

# P2P and cinematic movie distribution in Hungary

Working paper. Draft date: 7/15/2009

*Do not cite* or circulate without written permission from authors.

Comments welcome. Contact [bodo@mokk.bme.hu](mailto:bodo@mokk.bme.hu)

*Bodó Balázs, Budapest University of Technology and Economics<sup>1</sup>*

*Lakatos Zoltán, Eötvös Lóránd University of Arts and Sciences*

## Abstract

In this paper we analyze data collected from three of the biggest Hungarian bittorrent based file-sharing communities between 2008 May-June, and Hungarian cinema distribution data from the same period. We asked if the number of downloads for any given film correlates with ticket sales, revenues, the number of cinemas where the film was shown or any other statistical data on the traditional movie distribution infrastructure we had access to. Our results show that 3 out of 4 downloaded films were not available in Hungarian cinemas and only 1 out of 30 downloaded films actually screened when it was downloaded.

We have found that it is the time difference between releases that (and only that) defines to what extent a film is downloaded.

Although we could not find a direct, causal effect of the marketing related variables to the number of downloads, we have established that it marketing power defines what gets uploaded to p2p networks.

Finally, we have found no causal connection whatsoever between any of the cinematic and online popularity measures.

---

<sup>1</sup> This work started in 2006 with the publication of (Bodó, Halácsy, Korsós, Prekopcsák, & Szalai, 2007). Our colleagues of that study, Prekopcsák Zoltán and Korsós Milán of Kitchen Budapest continued to work with us in this study and proved to be of invaluable help throughout the work. Ducsay Tamás and Halász Péter, students at the Budapest University of Technology and Economics (BUTE) were instrumental in designing, implementing and running the crawler. Without them it would have been much more difficult to accomplish this study. We are also indebted to Veszeloovszki Zsolt and his staff at the online program guide port.hu for letting us use their database. Vincze Gábor, a PhD candidate at BUTE has helped us with infrastructure and with his insights for which we are grateful. And last but not least we would like to thank those anonymous hundreds who helped us in pairing torrents with movie titles. Thank You.

## Introduction

Though the unauthorized reproduction of someone else's creative output (for fame or for profit) is as old as creativity itself (Alford, 1995; Lendvai, 2008), it was the advent of the technologies of mass reproduction which has made it an everyday, mass phenomenon. Free riding on an already existing investment is always profitable. Even if there are firm institutions to curb this practice (such as copyright laws and treaties), the potential gains are so huge, that up until the era of file-sharing, we hardly see anyone resisting the temptation of reaping huge sums by copying without asking first. Even though the reasons that called pre-internet pirates into existence vary greatly (Bodó, in press) there is one thing that is common in all of them: their love of and quest for profit.

Online, peer-to-peer file swapping is unique in the sense that those who participate in it are not interested in the potential monetary gains. Of course there are many in the digital age who do unauthorized copying of CDs, DVDs, software, etc for profits, but this type of activity dwarfs in comparison with the number of those individuals who engage in the non-profit gift economy of online file-sharing.

If monetary incentives do not explain p2p file-sharing, we have to examine other factors that drive this activity. In this paper we hope to shed light on one factor among the many which may explain online file-sharing. That factor is the failure of traditional markets. By traditional markets we mean the well-established institutions that engage in the marketing, distribution, retailing, lending of cultural goods: libraries, cinemas, broadcasters, video rental outlets, etc. By the failure of these institutions we mean such a deficiency (lack of retail outlets, price, lack of variety, etc) that leaves a sizeable demand on the market unserved.

With the identification of such failures we hope to serve several aims. First we would like to explain why p2p file-sharing - something that many actors from the traditional markets see an undesirable, or even as criminal activity -, is so popular. Second, we hope that we can urge the traditional actors to do whatever they can to improve on those points in their businesses that contribute the most to these market failures. Third, we hope to help policy makers in devising a consumer and citizen friendly policy environment in which citizens, consumers on the cultural markets do not get prosecuted and punished for acting up if the traditional actors don't, or - due to structural deficiencies -, simply cannot.

In this paper we describe our findings from measuring the traffic of movies on three Hungarian, bittorrent based file-sharing networks between May and June, 2008. During this period we have tracked the new titles that appeared on these networks, the location of individual users and the instances of users downloading, seeding, uploading these titles. We are therefore able to tell who, downloaded what from where for how long. We compare this dataset to another set of databases that track the performance of the traditional movie distribution system: cinemas. We have mapped the geographical distribution of the cinema

network, and analyzed the distribution patterns that are defined by the producers and distributors of audiovisual works.

By matching these two datasets we were able to estimate to what extent file-sharing traffic can be explained by the way traditional markets operate.

## Previous research on the effects of file-sharing on traditional markets

We see file sharing networks as markets, with their own logics of supply and demand. It is clear, that the workings of these peer-to-peer markets differ greatly from the rules of traditional markets.

The source of these differences may be:

- **Price.** The zero price at file-sharing network is much lower than the price set by the competition of street pirates and lower than the monopolist price set by the traditional producers/distributors. Though traditional actors are convinced that one cannot compete with something that is free, it is clear that price alone cannot explain the existence of p2p markets.
- **The timing of putting goods on various markets.** While traditional actors utilize release windows to discriminate among different markets to extract the highest profits, file sharing networks honor the quickest releasers the highest, fuelling a race that makes new titles available in the same time at all territories worldwide. This competition can significantly lower the amount of time a potential consumer needs to wait before he or she can enjoy a work. As citizens in the global media sphere, consumers around the globe (from India to Brazil) are exposed to the marketing efforts in the most important western markets. This exposure generates some demand on markets which in most cases need to wait a considerable amount of time before the producers are willing to sell their wares there. P2p markets offer an instant gratification in such cases.
- **The size of the catalogue, the variety of supply.** Digital platforms, even though they promise to solve the bottlenecks of physical distribution, are yet to solve the problem of titles being out of print. Legal hurdles, lack of resources, business considerations hinder the release of all back catalogs in digital format. On the other hand, as file-sharers do the digitization, storage, transmission of titles they deem worthy, there is a good chance that any title that has at least one person to care for, will be available, bringing out of print works back into the market (Bodó, 2006; Freeman, November 2008)
- **File sharing is a social activity.** (Becker & Clement, 2006; Condry, 2004; J. Cooper & Harrison, 2001; M. N. Cooper, March 2005; Giesler &

Pohlmann, 2003; Huang, 2005; Hunter & Spitz, 2003; Keenan, November 2008; Manuel, 1993; Marshall, 2004; Rojek, 2005; Strahilevitz, 2003) In fact, the first file sharing service, Napster was created to solve the problem of fans chatting about music, but not able to show to each other what they are talking about. File sharing networks are indeed online communities organized around special interest p2p hubs.

Since 1999, when Napster, the first p2p technology appeared on the Internet, a growing body of research has emerged on the impact of file-sharing on traditional markets of cultural goods. (Becker & Clement, 2003; Blomqvist, Eriksson, Findahl, Selg, & Wallis, ; Dejean, 2008; "Digital Music Report 2006," 2007; Domon & Nakamura, 2007; "The Economic Impact of Counterfeiting and Piracy," 2008; Givon, Mahajan, & Muller, 1995; Ram D. Gopal, Bhattacharjee, & Sanders, 2006; R. D. Gopal & Sanders, 1998; Gu & Mahajan, 2004; Huang, 2005; IFPI, 2001, 2006; Liebowitz, 2006; Oberholzer-Gee & Strumpf, 2007; Peitz & Waelbroeck, 2006a, 2006b; Rob & Waldfogel, 2006; Sheikh, Rashed, Qudah, & Peace, 2006; Zentner, 2006) The results are more than ambiguous: there are (mostly industry sponsored) studies that link file-sharing to massive economic losses in the cultural industries, while others find little or no correlation between file-sharing activity and sales data. Still others find positive effects of file-sharing in the case of certain groups of artists, and in relation to overall social welfare. This variety of often contradicting findings only demonstrates that it is impossible to take file-sharing out from those cultural, economic, legal, social contexts in which the users of these services are situated. File-sharing *per se* might be a truly networked, global phenomena, but its impact on the traditional markets of culture are as local as those markets and its customers are.

We would like to add to this body of research by digging into the workings of a segment of the cultural markets of a post-communists country 20 years after its re-integration into the global flows of capital and culture.

## **Why cinemas and file-sharing?**

Though technically it would have been possible to measure the flow of any cultural goods swapped online, we have settled on focusing on films for several reasons.

Conducting such a study requires access to a wide variety of data beyond file-sharing traffic alone. In order to be able to assess the impact of file-sharing on traditional distribution channels, one naturally needs an exact picture of those markets: what is being sold on the market, at what price, where, for how long, with what success. In other words one needs detailed statistics on the production/distribution/consumption patterns on the traditional markets. The studies mentioned above rely heavily on publicly or commercially available data

on the workings of the given traditional market they are investigating. This is possible only because such data is collected and published by either governmental organizations, trade groups or third party organizations.

In Hungary there are very few such data sources. The market research databases are completely missing or are still underdeveloped, and even if individual members of the trade do have data on the shelf life of their wares, they are reluctant to share it with anyone. The problem of data scarcity on the working of traditional markets was the foremost factor determining which aspect of file-sharing we should deal with. We could not acquire access to any meaningful data on the music market: music publishers, industry members do not release useful data, and there is no third party data available either. On the other hand we were able to get access to the database of port.hu, an online program guide which contains screening information on each and every cinema in the country from 2000 onwards.

The second factor that determined our decision to measure film traffic was methodological. We had to decide between analyzing a sample of all file-sharing traffic, and trying to analyze the whole population of a selected field. The decision between focusing on films or focusing on music was determined by several factors: the number of titles to track, certain characteristics of the users and the characteristics of file-sharing hubs we could track.

In itself alone, the sheer number of titles could have been a determining factor. According to the Internet Movie Database there are around 1 million movie, TV and entertainment titles globally, while at cddb.com which collects music published in CD format, there are more than 6 million albums and over 80 million tracks. This nearly two orders of magnitude of difference (if we take the track as the unit of consumption) is further aggravated by the specificities of the markets of these two different cultural products. In the local movie markets the mainstream US producers share the market with a few European and Hungarian films. The overwhelming majority of these titles are from the last few years (Kanzler, 2009). The number of titles on the movie market at any given time is therefore limited to a small fraction of all possible titles. According to the port.hu database there were only 11805 different titles shown in Hungarian cinemas between 2000 and 2008, and 2008 has seen only 202 new releases.

Tracking music titles would have been a much more difficult task. To put the above number in perspective, on a popular Hungarian file-sharing service examining one randomly chosen user we have found more than 53.000 tracks in her classical and jazz collection alone. There is much more room for diversity in musical tastes, due to the lower cost of music production, free access to music through various online and offline channels, etc. This means that supply and demand on the traditional and on the file-sharing markets are made up of a narrower selection of fewer titles in the case of movies, and a wider selection of a larger number of titles in the music domain. This creates a significant difference in the number of movie and music titles which can be observed at any given

time. As a result we decided on tracking movies as the number of movies to track was more manageable.

Certain demographic variables have also contributed to this decision. Hungarian users have a relatively low command of foreign languages: according to a 2005 research (Szénay, 2005) only 9 percent of the 15-44 age group holds a certificate for English as a foreign language. This fact alone limits the demand for foreign language films, but certainly does not affect demand for foreign music, further aggravating the difference described above. The importance the language dimension is stressed by the early appearance (~2001) and the huge popularity of Hungarian online fansubbing communities that translate films, Tv series and produce freely downloadable subtitles.

Finally there are certain technical aspects that make tracking movies so much easier than tracking music. The prime vehicle for movie-sharing is the bittorrent protocol, due to its comparative advantage in terms of speed compared to other protocols, such as DC++, or direct downloading from a web server. Music sharing is partly track based, where bittorrent speed and efficiency factors are lost in comparison to other protocols. Music also has other important vehicles of digital transmission such as file-hosting services (like rapidshare) and streaming. These (non-p2p) alternatives are inconvenient for sharing large movie files, therefore movie sharers tend to concentrate around a few popular file sharing hubs, among which bittorrent hubs are clearly preferred due to the speed of the network. In turn music sharers are scattered among a number of protocols and services making them much harder to track.

In conclusion: by deciding to track movies on Hungarian file-sharing networks and in cinemas we were able to gather all the film sharing data from the 3 biggest Hungarian torrent networks and compare this data to a detailed program database which tracks film distribution in cinemas.

## Changes in the Hungarian movie distribution infrastructure

To put the current state of the Hungarian movie distribution infrastructure into perspective, one needs to go back to the decade before 1989. Due to the seemingly limitless state sponsorship and a strong cultural drive of the ruling party elite (György, 2005) Hungary enjoyed a dense network of libraries, cinemas, and other cultural institutions. Most villages had some kind of a multi-functional institution, a small cultural center that served as a concert or meeting hall, but could be converted into a screening hall as well. The high number of cinema screens (and libraries) during the eighties reflects this situation.

<i>Year</i>	<i>Number of libraries</i>	<i>Cinema screens</i>
1980	10498	3624
1981	10490	3552
1982	10272	3556
1983	10010	3700
1984	9580	3794
1985	9647	3745
1986	9320	3600
1987	9049	3279
1988	8731	2943
1989	8215	2608
1990	7350	1960
1991	6585	1025
1992	5848	697
1993	5264	638
1994	4727	595
1995	4468	597
1996	4248	558
1997	4092	594
1998	3908	628
1999	3786	604
2000	3585	564
2002		498
2004		464
2007		369

Source: Central Bureau of Statistics

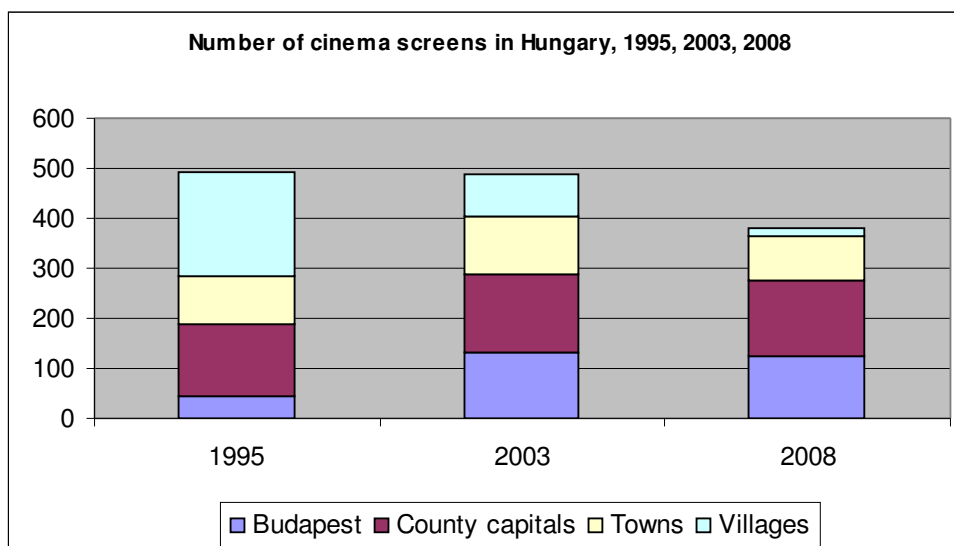
The collapse of the planned economy in 1989 put an end to the financial background of this network. The sudden collapse of the infrastructure was of course not limited of cinemas. It encompassed each and every field in the formerly state sponsored cultural industries including production and distribution of films, performances, books, etc.(Cserta, 2002)

Beyond the changes in the basic political and economic governing principles, several other factors also contributed to the post-1989 transformation of the movie distribution infrastructure.

1. With the disappearance of public funding, ticket prices rose rapidly which occurred alongside the sudden and dramatic drop in per capita GDP and therefore a drop in disposable income.
2. Profit oriented private companies replaced the access-conscious state distributors, releasing fewer copies to fewer cinemas, focusing their distribution efforts on high density markets only, denying local cinemas the possibility of displaying the latest releases.
3. Municipal owners of the cinemas had little funds to maintain, modernize the buildings and the equipment of movie theatres (Borsos, 2007. november), as a result these institutions quickly became run down and/or were privatized and put to other use.
4. In the second half of the 1990's , the rapid proliferation of cable TV, VHS and DVD players, later the cheap far-eastern home theatre sets posed a serious competition as well. At the end of 2008, 52% of the Hungarian adult population owned a PC/laptop, 59% reported owning a VHS player, 64% owned a DVD player, while 68% had a cable tv subscription. (Source: Szonda Ipsos, National Media Analysis, 2008 december)

These external factors resulted in rapid changes in the structure of the movie distribution infrastructure of the country:

- The number of screens literally decimated compared to the 1980s
- The remaining screens recessed to bigger urban centers leaving (in 2006) as much as 99% of villages, and 70% of towns without a cinema screen. On another level: in 57% of the Hungarian statistical regions there are no settlements with cinemas.(Borsos, 2007)



Source: Central Bureau of Statistics, National Film Bureau



- The screens in urban centers are more and more located in shopping malls, and operated by a handful of US owned companies. Such multiplexes controlled 49% of all screens, 50% of all seats, sold 76% of all tickets, and controlled 84% of all revenues in 2008. It goes without saying that multiplexes have a fundamental effect on what is being shown in cinemas, skewing movie supply towards popular US titles.
- Public subsidies aimed at reconstructing smaller “arthouse” cinemas that show movies outside of the mainstream culture did nothing to change the uneven distribution of cinemas and resulted in upgrading already established institutions without founding new ones. (Borsos, 2007. november)

In conclusion: in the last two decades movie theatres, along with other cultural retailers have receded to where effective, solvent demand was to be found: into urban centers. “The quick change in the economic and legal environment erodes the basic cultural supply. This is true in qualitative, content-wise terms, in terms of the physical state of infrastructure, costs of operation and in human resources, which is an especially serious problem because due to their cheap accessibility these institutions were mostly used by lower income social groups in need for an access to cultural goods.”(Bárdosi, Lakatos, & Varga, 2004) This process of regression proved to be a fatal one: the lack of solvent demand and adequate funding ruined the distribution infrastructure, and the collapse of the distribution infrastructure left those unserved who had been able to pay for these services, but who weren't numerous enough to be served economically.

The shift from independent cinemas with one or two screens to multiplexes in shopping malls also transformed the content that was shown in movie theatres. Multiplexes focus on the few most profitable titles, while those institutions that could serve midlist titles (to borrow a term from the publishing industry) have all but vanished. The lack of cinemas is a problem in itself, but it also generates another one: the lack of diversity in titles.

## **The structure of p2p file-sharing markets**

The structure of illegal online content (films, music, TV programs, e-books, software, etc) markets is a complex one, where p2p users sharing with and downloading from each other represent only the last step in an intricate and mostly hidden pyramid of middlemen, who participate in the process of acquiring, digitizing and distributing cultural items intended for official release. Based on Howe (January 2005) before a release hits the file-sharing networks, there are several groups whose participation is needed to make something widely accessible.

There are the insiders, „[i]ndustry and theater employees [who] run their own straight-to-video operations. Hackers looking for prerelease videogames target

company servers. And before that long-awaited CD hits Amazon.com, moles inside disc-stamping plants have already got a copy.”(Howe, January 2005) Then, release groups digitally repackage multi-gigabyte movie files for easy online distribution, rip CDs into mp3, or create cracks that bypass DRM. Many release groups have exclusive relationships with sites on the top of the distribution hierarchy. When a file appears on a so-called top-site, the distribution chain-reaction begins. Couriers step in to copy and transfer files from the top-sites to lower-level dump sites, and then from there to P2P networks and hubs. The couriers are working for such rewards as fame and respect, or for props from their peers and credits redeemable for goods on upper levels of the pyramid. (b-bstf, Summer 2004) The p2p using public mostly trades what is made available for them through these distribution channels. However, local (in many instances semi-amateur) release groups and individuals also participate in the digitization and publication processes, releasing mostly locally relevant titles to local hubs.

This structure of the underground cultural markets suggests two different factors that shape the supply of pirated online goods. On the one hand there is a steady stream of the global supply of the most recent titles. Weeks or months before the official release dates music, film, software is made available through the shadow distribution pyramid. At the same time, local releasers and individuals continuously release titles the local community deems important. Such local releases are either user-localized releases of global titles, such as vernacular fan-subbed releases, or different versions (DVD rips, TV rips, etc) of officially localized titles already on the market. The balance between the global and the local releases for a given hub is defined by the demographics and by the interests of the community that gathers around a specific hub.

File-sharing is undoubtedly a mass phenomenon and this massive demand for such services has called to life a wide variety of file-sharing protocols and each of these protocols support a number of different business models. Bittorrent for example is a file-sharing protocol, the method which describes how peers can connect and communicate with each other. Likewise, Gnutella, Freenet, etc. are similar protocols with different technical characteristics. Apart from a small number of these protocols, which are proprietary, software developers are able to develop a variety of client software to connect to the network.

Open protocols also enable the proliferation of services that coordinate the users using any given protocol. Sometimes called hubs, at other networks trackers, these services serve as meeting points for users who wish to share and download, or engage with each other in any other fashion. It is up to the service providers to decide what kind of business model they want to adopt. Some hubs operate in a truly communitarian fashion: the system administrators finance the operating costs of an open, ad-free service, as they believe they are engaged in a cultural / political mission. Others follow a closed, ad-supported model. The notorious Pirate Bay is a global, open torrent tracker, which exposing its users to advertising content, it is however unclear whether ad revenues cover the operating costs of the service. On the other end of the scale we find open, for-profit distributors: illegal warez servers which sell a flat rate access for a

relatively high monthly fee as well as authorized distributors who operate their services with the approval of rights-holders.

Services/clients	Open service (anyone can become a member)	Closed (invitation only) service
Non-profit (with no ad revenue / without membership fee / donations based)	Elite DC Hub service (DC protocol) Soulseek client (soulseek protocol)	Karagarga tracker (bittorrent protocol)
Ad supported	Piratebay tracker (bittorrent protocol) Mininova tracker (bittorrent protocol)	Bithumen tracker (bittorrent protocol) Ncore tracker (bittorrent protocol)
For profit (adware / spyware / membership fee)	Kazaa client (fasttrack protocol) Bearshare client (Gnutella protocol)	Stealth warez ftp servers

It should be clear though that - apart from the warez FTP servers, which not being a p2p service should not be included in this table anyway -, all of these services are free for the users in the sense that they cannot request and they don't receive any compensation neither from the tracker service nor from their fellow users for uploading or downloading content through these services. There are however several third party services which build on these networks, offering, for example server space for seeding files, enabling individual users to achieve a higher upload ratio thus more downloads for them. We do not take such services into account in this report, as their use is no way necessary to participate in file-sharing.

Apart from their business models we can differentiate file-sharing services according to their membership policies. An open service means that there are no registration requirements, or that gaining membership is easy. Closed services accept a limited number of users only, usually through invitation by an existing member. Openness has obvious advantages: the more connected people there are, the wider the catalogue and the faster the downloads are. On the other hand openness raises several issues: that of free-riding and that of the risk of being caught and litigated if rights-holders think the services breach - or help users breach - copyright.

Both of these factors have played a role in the proliferation of closed / secretive services in the last few years. Invitation only services offer a good solution to the problem of anonymous users free-riding on others and/or polluting the catalog with garbage. At the same time it offers some level of protection against rights-holders who hope to solve what they perceive as a threat to their business by taking legal actions against individual users as well as service providers. Exclusivity has always served a third, non-related function in the file-sharing

scene as well: the social hierarchy of the scene is created, maintained and measured by having access to certain sites: members of the most exclusive services are the highest ranking in the unofficial sub-cultural hierarchy.

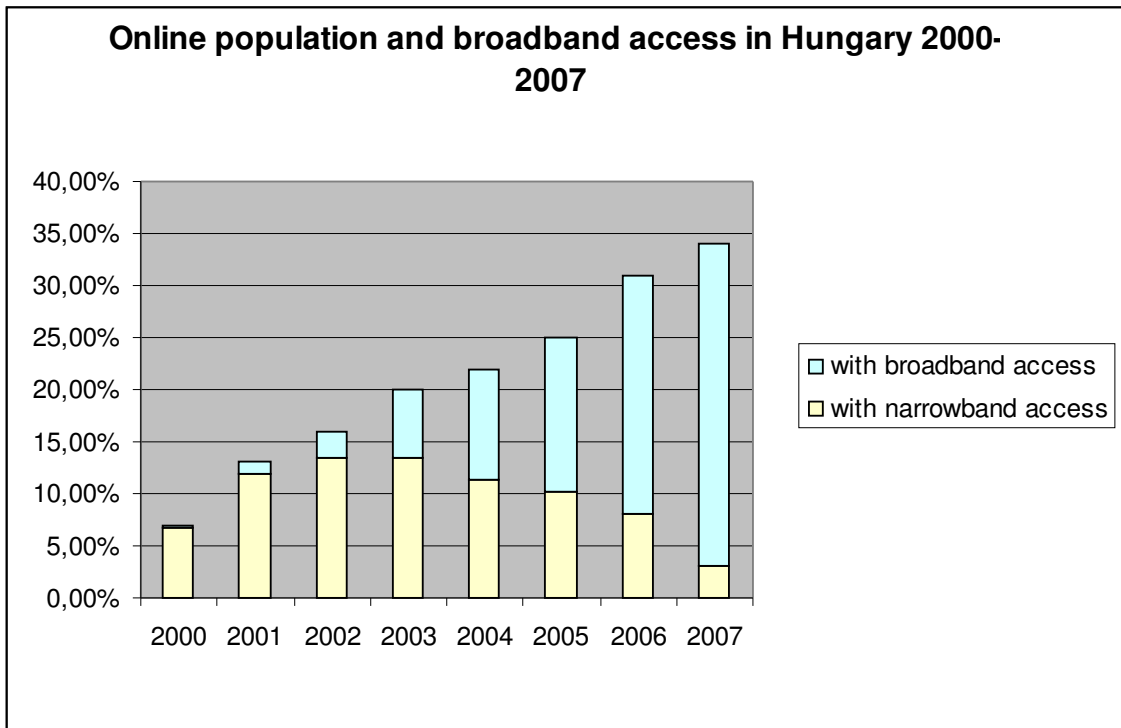
Apart from these third type of sites that wish to maintain their exclusivity at all costs, all other torrent tracker services need to balance their interests between having a relatively large userbase, and limiting the access to their services, therefore they all allow new users to join their services from time to time.

## The Hungarian file-sharing scene

Although sharing computer files is as old as computers themselves, the first mainstream file sharing applications emerged at the turn of the new millennium.

<b>First released</b>	<b>P2P Protocol</b>
July 1999	Freenet
September 1999	Napster
November 1999	Direct Connect
March 2000	Gnutella
September 2000	eDonkey2000
April 2001	BitTorrent

By 1999 all the necessary preconditions for wide-scale file-sharing were already in place: reasonable size individual digital libraries, an increasing level of PC penetration and bearable download times even with a modem. Nevertheless, it was at the colleges, equipped with broadband connections, where file-sharing first took off.



Source: National Communications Authority

College networks also hosted the few first Hungarian file-sharing hubs. The first Hungarian hub using the DC protocol started in 2001 and by 2002 as many as 6 hubs were running. All of these were on university networks, and only started to move to dedicated servers at commercial hosting services around 2004, when dealing with university network administrators became more difficult. ("Az Elite Hub történelme,") In the meantime residential broadband access started to gain momentum.

The first Hungarian bittorrent tracker, bitHUMen started in July, 2004 with a few hundred users. **(sct, 2009 február 01. 15:20)** Soon others followed. With the rapid growth of residential broadband access the number of trackers and the number of users increased rapidly. At the end of 2008 the top 10 Hungarian torrent trackers had the following registered user-base and peer numbers (blue represents trackers participating in our study):

Tracker	registered users	peers (downloaders + uploaders) on 2008. december 28.
nCore	78612	308330
Independence	68315	90327
Moobs	59989	125913
bitHUMen	51318	218731
Malacka	45067	24054
PREtorians	39979	39250
1st Torrent	37692	46811

GigaTorrents	35001	95085
Spiryt	33349	17898
Blue-Dragon	31826	66153

Source: <http://asva.info/2008-magyar-bittorrent-trackerei-2008-12-29.html>

This of course does not provide us with any real estimate on how many file-sharers there actually are in Hungary. Simply totaling the number of users for each tracker would result in taking into account those users more than once who are registered at multiple trackers. Others, who are not members of these sites would go unaccounted for. According to one estimate there were around 300.000 file sharers in Hungary in early 2008. (Turcsán, 2008. február 7. ) Based data from countries with similar population sizes (Huygen et al., 18 February 2009) we need to conclude that this figure of 300.000 downloaders must be a very conservative estimate.

## Trackers we tracked

We have decided to track 3 of the most popular Hungarian bittorrent trackers: bitHUMen, nCore and Independence. The choices were made based on reputation, stability, number of users/peers, number and type of titles and finally access. Information on these dimensions were based on interviews of and recommendations by community members.

BitHUMen service is the oldest Hungarian tracker with a solid reputation and a committed community. BitHUMen is also the world's 26<sup>th</sup> most sought after tracker in terms of the number of invitation requests found on the Internet (sharky, December 17, 08). It is relatively ad-free suggesting a non-profit operation.

NCore, another closed Hungarian hub we have tracked, ranks #40 in the same list. The specialty of this tracker is that it accepts releases from outside of "the scene", the unofficial circle of trusted release groups. Anyone can release on nCore, which means a wider variety of titles, but sometimes also lower quality and lower download speeds. nCore is also an ad based service.

Independence is a relative newcomer compared to the other two. Any user can register to the site, however registration is not free. Independence therefore reaches out to those users who cannot get into the other, more reclusive trackers and offers them a chance to buy themselves into a world they cannot otherwise have access to. This, and the site's strong emphasis on monetizing its user-base seemed to create a bad reputation for the site and its owner among Hungarian file-sharers, who seem to deem such an unabashedly commercial approach objectionable. Nevertheless exactly because of its relative openness we included it in our study.

Even if there is apparent (social) value in exclusiveness, all the torrent tracker services need to balance their interests between exclusivity and the advantages of a wide user-base, therefore they all let new users join the service from time to time. It is possible to join these services, even if it takes some time and effort. In the case of bitHUMen and nCore, we decided to track these closed /invitation only services, and we interpreted their entry barrier as a variable that separates casual file-sharers from those who engage in file-sharing in a more systematic fashion.

## Methodology of torrent traffic tracking

There are several approaches to measuring peer-to-peer file-sharing traffic (Chu, Labonte, & Levine, 2002; Gummadi et al., ; Guo et al., ; Pouwelse, Garbacki, Epema, & Sips, 2005; Saroiu, Gummadi, & Gribble, 2002; Schulze & Mochalski, 2008; Sen & Wang, 2002) utilizing deep packet inspection techniques, protocol level sampling or other approaches. Our decision to develop a new method was the result of a simple factor: as we do not have access to any data source that would let us connect IP addresses to settlement level geographical data. As geographic analysis is crucial in our research, we needed to come up with a different approach that enabled us to acquire user location data.

Luckily the most popular/influential torrent tracker services have all enabled their users to communicate on their user profile pages the settlement where they live. Not everyone has filled out this field in his/her profile, and there are significant differences in the list of settlements these services offered their users to choose from. Nevertheless in 40% of the cases we had proper location information which gave us enough torrent traffic data for the research to be feasible.

Our approach therefore focuses on the hubs that serve the Hungarian file-sharing community instead of monitoring the actual data flows over the network.

Besides serving as community hubs, providing users with self-identification and communication (forums, ratings, polls, etc) services, torrent trackers coordinate the p2p flock. They maintain the information about which user has which part of which file in the network. The users need to communicate with the tracker if they want to download or share something from the others, as it is the server which knows which user has the necessary piece of the file in question. Therefore the server knows, and publishes this type of information which is then relatively easy to gather.

Open-access hubs are easy to monitor, as they do not make an attempt to hide their activities. Closed hubs require more precaution, so the monitoring activity does not get detected by the administrators of the site. Such a monitoring effort raises several ethical issues. We have addressed these issues on several levels. First we gather only such data that is available for each and every ordinary member of the torrent tracker. We respected the privacy decisions of the site

administrators, and did not try to gather more information that was intended by them to be public. Also, we did everything we could to respect and protect the privacy of the individuals who use these services. We did not collect any information that could be used either by us or by other parties to connect the online user profiles with real life identities. On the other hand we did engage in a monitoring effort without the knowledge and consent on either the site administrators or the users. This was necessary as acquiring the same amount and depth of information from the administrators of these sites would have been impossible: either because they don't archive such information, or because they do everything they can to protect their communities. Before, during and after the data gathering period we have communicated clearly to the Hungarian file sharing scene that we are doing research on the effect of file-sharing on traditional markets. We were also trying to be present in the online discussion boards so we could personally answer any questions about the research.

In order to achieve a non-intrusive, difficult-to-detect monitoring of closed hubs we have developed the appropriate monitoring technology. The software has three main functionalities. (1) Its crawler collects data from the torrent hub, (2) the parser extracts relevant information, stores in a (3) database and instructs the crawler on which page to crawl next.

### **The crawler**

The crawler collects the traces of the data flow generated by the torrent communities. As the torrent-tracker is the information hub to which all users go for new downloads, it contains relevant information on what is accessible at any given moment through the given hub. The tracker also contains information on the health and status of the torrent files, it contains the profiles of the users, offering us a chance to extract some information on the users themselves. The crawler crawls these pages to extract every possible information available on the hub through the web interface.

The task of ensuring data quality requires that we crawl the hub (or parts of the hub) frequently. The status of freshly released popular titles (like a Hollywood blockbuster) changes quickly: dozens of users can appear and disappear from the downloading flock within minutes. This requires an intensive presence of the crawler on a site that tries to do everything it can to protect its users from scrutiny and possible prosecution. By dispersing the queries among a number of different proxies we were able to conduct an aggressive but non-intrusive, non-detectable monitoring.

### **The parser**

The parser has three tasks. First it extracts the relevant data from the files sent by the crawler. Second, it anonymizes user related data and dumps all the data into the database. It also adjusts the frequency by which the crawler needs to request a specific page. To avoid data loss we crawl popular, thus quickly



changing pages more often than pages of less popular or dead torrents. The parser adjusts the time of the next crawl of a page based on the amount of changes in the flock around the torrent since the last crawl.

## Data cleaning, title identification

The torrent files form the basis of the analysis, representing movie titles. This data needs further cleaning as there are several issues to be solved. Even in the case of global releases of global titles there might be several competing versions of the same title in various formats (VCD, DVDrip, CAM), released at different times, by different release groups.

For example the fourth installment in the Die Hard movie franchise is available under the following names: [07.11.17.Live.Free.Or.Die.Hard.Blu.Ray.All.Disk@Ht](#), *Die Hard - Quadrilogy. Untouched Box Set.Nordic*, *Die.Hard.SE.Trilogy.BOXSET.PAL.6DISC.DVDR-SPLiNTER*, *Die Hard Series*, *Die.Hard.4.0.Yippee.Ki.Yay.Edition.2DiSC.NORDiC.PAL.DVDR-ViSiON*, *DIE\_HARD\_4\_0. PAL. R2. SUBS DK,NO,SE.FI. DVD9*, *Live Free Or Die Hard (La Jungla De Cristal 4)*, *Die.Hard.4 x264.720p*, *Die Hard 4. 720p BluRay AC3-5.1 x264*, *Die Hard 4*, *Die Hard 4.0*, *Die Hard 4.0 (Live Free or Die Hard) 2007 DUTCH!*, *Die.Hard.4[2007]MultiSub.DvDR-Gothicmaster*.

It is clear that all of these versions contain the Die Hard 4 movie. The consolidation of these versions and establishing the connection with the titles stored in other datasets was a difficult task. We needed to consolidate more than 7000 movie torrent files that were uploaded within the examined timeframe with nearly 12.000 movie titles that track traditional markets. Proxies such as IMDB ID proved to be massively unreliable, and automatic pairing algorithms provided noisy results. Therefore we decided to crowdsource the task of pairing and asked the file-sharers themselves to participate in connecting torrent files with titles. The results were beyond every expectation: several hundred anonymous users finished this task in less than a week, with very few (less than 1%) errors.

Similar, but much smaller task was to standardize user location info.

## Additional Data sources

Our focus being the relationship between peer-to-peer trading and movie distribution of films, we had to conduct our data mining and analysis in a way that accounts for the multiple connections that exist between these two domains. As explained above, torrent files were identified making it possible to analyze not simple torrent but content-related trading patterns. Without meticulously linking torrent files to specific film titles, we could not have undertaken a content-oriented analysis of the peer-to-peer networks. The content oriented analysis was further supported by acquiring and cross referencing the following data sources:

Data from the p2p network <sup>2</sup>	Box office statistics <sup>3</sup>
--	------------------------------------

<sup>2</sup> (1) A database of the connections between users and torrent files. Each connection accounts for a different case, whereby users downloading several different torrent files appear in separate lines in the form of individual “transactions”. Both the users and the torrent files are identified. This is a comprehensive database as it includes

<ul style="list-style-type: none"> <li>- Location of downloader</li> <li>- Title of movie</li> <li>- Time of the download</li> <li>- Length of the download/upload activity</li> </ul>	<ul style="list-style-type: none"> <li>- Title of the movie</li> <li>- Date of release</li> <li>- No. of ticket sold in the year of release</li> <li>- Box office revenue (in HUF) in the year of release</li> <li>- No. of copies</li> </ul>
Cinema program guide <sup>4</sup>	IMDB (where available)
<ul style="list-style-type: none"> <li>- Location of the cinema</li> <li>- Title of the movie</li> <li>- No. and date of screenings</li> </ul>	<ul style="list-style-type: none"> <li>- Title of the movie</li> <li>- User rating score</li> <li>- No of rating votes</li> <li>- Thematic categories</li> </ul>

Using these sources, we were able to analyze in detail the relationship between a film's peer-to-peer and cinema distribution within a two months timeframe between May 1<sup>st</sup> and June 30<sup>th</sup> 2008.

## Results

### Basic user statistics

Various estimates put the overall number of Hungarian file-sharers between 3-600.000 users. During our overall monitoring period (between March and December, 2008) we have encountered 187.000 users on three trackers. The number of individuals behind these online avatars is probably smaller due to overlapping user-bases. In the time-window we encountered 63.000 users.

We were able to identify the location of the users in 24.000 (37%) cases. If we compare this data to the distribution of Hungarian citizens and Hungarian broadband subscribers among different settlement types, we need to conclude that the p2p users we were able to identify are tend to concentrate in the capital and in the biggest urban centers and we can find them underrepresented in small towns and villages. The fact that the village residents are underrepresented in our study can be explained to a certain degree by the fact that the sites in our study offered only limited lists of settlements for their users to choose from, and these hardly include the smaller settlements. Nevertheless, this skewedness was

---

all the transactions that occurred on the three selected Hungarian trackers between May 1st and June 30th 2008.

(2) A database with information on torrents: size, upload and creation date, tracker, etc. using unique identification. (3) A table linking torrent files to film titles. (Films, just like torrent files, have a unique identification code.) (4) A user database with all available information on the users registered at the three torrent trackers in our data collection. User location is of particular importance as it allows for examining the geographical aspect (availability of films downloaded within users' catchment area). (5) A geographical database with elementary information on settlements, including distances between pairs of them. This is especially useful when exploring the relative shortage of films within one's catchment area.

<sup>3</sup> (1) Detailed box office data (release dates, number of tickets sold, revenue, copies, etc.) for Hungarian releases between February 2004 and December 2008.

<sup>4</sup> (1) Screening dates, times, cinema locations for each film screened in Hungary after the year 2000. (2) A cinema database with the geographical parameters of the cinemas including the name of its location, seats, etc.

observable even if the tracker (nCore) had an extensive selection of settlements to offer.

Settlement type	Hungary (in thousands)		Broadband subscribers in June 2007 (in thousands)		P2P users with known location	
	Sum		Sum		Sum	
1 Budapest	1 778	17%	334	30%	7259	30%
2 County capital	1 821	18%	234	21%	8055	34%
3 Town	3 395	33%	343	31%	7044	30%
4 Village	3 188	31%	194	18%	1487	6%

Sources: National Bureau of Statistics, National Communications Authority

This user-distribution hints that the diffusion of (invitation-based) file-sharing is more successful in urban centers with rich and complex social interactions. As membership on these services can be acquired through the invitation of an active member, without the proper social network it is difficult to get into these services. Living in a relatively poor media environment with limited access to different forms of entertainment might be a motivational force to seek out alternative access channels on the internet, but even if such a drive exists, we could not confirm its existence in this study. The file-sharing population in our study lives in places with relatively rich and varied access to different media and other forms of entertainment.

The relative richness of the media environment is also stressed by the fact that only 21% of the users live in a settlement where there was not a single cinema screening in our time-window. This figure rises to 24% if we only include cinemas with more than one screenings per week.

### Basic movie statistics

We identified 4838 films in our study, which we sorted into five categories according to their download and movie availability. As for the latter we defined three categories: one includes films that were screened when we tracked the downloads. Another category includes those films of which we know the cinematic distribution data. If we could not find such data for a film, we regard it as 'Not screened' even if it was screened but only a long time ago (i.e. prior to 2004).

The distribution of films among the categories is shown in the following table.

	Downloaded	Not downloaded
Screened in the download window	152 (3,1%)	592 (12,2%)
Screened sometime before the download window	776 (16%)	627 (13%)
Not screened	2691 (55,6%)	N/A

Sum:	3619 (74,8%)	1219 (25,2%)
------	-----------------	-----------------

We found that nearly 75% of all the *downloaded* films are in the 'Not screened' category. Such a high proportion could be an indicator of the importance of peer-to-peer networks in the diffusion of cinematographic content, as it suggests that one main motivation behind downloading is *scarcity* on the legitimate market. But before accepting this conclusion we should note, that we lack conclusive data about the DVD (sales and rental), television (broadcast, cable, satellite and IPTV) and legitimate online distribution channels therefore we cannot readily accept (or reject) the scarcity model on the content side.

However it is clear that release windows strictly define when, in what format is the content available on the market and for how long. The traditional system of audiovisual content distribution therefore places strict rules on the accessibility of content. Accessibility can be limited temporally: DVD distribution rights expire, broadcast dates pass. It can be limited geographically: the closest retail outlet might be inconveniently far away. P2P downloading can bridge both types of limitations.

The basic statistics (see next page), however, do not support the temporal scarcity suggestion. The mean life-span of torrents is significantly lower than the mean life-span of movies in the cinema in every category. This is clearly the result of the file-sharing technology we have observed.

Unlike the DC++ file-sharing hubs, that usually prescribe a minimum amount of data to be offered in a shared library, bittorrent trackers require that the user balances his/her upload/download ratio around 1.0. This technical setup has some serious implications on how content is distributed and consumed on each network. Users around DC++ hubs form large, searchable archives, where the amount of data shared is in itself a source of pride and social recognition. Bittorrent, on the other hand discourages the emergence of large individual shared libraries as such large libraries offer little reward in terms of the valuable upload ratio. As most of the downloads of a new torrent file concentrate to the first few days of its lifespan, those who wish to gain some upload credit, need to be able to offer, therefore download titles that are or will be downloaded by others. As the reward comes from serving a title to the highest number of users in as short time as possible, the lifespan of the titles on bittorrent networks are short, with the overwhelming majority of the downloads occurring in the first few days of the torrent's existence. These findings apply to the observed closed, Hungarian torrent trackers. Open, global trackers, like The Pirate Bay have larger user-bases therefore the chances to find older content are higher. We can also expect that despite the language difficulties, at least some of the local users use these latter, global services, and get content that is not available on local trackers from those services. We, however are unable to monitor such activity, therefore we can only suspect the existence of such usage pattern.

Geographical scarcity, however, might still apply: of all the distribution channels (apart from terrestrial broadcasting) internet endpoints are the densest in the geographical sense. The data to confirm the geographical scarcity is readily available, and some preliminary tests confirmed the significance of this dimension, however, we currently lack the detailed analysis of this variable.

## Detailed analysis

The detailed statistics of the 5 categories are shown in the following table:

**Macro-statistics of theatrical distribution and peer-to-peer traffic  
(base=all transactions)**

Film categories		<b>1 Within time frame: screened AND downloaded</b>	<b>2 Within time frame: screened AND NOT downloaded</b>	<b>3 Prior to time frame: screened AND downloaded</b>	<b>4 Prior to time frame: screened AND NOT downloaded</b>	<b>5 DID NOT screen AND downloaded</b>	Total
<b>Torrent life-span (days)</b>	Mean	<b>79</b>	0	40	0	24	22
	Maximum	<b>617</b>	0	439	0	440	617
	Sum	<b>11970</b>	0	30998	0	<b>65771</b>	108739
	Std Deviation	<b>99</b>	0	41	0	33	39
<b>Number of downloads</b>	Mean	<b>1042</b>	0	190	0	129	135
	Maximum	<b>9108</b>	0	2579	0	6736	9108
	Sum	<b>158358</b>	0	147357	0	<b>346844</b>	652559
	Std Deviation	<b>1874</b>	0	247	0	335	461
<b>Film life-span (days)</b>	Mean	110	<b>126</b>	103	102	.	110
	Maximum	<b>418</b>	417	409	410	.	418
	Sum	16757	<b>74525</b>	<b>79959</b>	63955	.	235196
	Std Deviation	<b>137</b>	136	108	107	.	118
<b>Screenings</b>	Mean	<b>2636</b>	434	1785	1487	0	615
	Maximum	<b>12588</b>	11789	<b>14008</b>	11284	0	14008
	Sum	<b>400623</b>	256689	<b>1384916</b>	932073	0	2974301
	Std Deviation	<b>2735</b>	1161	2356	2196	0	1604
<b>No. of theaters where film screened</b>	Mean	<b>63</b>	19	<b>69</b>	60	0	23
	Maximum	249	<b>286</b>	<b>301</b>	<b>301</b>	0	301
	Sum	9635	<b>11087</b>	<b>53533</b>	37326	0	111581
	Std Deviation	49	35	<b>72</b>	69	0	50
<b>Revenue (million HUF)</b>	Mean	<b>93</b>	20	68	58	0	10
	Maximum	<b>676</b>	405	<b>686</b>	556	0	686
	Sum	<b>8270</b>	2340	<b>14613</b>	8367	0	33590
	Std Deviation	<b>126</b>	47	96	91	0	46

<b>Tickets sold</b>	Mean	<b>99064</b>	27107	78673	68703	0	11928
	Maximum	<b>853926</b>	501098	826129	610135	0	853926
	Sum	<b>8816665</b>	3144467	<b>16914660</b>	9961981	0	3883773
	Std Deviation	<b>149989</b>	58240	116819	106648	0	53849
<b>Tickets per screening</b>	Mean	<b>24</b>	17	23	21	.	21
	Maximum	<b>121</b>	55	101	70	.	121
	Sum	<b>2136</b>	1921	<b>4895</b>	3095	.	12046
	Std Deviation	<b>17</b>	8	13	11	.	13
<b>Copies</b>	Mean	<b>18</b>	7	17	16	0	2
	Maximum	<b>41</b>	<b>41</b>	<b>43</b>	<b>43</b>	0	43
	Sum	<b>1690</b>	828	<b>2917</b>	1773	0	7208
	Std Deviation	<b>11</b>	8	9	10	0	7
<b>Time between cinematic and p2p releases (weeks)</b>	Mean	104	122	239	<b>244</b>	.	199
	Maximum	<b>410</b>	<b>410</b>	<b>410</b>	<b>410</b>	.	410
	Sum	15882	72239	<b>185246</b>	153205	.	426572
	Std Deviation	<b>138</b>	135	129	124	.	142
<b>Time between last screening and time frame (weeks)</b>	Mean	0	0	136	<b>143</b>	.	91
	Maximum	0	0	408	<b>409</b>	.	409
	Sum	0	0	<b>105420</b>	89353	.	194773
	Std Deviation	0	0	<b>120</b>	117	.	116

From these data we can formulate some hypotheses.

### **H1: The more recent a film is, the more it is downloaded.**

Even though Cat. 5 has the most overall downloads, it is only due to the high number of films in this category. The average number of downloads is the highest (1042) in those cases when the film is available in the cinemas as well. Relatively recently screened films (in Cat. 3) have significantly lower appeal (190), while films without cinematic support have the fewest downloads in average. What is true for the demand side is also true for the supply side: we see the same ranking in regard to the life-span of the torrents. Users seem to seed recent films longer.

### **H2: Marketing power matters**

If we look at the market data of cinematic distribution, we can observe significant deviations in Cat. 2. It seems that three factors, the number of copies the film starts with, the number of theatres it is shown and the number of screenings are significantly lower in Cat. 2. These factors are controlled by the distributor and

they strongly correlate with the size of the marketing budget of a movie. A saturation release in the Hungarian movie market means opening a film with 30-40 copies, and screening it in around 250-300 cinemas 3-400.000 times. It seems that such films tend to show up on the p2p networks as well, populating Cat. 1, while films with narrow releases, fewer copies, more limited geographical reach populate Cat 2.

Data from Cat 3&4 suggests that the effect of marketing power diminishes over time as less marketed films eventually show up on the p2p networks.

The correlation table (seen in the next page) helps us formulate a third hypothesis:

**H3: There is a connection between the popularity of a movie in the cinemas (in terms of the number of ticket sold) and the popularity on the p2p networks.**

The correlation table for the entire population shows a statistically significant, weak positive correlation between box office popularity and the number of downloads, which might be a result of the aforementioned marketing effect. Besides, in other sub-sections of the data, we should look for negative correlation, to check if the p2p and the cinema markets supplement each other.

The correlation table for the whole population has few surprises. The number of downloads is positively correlated with the number of screenings, theatres, copies and tickets sold. These correlations are, however, weak.

The relatively strong, positive correlation between the time since a movie was last screened and the number of theatres, and the copies shows the contraction of the Hungarian cinema network: in the past movies started with more copies and were shown in more theatres.



### Correlations of theatrical distribution and peer-to-peer traffic (base=Cat. 1,2,3,4)

		Torrent life-span (days)	Number of downloads	Film life-span (days)	Screenings	No. of theaters where film screened	Revenue (million HUF)	Tickets sold	Tickets per screening	Copies	Time between cinematic and p2p releases (weeks)	Time between last screening and time frame (weeks)
Torrent life-span (days)	Pearson Corr. Sig. (2-tailed) N	<b>1</b> 4838										
Number of downloads	Pearson Corr. Sig. (2-tailed) N	<b>0,654</b> 0,000 4838	<b>1</b> 4838									
Film life-span (days)	Pearson Corr. Sig. (2-tailed) N	<b>-0,073</b> 0,001 2147	-0,093 0,000 2147	<b>1</b> 2147								
Screenings	Pearson Corr. Sig. (2-tailed) N	<b>0,126</b> 0,000 4838	<b>0,116</b> 0,000 4838	<b>-0,043</b> 0,046 2147	<b>1</b> 4838							
No. of theaters where film screened	Pearson Corr. Sig. (2-tailed) N	<b>0,073</b> 0,000 4838	<b>0,054</b> 0,000 4838	<b>0,000</b> 0,994 2147	<b>0,883</b> 0,000 4838	<b>1</b> 4838						
Revenue (million HUF)	Pearson Corr. Sig. (2-tailed) N	<b>0,097</b> 0,000 3256	0,075 0,000 3256	0,104 0,014 565	<b>0,880</b> 0,000 3256	0,736 0,000 3256	1 3256					
Tickets sold	Pearson Corr. Sig. (2-tailed) N	<b>0,083</b> 0,000 3256	<b>0,056</b> 0,002 3256	0,147 0,000 565	0,867 0,000 3256	0,733 0,000 3256	<b>0,991</b> 0,000 3256	1 3256				
Tickets per screening	Pearson Corr. Sig. (2-tailed) N	<b>0,068</b> 0,108 565	0,002 0,967 565	0,206 0,000 565	0,599 0,000 565	0,562 0,000 565	<b>0,804</b> 0,000 565	<b>0,828</b> 0,000 565	1 565			

Copies	Pearson Corr. Sig. (2- tailed) N	<b>0,131</b> 0,000 3178	<b>0,156</b> 0,000 3178	-0,123 0,007 487	<b>0,931</b> 0,000 3178	<b>0,942</b> 0,000 3178	<b>0,757</b> 0,000 3144	<b>0,737</b> 0,000 3144	<b>0,512</b> 0,000 453	1 3178		
Time between cinematic and p2p releases (weeks)	Pearson Corr. Sig. (2- tailed) N	<b>-0,044</b> 0,040 2147	<b>-0,131</b> 0,000 2147	<b>0,606</b> 0,000 2147	<b>-0,008</b> 0,697 2147	<b>0,171</b> 0,000 2147	<b>0,086</b> 0,041 565	<b>0,122</b> 0,004 565	<b>0,174</b> 0,000 565	0,040 0,375 487	1 2147	
Time between last screening and time frame (weeks)	Pearson Corr. Sig. (2- tailed) N	<b>0,022</b> 0,319 2147	<b>-0,063</b> 0,004 2147	<b>-0,273</b> 0,000 2147	<b>0,034</b> 0,118 2147	<b>0,207</b> 0,000 2147	<b>0,003</b> 0,949 565	<b>0,004</b> 0,923 565	<b>0,007</b> 0,873 565	<b>0,198</b> 0,000 487	0,599 0,000 2147	1 2147

\*\*

Corr. is significant at the 0.01 level (2-tailed).

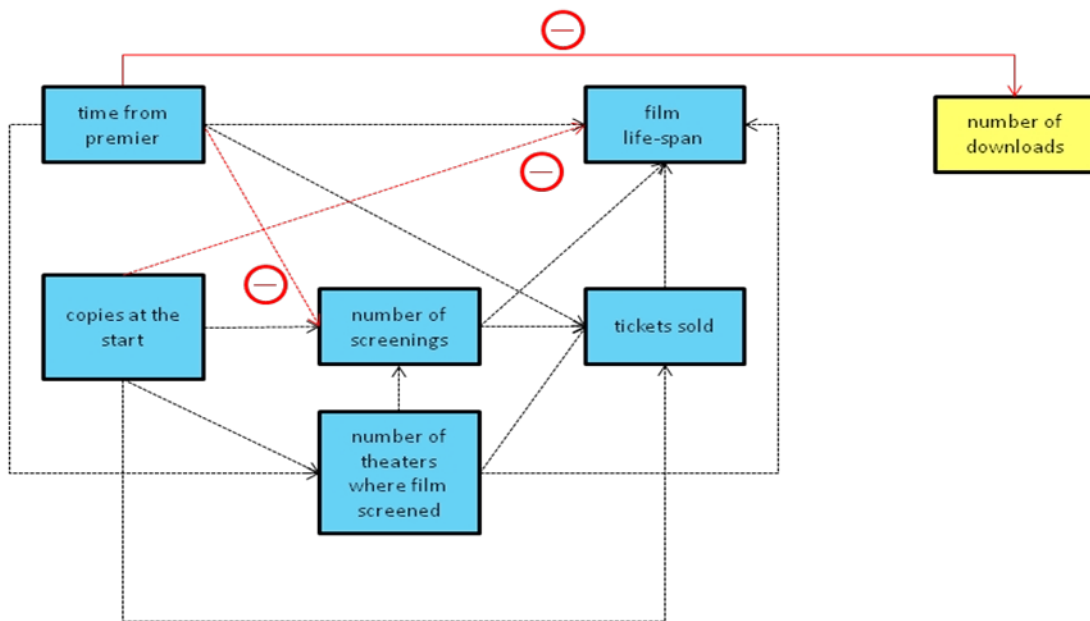
\*

Corr. is significant at the 0.05 level (2-tailed).

We have built a regression model to directly measure the effect of the aforementioned variables to the number of downloads. We have defined a path model that tries to explain the number of downloads with the distributor-controlled factors, such as the number of copies and the time between the movies' official and p2p release date.

**Path model for downloads (between April and June 2008)**

*variant: film categories=1, 2, 3, 4, 5 (n=4838)  
R<sup>2</sup>=0,034*



**Regression coefficients for the explanatory model of a film's popularity among downloaders (base=Cat. 1,2,3,4,5) R<sup>2</sup>=0,034**

Model	Unstandardized Coefficients		Standardized Coefficients (Beta)	t	Sig.	95% Confidence Interval for B	
	B	Std. Error				Lower Bound	Upper Bound
1 (Constant)	248,17	75,16		3,30	0,00	100,47	395,87
t_from_premier Time between cinematic and p2p releases (weeks)	-0,33	0,16	-0,10	-2,08	0,04	-0,64	-0,02
rkopia Copies (residual)	36,85	20,36	0,12	1,81	0,07	-3,16	76,85

rsum_cin	No. of theaters where film screened (residual)	-53,1 2	30,79	-0,1 0	-1,7 2	0,0 9	-113,63	7,40
rsum_scr	Screenings (residual)	4,71	16,61	0,0 2	0,2 8	0,7 8	-27,93	37,36
rnezoszam	Tickets sold (residual)	-16,0 0	11,56	-0,0 9	-1,3 8	0,1 7	-38,72	6,72
rf_life_span	Film life-span (residual)	-4,79	65,23	0,0 0	-0,0 7	0,9 4	-132,98	123,41

a Dependent Variable: d\_torrent\_max  
Number of downloads

In the regression model the weak correlations disappear, the residual effects of the independent variables on the number of downloads as the dependent variable are statistically insignificant apart from a weak negative effect of the time that elapsed between the cinematic and P2P releases. Recent films have somewhat higher downloads and as time passes, demand on the p2p networks fades with the memories of the users. This supports our H1 hypothesis.

Interestingly, we could not find any causal relationship between movie popularity in cinemas and movie popularity on p2p networks, therefore in this population we need to reject H3.

We should remember, however, that for the films in Cat 3, 4 and 5 the DVD market availability probably plays a significant role in the fate of p2p downloads, therefore it makes little sense trying to explain the number of downloads in these categories with cinematic distribution data only.

As a result, we limit our further analysis to the first two categories, i.e. to those films that were screened in the p2p observation timeframe.

As a first step we confirmed that the number of copies, i.e marketing power is the relevant factor in determining whether the film will be available on p2p networks parallel with the cinematic distribution. We did this by running a regression in which the dependent variable was a binary variable of the film being downloadable on the P2P networks. This regression confirmed that films with copies below a certain threshold such as niche films, art-house movies, films with limited marketing budgets, locally produced films are slow to get to the p2p networks. The limited marketing efforts fail to generate enough interest from the Hungarian release-scene to re-release a foreign torrent of the film, or to produce an original local release from cinema sources. In these cases releasers just wait for the DVD to come out and rip that long after the cinema-life of the film is over.

With this step we narrowed our study to those films which screened and were downloaded in the same time.

### Correlations of theatrical distribution and peer-to-peer traffic (base=Cat. 1)

		Torrent life-span (days)	Number of downloads	Film life-span (days)	Screenings	No. of theaters where film screened	Revenue (million HUF)	Tickets sold	Tickets per screening	Copies	Time between cinematic and p2p releases (weeks)	Time between last screening and time frame (weeks)
Torrent life-span (days)	Pearson Corr. Sig. (2-tailed) N	<b>1</b> 0,000 92										
Number of downloads	Pearson Corr. Sig. (2-tailed) N	<b>0,703</b> 0,000 92	<b>1</b> 0,000 92									
Film life-span (days)	Pearson Corr. Sig. (2-tailed) N	<b>-0,253</b> 0,015 92	-0,299 0,004 92	<b>1</b> 0,000 92								
Screenings	Pearson Corr. Sig. (2-tailed) N	<b>0,060</b> 0,572 92	<b>-0,078</b> 0,458 92	<b>0,098</b> 0,352 92	<b>1</b> 0,000 92							
No. of theaters where film screened	Pearson Corr. Sig. (2-tailed) N	<b>-0,018</b> 0,867 92	<b>-0,130</b> 0,215 92	<b>0,335</b> 0,001 92	<b>0,827</b> 0,000 92	<b>1</b> 0,000 92						
Revenue (million HUF)	Pearson Corr. Sig. (2-tailed) N	<b>-0,003</b> 0,975 86	-0,127 0,245 86	0,282 0,008 86	<b>0,837</b> 0,000 86	0,693 0,000 86	1 0,000 86					
Tickets sold	Pearson Corr. Sig. (2-tailed) N	<b>-0,018</b> 0,868 86	<b>-0,144</b> 0,186 86	0,320 0,003 86	0,811 0,000 86	0,678 0,000 86	<b>0,976</b> 0,000 86	1 0,000 86				
Tickets per screening	Pearson Corr. Sig. (2-tailed) N	<b>-0,068</b> 0,532 86	-0,130 0,234 86	0,351 0,001 86	0,501 0,000 86	0,491 0,000 86	<b>0,802</b> 0,000 86	<b>0,846</b> 0,000 86	1 0,000 86			

Copies	Pearson Corr. Sig. (2- tailed) N	<b>0,065</b> 0,537 92	<b>-0,006</b> 0,958 92	-0,033 0,752 92	<b>0,809</b> 0,000 92	<b>0,707</b> 0,000 92	<b>0,685</b> 0,000 86	<b>0,646</b> 0,000 86	<b>0,477</b> 0,000 86	1 0,000 92		
Time between cinematic and p2p releases (weeks)	Pearson Corr. Sig. (2- tailed) N	<b>-0,256</b> 0,014 92	<b>-0,295</b> 0,004 92	<b>0,999</b> 0,000 92	<b>0,087</b> 0,409 92	<b>0,328</b> 0,001 92	<b>0,271</b> 0,012 86	<b>0,308</b> 0,004 86	<b>0,352</b> 0,001 86	-0,035 0,737 92	1 0,000 92	
Time between last screening and time frame (weeks)	Pearson Corr. Sig. (2- tailed) N	<b>.(a)</b> . 92	<b>.(a)</b> . 92	<b>.(a)</b> . 92	<b>.(a)</b> . 92	<b>.(a)</b> . 92	<b>.(a)</b> . 86	<b>.(a)</b> . 86	<b>.(a)</b> . 86	<b>.(a)</b> . 92	.(a) . 92	.(a) 0,000 92

\*\*

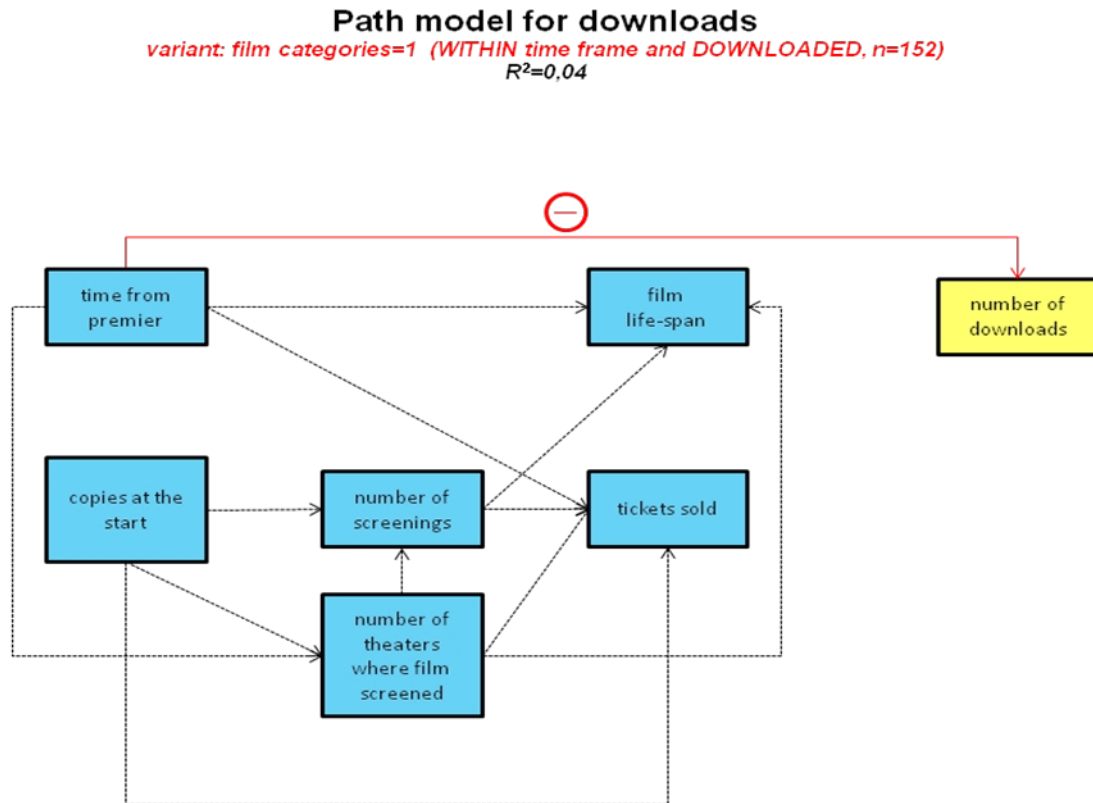
Corr. is significant at the 0.01 level (2-tailed).

\*

Corr. is significant at the 0.05 level (2-tailed).

In the correlation table we find a weak, negative correlation between the number of tickets sold and the number of downloads, which, unlike in the previous case, suggests a substitution effect. On the other hand the fact that the number of screenings and the number of theatres are negatively correlated with the number of downloads suggests that a scarcity effect is in the background.

The regression model, however confirms neither of these suggestions.



**Regression coefficients for the explanatory model of a film's popularity among downloaders (base=Cat. 1)**  
 $R^2=0,041$

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
	B	Std. Error				Lower Bound	Upper Bound
1 (Constant)	1428,77	218,84		6,53	0,00	993,18	1864,36
t_from_premier Time between cinematic and p2p releases (weeks)	-9,48	3,17	-0,33	-2,99	0,00	-15,79	-3,18
rkopia Copies (residual)	-24,57	176,34	-0,02	-0,14	0,89	-375,57	326,43

rsum_cin	No. of theaters where film screened (residual)	-45,65	179,5	-0,0	-0,2	0,8	-402,93	311,62
rsum_scr	Screenings (residual)	-100,0	178,2	-0,0	-0,5	0,5	-454,86	254,78
mezoszam	Tickets sold (residual)	-28,22	157,4	-0,0	-0,1	0,8	-341,53	285,10
rf_life_span	Film life-span (residual)	-225,7	189,6	-0,1	-1,1	0,2	-603,16	151,63
a	Dependent Variable: d_torrent_max Number of downloads	7	0	3	9	4		

Again, only the time difference between the release dates has a causal effect on the number of downloads: most recent films are downloaded the most. The marketing power only influences what gets uploaded, but has no effect on the downloads. Film popularity, overall film quality (expressed in box office revenue) is also insignificant: we cannot pinpoint the effects of word-of-mouth information dissemination.

The explanatory power of the model has risen compared to the model that included all 5 categories, although it is still very low. This low explanatory power only reinforces our belief that the p2p markets and the cinema market are in fact two separate markets. Though there are common factors, such as marketing power, that control demand on both markets, we could find little communication between the two domains.

As for our hypotheses: we have found strong evidence both in the narrow and in the wide model that it is the time difference between cinematic and P2P releases that really defines to what extent a film is downloaded. H1 is therefore accepted.

Although we could not find a direct, causal effect of the marketing related variables to the number of downloads, we have established that it certainly defines what gets uploaded to p2p networks. The number of copies define what becomes available on p2p networks, therefore we can accept H2.

As for the connection between cinematic and online popularity: we have found no causal connection whatsoever between any of the cinematic and online popularity measures. H3 is therefore rejected.

## Conclusion

Cinematographic supply has a dual role in shaping downloading activity:

- its shortages expand its horizon
- its marketing efforts define its focus.

Box office numbers however fail to explain peer-to-peer demand for movies. While peer-to-peer supply is triggered by media presence, the p2p demand has little connection with the cinematic market. This lack of correlation suggests that the cinema distribution market has little to fear from the downloaders. Though p2p users react to the same incentives as cinemagoers, the two markets do not substitute each other.



Though the P2P market might have a significant effect on the DVD market, we suspect something altogether different than a direct substitution with either of the traditional distribution channels. What we witness here is the birth of a new distribution format. It is not TV, it is not really the infinite video library of The Pirate Bay, and it does not quite resemble the online video outlets either.

As the amount of archival content is limited, on these p2p networks it is clearly not the search activity that links suppliers (uploaders with a specific title) with the demand (prospective downloaders). Instead, the focal point of the user activity on a bittorrent hub tends to be the page which lists the latest, newest torrent files available through the tracker. Each hub offers a continuous stream of new content and users decide which they will download. This consideration can be strategic (if they download a title only to gain from sharing it to others) or can reflect a genuine interest in the title. Nevertheless, the traffic on each tracker is defined by the rhythm of new uploads. In this sense the users of torrent trackers resemble to a crowd of TV watchers, who consume what the programmer (those who control what gets uploaded) offers to them. On some sites the programming is democratic, as there are no restrictions on uploading. On other sites, users enjoy the selection of trustworthy release groups. Some sites even specialize along cultural, thematic, linguistic niches, setting up their own, thematic p2p channels.

The strong competition among a plethora of torrent trackers suggests that the true value of a p2p hub lies in its power to offer an attractive content bundle.

**References**