ns	-	+
Advertised=1 Control=0 Private=0 Specialized=0	0,7 . log(registered)	0,31 . log(registered)	ns . registered + 9,09E-13 . registred <sup>2</sup>
Net effect of “advertised”	-	+	+
Specialized=1 Control=0 Private=0 Advertised=0	0,46 . log(registered)	0,16 . log(registered)	-5,78E-06 . registered + 6,02E-11 . registred <sup>2</sup>
Net effect of “specialized”	-	+	+

The result is more balanced when considering the availability of files as a dependant variable, being part of an “advertised” and/or a “specialized” community increases the benefit of an additional user in term of availability of files.

The most positive impact of incentives mechanisms concerns the proportion of voluntary contribution in the community. Only the “control” dummy has a negative impact on the quality of voluntary contribution when the size of the community rises. Otherwise, “private”, “advertised” and “specialized” communities seem to provide best incentives to favor voluntary contribution in large group. The proportion of files shared even increases



exponentially with the community enlargement when there is a specialization in a particular category of content.

We also have to note that interaction effects exhibit in table 7 are not exclusive but cumulative which mean, for example, that a community specialized in a particular content and which advertises its catalog in public meta-search engine increases the benefit of an additional members on the proportion of files shared as compared to a community which only provides one of the two incentive.

## 5. Conclusion

In this paper we have 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 have collected data on the activity of 42 private and semi-private bitTorrent communities. Our finding reveals that the collective provision in these communities can be analyzed as a pure public good. Indeed, the amount of collective good increases with the number of registered users whereas the individual propensity to contribute decreases with the community enlargement. We also show that the rules designed by the administrators of the community have a significant impact on the performance of the community and on their sustainable size.

## References:

- Adar, E. and B. Huberman (2000) "Free Riding on Gnutella", *First Monday* 5.
- Asvanund, A., K. Clay, R. Krishnan and M.D. Smith (2004) "An Empirical Analysis of Network Externalities in Peer-to-Peer Music-Sharing Networks", *Information Systems Research* 15, 155-174.
- Asvanund A., R. Krishnan, M.D. Smith and R. Telang (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. and J.N. Katz (1995), "What to Do and Not to Do With Time-Series Cross Section Data", *The American Political Science*, Vol. 89, 634-647.
- Bergstrom, T., Blume, L., Varian, H., 1986. On the private provision of public goods. *Journal of Public Economics* 29, 25-49.

- Bharambe A.R., C. Herley, and V.N. Padmanabhan (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., S. Lin and W.R. Reed (2006) “A Monte Carlo Estimation of the Efficiency of the PCSE Estimator” Working paper, Department of Economics, College of Business and Economics University of Canterbury.
- Cohen B. (2003), “Incentives Build Robustness in BitTorrent”, *proc. of International Workshop on peer-to-peer Systems*.
- Dang-Nguyen, G., and T. Pénard (2007) “Network Cooperation and Incentives Inside Online Communities” in *Internet and Digital Economy* by E. Brousseau and N. Curien (Eds.), Cambridge University Press.
- Drukker, D.M. (2003) “Testing For Serial-Correlation in Linear Panel-data Models”, *The Stata Journal* 3, 168-177.
- Esteban, J., and D. Ray (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., L. Wang, D Zhang and Y. Liu (2007) “A Dynamic Quota-Based Peers Selection in Bit Torrent”, *Proc. of the Sixth International Conference on Grid and Cooperative Computing*.
- Isaac, R.M., and J.M. Walker, (1988) “Group Size Effect in Public Good Provision: The Voluntary Contributions Mechanism”, *The Quaterly Journal of Economics* 103, 179-99.
- Isaac, R.M., J.M. Walker, and A.W. Williams (1994) “Group Size and the Voluntary Provision of Public Goods Experimental Evidence Utilizing Large Group”, *Journal of Public Economics* 54, 1-36.
- Krishnan, R., M.D. Smith, Z. Tang and R. Telang (2007) “Digital Business Models for Peer-to-Peer Networks: Analysis and Economic Issue,” *Review of Network Economics*, 6, 194-213.
- Krishnan, R, M.D. Smith and R. Telang (2003) “The Economics of Peer-to-Peer Networks”, *Journal of Information Technology Theory and Applications*, 5:3, 31-44.

- Legout A., G. Urvoy-Keller and P. Michiardi (2005), “Rare First and Chock Algorithms are Enough”, Proc. 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., and R.E Ames (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., and A. Temimi (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.
- Yue Y., C. Lin and Z. Tan (2006), “Analysing the Performance and fairness of Bit-Torrent Like Protocol Using a general Fluid Model”, *Computer Communication*, Vol. 29, 3946-3956.
- Zhang, M., and F. Zhu (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	speciali zed	cont rol	adverti sed	saturated
<a href="http://www.captain-tracker.fr/index.php">http://www.captain-tracker.fr/index.php</a>	1	general	0	0	1	0	0
<a href="http://www.sharing-torrents.com/">http://www.sharing-torrents.com/</a>	2	general	0	0	1	1	0
<a href="http://leparrain.mine.nu/torrents.php">http://leparrain.mine.nu/torrents.php</a>	3	general	0	0	0	0	0
<a href="http://www.unlimited-tracker.net/">http://www.unlimited-tracker.net/</a>	4	general	0	0	1	1	0
<a href="http://www.nhltorrents.co.uk/">http://www.nhltorrents.co.uk/</a>	5	sport	0	1	1	1	0
<a href="http://xtremewrestlingtorrents.net/statistic.php">http://xtremewrestlingtorrents.net/statistic.php</a>	6	sport	0	1	1	0	0
<a href="http://www.dimeadozen.org/index.php">http://www.dimeadozen.org/index.php</a>	7	music	0	1	0	1	1
<a href="http://www.indietorrents.com/index.php">http://www.indietorrents.com/index.php</a>	8	music	1	1	1	0	0
<a href="http://shnflac.net/index.php">http://shnflac.net/index.php</a>	9	music	0	1	0	0	1
<a href="http://jamtothis.com/">http://jamtothis.com/</a>	10	music	0	1	1	0	0
<a href="http://www.browntracker.net/browse.php">http://www.browntracker.net/browse.php</a>	11	music	0	1	0	0	0
<a href="http://anvilofsound.com/">http://anvilofsound.com/</a>	12	music	0	1	0	1	0
<a href="http://mixes.dfx.at/index.php">http://mixes.dfx.at/index.php</a>	13	music	0	1	1	0	0
<a href="http://asiandvdclub.org/">http://asiandvdclub.org/</a>	14	cinéma	0	1	1	0	0
<a href="http://alt.bitworld.to/browse.php">http://alt.bitworld.to/browse.php</a>	15	general	0	0	1	1	1
<a href="http://www.araditracker.com/">http://www.araditracker.com/</a>	16	general	0	0	1	0	0
<a href="http://www.titaniumtorrents.net/">http://www.titaniumtorrents.net/</a>	17	general	1	0	1	1	0
<a href="http://dididave.com/">http://dididave.com/</a>	18	general	0	0	1	0	0
<a href="http://www.quebectorrent.com/">http://www.quebectorrent.com/</a>	19	general	0	0	1	1	0
<a href="http://cinemageddon.org/">http://cinemageddon.org/</a>	20	cinéma	0	1	1	1	0
<a href="http://www.blades-heaven.com/index.php">http://www.blades-heaven.com/index.php</a>	21	general	0	0	1	1	0
<a href="http://www.puretna.com/">http://www.puretna.com/</a>	22	adult	0	1	0	0	0
<a href="http://www.kingdomxxx.com/">http://www.kingdomxxx.com/</a>	23	adult	0	1	0	0	0
<a href="http://www.emporium.us/">http://www.emporium.us/</a>	24	adult	0	1	1	0	0
<a href="http://www.pornevo.com/">http://www.pornevo.com/</a>	25	adult	0	1	1	0	1
<a href="http://www.cinema-obscura.com/">http://www.cinema-obscura.com/</a>	26	adult	0	1	1	0	0
<a href="http://www.underground-gamer.com/">http://www.underground-gamer.com/</a>	27	vidéo game	0	1	1	1	1
<a href="http://www.pleasuredome.org.uk/">http://www.pleasuredome.org.uk/</a>	28	vidéo game	0	1	1	1	0
<a href="http://my-gamebox.com/">http://my-gamebox.com/</a>	29	vidéo game	0	1	1	1	0
<a href="http://thepeerhub.com/">http://thepeerhub.com/</a>	30	general	0	0	1	1	0
<a href="http://bitnation.com/index.php">http://bitnation.com/index.php</a>	31	general	1	0	1	1	1
<a href="http://p2pworld.ulmb.com/">http://p2pworld.ulmb.com/</a>	32	general	0	0	1	1	0
<a href="http://torrent-hackers.co.uk/">http://torrent-hackers.co.uk/</a>	33	general	0	0	1	1	0
<a href="http://www.torrentbits.ro/index.php">http://www.torrentbits.ro/index.php</a>	34	general	0	0	1	0	0
<a href="http://www.sport-scene.net/">http://www.sport-scene.net/</a>	35	sport	0	1	1	0	0
<a href="http://www.sportbit.org/">http://www.sportbit.org/</a>	36	sport	0	1	1	0	1

<a href="http://www.prosporttorrents.net">http://www.prosporttorrents.net</a>	37	sport	0	1	1	1	0
<a href="http://www.mamitracker.com/index.php">http://www.mamitracker.com/index.php</a>	38	general	0	0	0	1	0
<a href="http://zombitracker.the-zomb.com/">http://zombitracker.the-zomb.com/</a>	39	music	0	1	1	0	0
<a href="http://cinematik.net/">http://cinematik.net/</a>	40	cinéma	1	1	1	0	1
<a href="http://www.zinebytes.org/">http://www.zinebytes.org/</a>	41	e-learning	0	1	1	0	1
<a href="http://www.mytracker.ru/index.php">http://www.mytracker.ru/index.php</a>	42	general	0	0	1	0	0
<a href="http://linuxmafia.net/">http://linuxmafia.net/</a>	43	general	0	0	1	1	0
<a href="http://zerotracker.com/index.php">http://zerotracker.com/index.php</a>	44	general	0	0	1	0	0

## Annex 2

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

	Dep. Var= log(unique files)					
	OLS (1a)	GLS (1b)	OLS (2a)	GLS (2b)	OLS (3a)	GLS (3b)
<b>Log(registered)</b>			0.67 (73.51)***	0.05 (8.66)***		
<b>Registered</b>	2.43e-02 (132.27)***	1.06e-02 (7.66)***			1.70e-01 (18.61)***	3.72e-04 (0.77)
<b>Registered<sup>2</sup></b>					-8.88e-08 (16.09)***	1.82e-08 (3.17)***
<b>Control</b>	-1772.58 (5.81)***	-2294.24 (10.65)***	-0.37 (7.13)***	-0.56 (11.46)***	-3762.61 (14.75)***	-8653.65 (5.39)***
<b>Private</b>	-1158.42 (5.59)***	744.00 (3.04)***	0.72 (22.97)***	0.62 (11.62)***	1458.68 (10.44)***	-13096.13 (2.11)**
<b>Advertised</b>	-2589.04 (6.43)***	-1202.97 (10.12)***	-0.13 (3.44)***	-0.28 (6.50)***	-812.10 (3.87)***	2918.08 (1.65)*
<b>Specialized</b>	-2528.10 (5.71)***	-1116.86 (9.51)***	-0.25 (6.97)***	0.27 (6.51)***	-2849.88 (8.56)***	6081.48 (3.58)***
<b>Constant</b>	8035.01 (16.11)***	4933.10 (21.59)***	1.47 (13.22)***	7.59 (93.15)***	5386.79 (14.72)***	-12813.23 (7.27)***
<b>Observations</b>	5097	5097	5097	5097	5097	5097
<b># of communities</b>		42		42		42
<b>R2</b>	0.43		0.51		0.6	
<b>Log-likelihood</b>		-23778		13724		-20889

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

Dep. Var= log(files shared/unique files)						
	OLS (1a)	GLS (1b)	OLS (2a)	GLS (2b)	OLS (3a)	GLS (3b)
<b>Log(registered)</b>			0.18 (25.03)***	0.16 (94.67)***		
<b>Registered</b>	6.29e-07 (5.89)***	1.74e-06 (26.26)***			6.37e-06 (6.31)***	9.04e-06 (18.09)***
<b>Registered<sup>2</sup></b>					-3.50614e-12 (5.82)***	-4.67296e-12 (14.56)***
<b>Control</b>	-3.66078 (12.77)***	-1.56912 (13.41)***	-0.31976 (7.24)***	-0.15571 (11.70)***	-3.74 (13.05)***	-2.83 (22.51)***
<b>Private</b>	0.19 (1.05)	-0.58 (6.05)***	0.21 (5.51)***	0.07 (3.37)***	0.30 (1.61)	-0.62 (6.30)***
<b>Advertised</b>	-2.03 (10.82)***	-1.22 (23.20)***	-0.45 (14.52)***	-0.46 (52.55)***	-1.96 (10.22)***	-1.12 (21.67)***
<b>Specialized</b>	-1.11 (5.47)***	-0.55 (11.88)***	-0.19 (6.72)***	-0.17 (18.37)***	-1.12 (5.53)***	-0.65 (14.20)***
<b>Constant</b>	9.23 (23.50)***	5.91 (44.97)***	-0.07 (0.65)	-0.12 (4.90)***	9.12 (22.98)***	7.04 (55.96)***
<b>Observations</b>	5097	5097	5097	5097	5097	5097
<b># of communities</b>		42		42		42
<b>R2</b>	0.11		0.18		0.11	
<b>Log-likelihood</b>		-11459		-3809		-11351

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

	Dep. var= log(proportion of file sharers)					
	OLS (1a)	GLS (1b)	OLS (2a)	GLS (2b)	OLS (3a)	GLS (3b)
<b>Log(registered)</b>			-0.03 (11.58)***	-0.03 (33.42)***		
<b>Registered</b>	-7.46e-10 (-0.18)	-1.34e-08 (5.14)***			-1.44e-06 (29.55)***	-1.51e-06 (43.52)***
<b>Registered<sup>2</sup></b>					8.79e-13 (29.43)***	9.20e-13 (43.87)***
<b>Control</b>	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)***
<b>Private</b>	0.04092 (8.86)***	-0.00 (0.76)	0.04 (8.57)***	0.00 (0.51)	0.01 (3.44)***	-0.02 (8.19)***
<b>Advertised</b>	0.08 (18.88)***	0.07 (45.32)***	0.11 (15.08)***	0.06 (27.63)***	0.06 (14.52)***	0.04 (19.40)***
<b>Specialized</b>	0.08 (16.90)***	0.04 (19.52)***	0.15 (20.10)***	0.07 (22.63)***	0.08 (18.91)***	0.03 (16.58)***
<b>Constant</b>	0.64 (73.88)***	0.69 (216.95)***	-0.26 (8.04)***	-0.08 (7.57)***	0.67 (74.38)***	0.70 (204.69)***
<b>Observations</b>	5097	5097	5097	5097	5097	5097
<b># of communities</b>		42		42		42
<b>R2</b>	0.16		0.22		0.28	
<b>Log-likelihood</b>		5414		4193		5676



## Annex 3

**Table 8: Estimation of equation M4, M5 and M6 using FGLS estimation**

	FGLS Dep. Var= Unique files		FGLS Dep.var = Files Shared/ unique files		FGLS Dep.var = Prop. files shared	
<b>Log(registered)</b>	0,82 (27,17)***		0,02 (1,88)*		-0,02 (7,92)***	
<b>Registered</b>		-1,61E-01 (17,35)***		2,15E-05 (2,25)**		-1,25E-06 (3,95)***
<b>Registered<sup>2</sup></b>		3,25E-06 (73,41)***		-8,18E-11 (1,66)*		-3,85E-12 (2,88)***
<b>Control</b>	3,6241 (16,70)***	-865,78 (5,61)***	-0,10287 (-1,06)	-4,25 (23,13)***	0,65953 (26,03)***	0,16 (45,71)***
<b>Log(registered) X control</b>	-0,43 (21,02)***		0.0 (-0,31)		-0,06 (26,60)***	
<b>Registered X control</b>		4,02E-03 -0,82		-1,42E-05 (4,44)***		-4,03E-07 (2,41)**
<b>Registered<sup>2</sup> X control</b>		-6,98E-10 -0,23		8,52E-12 (4,29)***		3,31E-13 (3,34)***
<b>Private</b>	0,97 (4,07)***	-244,04 (2,24)**	2,8 (7,95)***	1,38 (2,73)***	-0,05 (-1,04)	0,09 (5,62)***
<b>Log(registered) X private</b>	-0,04 (-1,64)		-0,3 (7,89)***		0,01 (-1,57)	
<b>Registered X private</b>		1,96E-01 (7,90)***		-1,98E-04 (2,08)**		-1,99E-05 (7,09)***
<b>Registered<sup>2</sup> X private</b>		-6,46E-06 (7,25)***		4,51E-10 -0,13		7,34E-10 (7,33)***
<b>Advertised</b>	0,99 (3,23)***	-1264,93 (13,77)***	-3,09 (45,52)***	-3,26 (30,63)***	-0,09 (4,12)***	0,05 (14,40)***
<b>Log(registered) X advertised</b>	-0,12 (3,87)***		0,29 (41,46)***		0,01 (6,76)***	
<b>Registered X advertised</b>		9,55E-02 (10,74)***		2,47E-04 (19,60)***		-4,53E-06 (13,35)***
<b>Registered<sup>2</sup> X advertised</b>		9,41E-07 (8,95)***		-2,92E-09 (24,55)***		6,41E-11 (18,01)***
<b>Specialized</b>	3,61 (12,51)***	-136,98 -1,5	-1,42 (18,94)***	-1,03 (11,42)***	-0,23 (10,14)***	0,04 (12,63)***
<b>Log(registered) X specialized</b>	-0,36 (12,13)***		0,14 (17,78)***		0,04 (14,28)***	
<b>Registered X specialized</b>		1,66E-01 (20,86)***		4,03E-06 (-0,45)		-2,82E-07 (-1,18)
<b>Registered<sup>2</sup> X specialized</b>		-3,24E-06 (73,30)***		6,71E-11 -1,37		4,76E-12 (3,62)***
<b>Constant</b>	0,1 (-0,37)	3,19E+03 (18,41)***	1,09 (9,20)***	8,60E+00 (40,88)***	-0,16 (5,22)***	7,07E-01 (153,24)***
<b>Observations #</b>	5097	5097	5097	5097	5097	5097
<b>communities</b>	42	42	42	42	42	42
<b>Log-likelihood</b>	13262	-27585	-3857	-10728	4776	6280

