



**HAL**  
open science

## Film factory losses: is BitTorrent a major responsible?

Praboda Rajapaksha, Reza Farahbakhsh, Noel Crespi, Roberto Minerva

### ► To cite this version:

Praboda Rajapaksha, Reza Farahbakhsh, Noel Crespi, Roberto Minerva. Film factory losses: is BitTorrent a major responsible?. SNAMS 2019: 6th International Conference on Social Networks Analysis, Management and Security, Oct 2019, Grenade, Spain. pp.181-188, 10.1109/SNAMS.2019.8931858 . hal-02363514

**HAL Id: hal-02363514**

**<https://hal.science/hal-02363514>**

Submitted on 14 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Film Factory Losses: is BitTorrent a Major Responsible?

Praboda Rajapaksha<sup>\*†</sup>, Reza Farahbakhsh<sup>\*</sup>, Noël Crespi<sup>\*</sup>, Roberto Minerva<sup>\*</sup>

<sup>\*</sup>*Institut Polytechnique de Paris, Telecom SudParis, CNRS Lab UMR5157, Evry, France.*

<sup>†</sup>*Uva Wellassa University, 90000, Badulla, Sri Lanka*

{praboda.rajapaksha, reza.farahbakhsh, noel.crespi, roberto.minerva}@telecom-sudparis.eu

**Abstract**—Online piracy is an important challenge in the motion-pictures industry. Several studies have claimed that unauthorized content in online venues are reducing substantially the Box Office revenues while few other studies were not in favor of this claim. To understand better the impact of this phenomena early years, a study based on a large dataset is needed to analyze how different portals (e.g. BitTorrent) permitted movie downloads that help to increase revenue movie industry and it is also important to identify what type of movies were mainly affected. This paper aims to answer this question based on a dataset containing almost 15 million records obtained from around 3.25 million torrents’ data collected from the BitTorrent portal and their detailed movie related records extracted from IMDB. In this study, we observe (i) the impact of online movie downloads on movie revenues in early years, which predominantly affect on low budget and independent movies, (ii) the correlation between screen period of movies in cinema to the availability of torrents, (iii) the fake torrents that are injected to the portals before and during the screen period of a movie. Apart from that this work analyses the movie viewer’s feedbacks gathered from a questionnaire survey on user opinion and experiences about online movie downloads. We explored that, people used to be aware more about the online downloads and their related portals after introducing anti-piracy laws than before. We also suggest several other ways to help reducing online download rates of movies.

**Index Terms**—BitTorrent, motion pictures industry, movie’s BoxOffice, online piracy, anti-piracy law.

## I. INTRODUCTION

There exists a number of online portals such as BitTorrent allowing people to share and download media content, for instance video (movies, music, TV shows etc.) and audio files over the Internet. Previous studies have shown that a large portion of content available in these portals were videos [1] and unauthorized [2]. The main after-effects of sharing media files via these portals is that the community can stream or download without any charge or at a minimal cost. As a result, movie industry players arguing that their revenues are declining with the distribution of pirated copies in an online venues. As a consequence, the impact of unauthorized multimedia content shared in online platforms is a hot and debatable topic. In the past decades, number of researchers have analyzed the impact of online media content downloads to the movie revenue. Few studies have reported that this impact was very low yet other studies have proven that the online downloads have declined movies’ boxoffice revenue. In addition, few other studies have argued on benefits of online

file sharing (e.g. distribution of the pre-release hype of a movie) and elaborated that this phenomenon had increased the attraction of the movie viewers to visit the movie in cinema [3].

In this paper, we investigate the impact of keeping movies in online platforms by allowing them to download and its effect on movie budget and revenues. The analysis will be done on movies revenue losts with respect to the non-copyrighted compliant downloads. However, trying to study all the non-copyrighted content is too ambiguous as there exists many different online file sharing platforms, such as Cyberlockers (e.g. Uploaded, WUupload, Letitbit, etc) and streaming services (e.g VideoZZ). Hence in this study, we only focus on BitTorrent, one of the most successful file-sharing technologies during last decade.

We implemented an advanced data crawler for collecting a very large dataset from BitTorrent in order to monitor available torrents belongs to 18 major trackers for a period of 47 days. We collected 15 million(M) torrent records and among them, 3.25M records are unique torrents. We filtered and identified 926K qualified torrents that include 241K torrent files associated with movies. Their infohashes are used to crawl TorrentZ web portal to collect more information about each movie. BoxOffice information (production budget, BoxOffice revenue, etc.) of each identified movie is also collected from IMDB portal and other sources for analyzing correlation between movie downloads with their budget and worldwide gross. To sum up, in this study we implemented four different crawlers to obtain required information about unauthorized online movies. In addition, we conducted one questionnaire survey among 500 participants to understand their awareness about online piracy and what can be a motivation for them to have legal access to the movies instead of downloading them unlawfully.

The main contributions of this study are:

- (i) propose a methodology to identify the impact of online downloads on the film industry and their revenue. This methodology can be generalized across other domain such as music and software market.
- (ii) obtain a large dataset of torrents that are associated to movies, music, and software (available for further research<sup>1</sup>).

<sup>1</sup>Please contact authors of the paper to obtain our dataset for further experiments.

(iii) introduce a set of observations that helps to mitigate movie lost based on the solutions analyzed from the feedback received for a large questionnaire.

## II. RELATED WORK

A large number of literature have investigated the impact of online piracy on different sectors including movie and music markets. Most of these studies were based on small datasets collected from questionnaires or/and surveys and only a few of them have used a real dataset collected from movie portals. Based on our knowledge, limited number of studies have considered very large amount of information to analyze music industry, DVD trades, and box-office gross.

Online piracy is associated with sharing copyrighted content without authors consent in the Internet. Researchers use various approaches to find the relationship between legal and piracy content. A book published by Gunter et al. [4] presented a detailed description and a survey on how movie success can be predicted at the Box Office. They briefly reviewed the history of movie-making, the growth of movie-going and global trends in movie attendance in understanding what factors underpin the success of a new movie. In another study [5], authors have measured the relative impact of movie piracy on motion picture industry and shown that, if piracy could be eliminated then the Box Office revenue would increase by 15%. Hollywood movies are released first in US and subsequently returned for further screening in other countries. BitTorrent and other similar portals distribute those movies illegally to other countries before their official release. An analysis to this phenomena is the study carried out by [6], across 17 countries. They have shown that longer international release lags declined Box Office revenues by at least 7% in the presence of pre-release of movies. A very recent work [7] studied an emerging streaming cyberlocker ecosystems (e.g. openload.co, thevideo.me and vidzi.tv.) and discovered that these environments actively involved in copyright infringement with an aggressive injection of recent releases of movies. As these cyberlockers ease of use attract a large number of users similar to traditional systems such as Gnutella and BitTorrent. Therefore Motion Picture Association of America (MPAA) have shifted their efforts towards shutting down the cyberlockers themselves. The study done by Marc R. [8] used multiple data sources to collect more information about movies Box Office related information and unauthorized CAM copy downloads from BitTorrent. The results have proved that CAM downloads had a low probability of equating to a lost sale. Apala et al. [9] predicted movies Box Office performance using data collected from social media such as Twitter, YouTube and IMBD portal.

Several major actions have been taken by governments and major content providers and companies to reduce the availability of the illegally published copyrighted content in online platforms. As an example, Danaher et al., [10] investigated the impact of ceasing the major piracy sites such as Megaupload and associated Web portals that are used for movie sales. Their study shown that, because of these close

down events, revenues of three major motion picture company increased by 6.5-8.5%. Apart from that, the study [2] identified how antipiracy actions effect on online piracy by considering closure of Megaupload and the implementation of the French antipiracy law (Hadopi). They have shown that Megaupload closure diminished the activities of professional BitTorrent publishers who are running their own BitTorrent portals.

Considering the stated literature, this paper aims to look in to this problem from a deeper perspective by using a large scale dataset collected from several relevant sources including BitTorrent and IMDB. We use a novel simple methodology to measure this impact and provide some guidelines on how this impact can be mitigated based on the survey results.

## III. DATA COLLECTION METHODOLOGY

This section describes our data collection methodology and different steps used for preparing a final dataset including data filtering, classification of movies, and procedure of collecting movies metadata.

### A. Raw data collection

In order to collect movies information from BitTorrent, we have implemented an advanced crawling tool that connects to different trackers. Tracker keeps track of the locations of the file copies and coordinate among those files efficiently, and reassemble copied files using trackers' scrape mode. A scrape or tracker scrape is a request sent by a BitTorrent client to a tracker and then a connection to the tracker is established, information is exchanged and finally the connection is closed.

Two filtering steps are executed to select a most suitable tracker list from more than 100 unique trackers. Firstly, we investigated from the dataset crawled from BitTorrent, a group of URLs or peer machine addresses referred to a same tracker (analyzed using dig utility) or on the other hand, for instance *publicBT* tracker refers to a wide number of URLs. Thus we filtered unique trackers among them. Secondly, we identified and excluded a group of trackers those who restricted to access via scrape mode as our crawler works only with the scrape mode. We also excluded some trackers available in our dataset as their scrape file was in a very strange codification. Finally, we identified the most appropriate 18 trackers and queried following information about the torrent lists for each trackers. i) infohash (a unique ID of a torrent file); ii) number of downloads; iii) number of seeders at present (i.e. torrent files those who have completed one download and in swarm to send pieces of the content to others); and iv) number of leechers at present (i.e. torrent files those who have not completed its download).

This crawling process executed 6 times per day in every 4 hours to collect scrape file information (around 15 million records of 3.25M infohashes) and store details of number of downloads, seeders, and leechers. As a summary, table I lists the names of 18 trackers, number of torrents, and number of movies available for each tracker (identified in Section III-C). Next section explains how we collected information of each

torrent file such as file name and type of the content (movie, music, etc.).

### B. Data filtering

We have collected around 3.25M unique torrent information and among them, a large number of torrents attracted very few users (in many cases number of downloads, seeders and leechers are 0). Therefore, following policies are used to filter torrents that are having considerable amount of downloads, seeders and/or leechers.

- i) Infohashes having more than 500 downloads and
- ii) Torrents having more than 10 seeders or 10 leechers if they do not have any downloads.

After these filtering steps only 926K torrents were identified in order to analyze torrents with reasonable amount of metadata.

### C. Modeling number of Downloads for a Tracker

As was mentioned earlier, tracker's response includes number of downloads, seeders, and leechers. However, few trackers (4 out of 18 in our study) do not provide real number of downloads as they reset download count frequently yet gives correct number of seeders and leechers. However our inspections on the snapshots with valid number of downloads shown that number of leechers in each timestamp is almost equal to the number of downloads. Therefore we use the value of number of leechers to represent number of downloads for torrents those who have not provided download count. Following steps are also used to calculate number of downloads per torrent when a snapshot missed the value of download count but, have a valid downloads counts in the early snapshots.

- i) If #downloads is zero for the present crawling period, then the previous value (snapshot of 4 hours ago) is considered.
- ii) If #downloads is smaller than previous value (when the tracker reset #downloads), sum of those two values are considered.

### D. Content classification

Another crawling application is implemented to query Torrentz portal indexing the infohashes. The main idea of data collection in this phase is to differentiate type of the content (e.g. Film, TV Show, Game, etc) of each torrents. This tool collects information such as name, date when it was first available in one of the torrent portals, type (e.g. movie, video, video TV, game, audio, software, picture, porn) etc. For example, infohash: 0000bf2359476d55166d891c22d61cc48af6df1a has details such as infohash name: "The Girl With The Dragon Tattoo" (2011), torrent category: movie and added date to the portal: 2012-01-18.

Table II summarizes the distribution of torrents based on the obtained data from TorrentZ portal. It shows a large number of torrents (more than 50%) are from movies and videos categories. The analysis of our study is based only on the movie category which includes 241, 129 movie torrents. As a

TABLE I  
TRACKERS' INFORMATION AND INITIAL NUMBER OF COLLECTED TORRENTS FROM EACH TRACKER

Tracker-Name	#Torrents	%Movies
blazing	9,783	31
podtropolis	49,698	40
eztv	580,555	25
firebit	8,614	41
publicbt	579,581	25
vtv	579,354	25
harry	568,022	25
torrentbay	238,897	32
fr33dom	184,499	32
xxx	34,537	15
exodus	279,092	29
opensharing	24,623	62
yify	11,000	47
hotplug	31,577	38
9you	6,498	34
elitezones	2,218	17
mytracker	12,741	37
anime	51,058	21
<b>Total</b>	<b>3,252,347</b>	<b>Avg=32%</b>

TABLE II  
TORRENT CATEGORIES (BASED ON TORRENTZ PORTAL CATEGORIZATION)

Category	#Torrents	%Torrents
Movie	241,129	26.03
Video	240,956	26.01
None	240,652	25.98
Audio	92,321	9.97
Software	43,369	4.68
Games	30,238	3.26
Ebooks	27,747	3
Porn	1,078	0.12
Pictures	8,931	0.96
<b>Total</b>	<b>926,421</b>	<b>100</b>

potential future work, the research question of this paper can be studied for other categories such as game and Software.

### E. Collecting movies information

For each identified movie, we collect movie budget and revenue information by using available APIs from IMDB portal (filter by film name or IMDB-ID) and obtained more details such as IMDB unique ID, movie title, production year, released date, country, director, actors, etc. In addition, another crawler is implemented to collect box-office information from:

- (i) "IMDB business" portal,
- (ii) "The-Numbers" portal, and
- (iii) "Worldwide Box Office" portal.

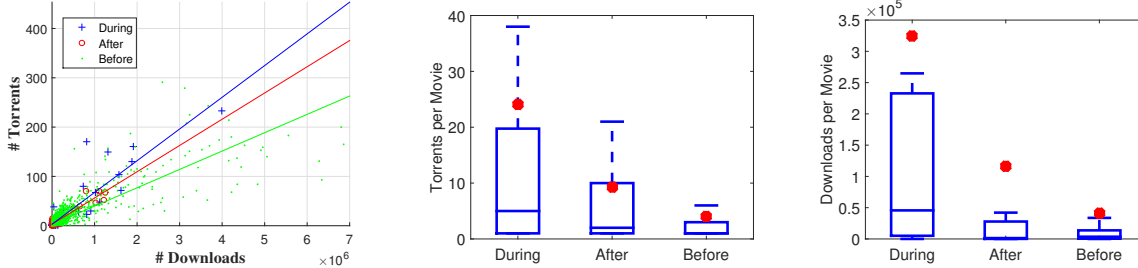
All Box Office information acquired from above web portals are amalgamated to gather accurate information. This approach allows us to collect complete Box Office details for each IMDB-ID including movie's budget, domestic and overseas sales, worldwide gross revenues, domestic DVD sale, released date of the movie, etc.

### F. Final Datasets

Table III summarizes our dataset collected during 47 days starting from 2012-02-17 and we categorized movies into the 3 groups: during, after and before.

TABLE III  
MAIN FIGURES OF THE DATASET AFTER FILTERING AND CONTENT IDENTIFYING STEPS

Facts	Figures
Start date of data collection campaign	17/02/2012
Crawling period (#days)	47
#snapshot collected per day	6 (Q4h)
#trackers	18
#snapshots (total collected infohashes)	15,079,905
#Unique torrents (unique infohashes)	3,252,347
#torrents (after filtering step)	926,421
#torrents corresponding to movies (TorrentZ step)	241,129
#torrents corresponding to movies (IMDB step)	59,500
#movies (unique movies identified from IMDB step)	14,823
#movies (& #torrents) in <i>During</i> category -“released during the crawling period”	81 (1,240)
#movies (& #torrents) in <i>After</i> category -“released after the crawling period”	124 (682)
#movies (& #torrents) in <i>Before</i> category -“released before the crawling period”	14,618 (57,578)



(a) Linear distribution of #torrent vs #download (b) Distribution of #torrents per Movie (c) Distribution of #download per Movie  
Fig. 1. Availability of torrents (#torrents) vs. popularity of torrents (#downloads) across movies (the red dots show the average value)

(1) *During*: All movies with the released during 47 days of our crawling period (2012-02-17 to 2012-04-04) and all associated torrents to these movies that are uploaded after movie’s released date (1,240 torrents, 81 movies).

(2) *After*: All movies with the released date after crawling period (682 torrents for 124 movies). This category includes all torrents uploaded before a movie is released. Our investigation shows a large portion of this types of torrents were available in the online portals.

(3) *Before*: All movies with the released date before crawling period (57,578 torrents, 14,618 movies).

Torrents that are uploaded to the BitTorrent portal might publish (i) before releasing a movie or (ii) after releasing a movie, and this study is based on the later category.

#### IV. NO.OF DOWNLOADS VS. MOVIES’ BOX OFFICE

In this section we investigate how movie downloads affect on movie’s worldwide gross and production budget.

##### A. Availability vs. Popularity

First, we explore a relationship between availability of the torrents in BitTorrent and their popularity (number of downloads) for *During*, *After*, and *Before* categories as depicted in Figure 1. Figure 1(a) shows a linear relationship between total number of movie downloads and torrent uploads. The first observation is that, many movies with large number of downloads (>4M) are from the *Before* category ( $r=0.84$  where ‘r’ represents correlation coefficient). Based on the linear regression analysis,  $r^2$  for *Before* category is 0.706 ( $r^2$  represents coefficient of determination), which indicates

that 70.6% of the total variation in number of torrents can be interpreted by the linear relationship between number of downloads and number of torrents. This is reasonable since after a movie is released, there is a higher chance of increasing the number of torrents uploaded to the BitTorrent.

Figures 1(b) and 1(c) show distributions of number of torrents and downloads per movie. In the *During* category, there were more uploaded torrents and more downloads compared with other categories. This indicates that, recently released movies were more popular among content publishers and consumers. On average, 24 torrent files were uploaded to BitTorrent in *During* category which is almost 3 times more than the *After* category (9 Torrents in Average) and 6 times more than the *Before* category (4 Torrents in average). We also identified top 3 movies having the highest number of torrents than other movies that are associated to the *Before* category (“*Journey 2: The Mysterious Island*” - 291 “*Ghost Rider: Spirit of Vengeance*” - 279 “*Mission: Impossible - Ghost Protocol*” - 243). Another interesting observation from Figure 1(a) is that many movies having at least 200 torrents exhibited less than 2M downloads. As shown in Figure 1(c) most movie downloads occurred in the *During* category (average - 0.35M downloads) indicating many movies were downloaded just after the release.

As indicated in Table III, the largest number of movies and number of torrents are associated to the *Before*, but average movie download count in this category is 411k, which is 10 $\times$  and 3 $\times$  lower than the *During* and *After* categories, respectively. Maximum number of downloads per movie is from *Before* category(almost 6.9M), and these were widely

popular and already released movies.

To summarize here, although newly released movies were popular (number of downloads) and available (number of torrents) in BitTorrent some other movies are very popular, but not yet released.

### B. Worldwide Gross vs. Budget and Downloads

This section focuses on the impact of movie downloads to the worldwide gross, considering their initial production budget. Figure 2 shows a 3D plot of the movies worldwide gross in compared to number of downloads and their budget. This graph clearly indicates the effect for the worldwide gross in different movies with respect to their number of downloads and budget. We noticed that movies having a high worldwide gross were the once with very large budget and had few number of downloads. Figure 2 also indicates that movies with low budget and/or less number of downloads had much lower worldwide gross. Movies with a large number of downloads were less costly or independent movies. Regression coefficient- $r^2$  between movie budget and worldwide gross is 0.435, means only 43.52% of the movie's worldwide gross can be predicted using its budget. One reason behind this might be the copyright controls issued by movie production companies on popular movies with the goal to reduce unauthorized copies in online platforms. Besides these facts, there is no any liner relationship between movie's worldwide gross and movie downloads.

In summary, low budget movies were downloaded more than the other movies and had very low worldwide gross, and their revenues had impacted a lot due to online downloads.

On average, many movie downloads and torrent uploads are from the *During* category. Therefore, in-order to have a better understanding of the above analysis, we look more in detail to the *During* category as shown Figure 3. In contrast with the analysis performed for all categories (Figure 2), results from the *During* category presents that (Figure 3), higher the number of downloads per movie higher the worldwide gross ( $r^2=0.735$ ), but still a many number of low budget movies were also shown to have large number of downloads.

## V. POTENTIAL LOSS ANALYSIS

This section analyses movies losses (absolute and relative losses) based on BitTorrent downloads focusing on the *During* category.

### A. Absolute Loss

In our hypothesis, in order to study absolute loss we assume one BitTorrent download as a potential audience who could pay for a movie in cinema. Therefore BitTorrent downloads are proportional to  $\sum_{j=1}^n X_i^j$ ; where cinema viewer 'i' visit theater for  $j^{\text{th}}$  time (up to n times) and this is equivalent to a user who may view the same movie multiple times from online platforms. The analyses are focused on the *During* category as many downloads were occurred during that period while screening the movie and few month later. Despite that, some movies in the *During* category continue their screening after

our crawling period, but 62% of the identified movies in this category were released before the first half of the crawling period. As a result, the probability of identifying a movies' download patterns is very high just after a movie is released.

According to the statistics of the The Hollywood Reporter, average cinema ticket price during the second quarter of 2012 (our crawling period) was 8.12\$. Hence, we consider 8.12\$ as the ticket price of movies in-order to calculate the total amount of losses using above hypothesis as follows:

$$Absolute\ Loss = \sum_{j=1}^n X_i^j \times Ticket\ Price$$

Figure 4 elucidates the distribution of the absolute losses for the *During* category. We assume that, each user stream/download any video only once ( $j=1$ ). The average movies' potential loss is around 4.81M\$ in the *during* category. However, several movies appeared to have a huge loss; e.g. "*John Carter*" movie had the largest absolute loss, around 32M\$ (Maximum value), and 4M online downloads, where the loss of this movie can be considered as approximately 1/8 of its worldwide gross. This signifies that, even though a movie had high revenue, still there is a possibility to increase its income by limiting online downloads.

In summary, movies can increase their income if people visit cinema instead of downloading them from online portals or else online portals can charge from the users per stream/download by keeping them as legitimate copies.

### B. Relative Loss

Absolute loss analysis is important to measure the relative impact of BitTorrent downloads on movie revenues and we define *HIR* (Harmful Income Ratio) in-order to understand this impact.

$$Harmful\ Income\ Ratio\ (HIR) = \frac{AbsoluteLoss}{ActualGross}$$

$$Actual\ Gross = Worldwide\ Gross - Movie\ Budget$$

Negative values of *HIR* metric represent movies with very large revenue losses and usually their worldwide gross is lower than its budget. On the other hand, positive value of *HIR* represent movies with high worldwide gross and might have higher absolute loss or lower actual gross compared to other movies. If the *HIR* value is approximately equal to 0, then a movie might has considerably a large worldwide gross regardless of the number of downloads. Also, if the *HIR* is 1, then a movie might has similar figures for the actual gross and absolute loss indicating that the movie had lost an equal amount of money as its revenues. We calculated *HIR* for the most affected movies in the *During* category identified from their absolute losses and analyzed correlation between *HIR* with number of downloads, worldwide gross and movies' budget. The results are shown in Figure 5.

Figure 5 illustrates the relationship between movie's budget, number of downloads, worldwide gross, and *HIR* in a 4D graph. The *HIR* values are represented by different colors in the graph and shows the variability of *HIR* with other

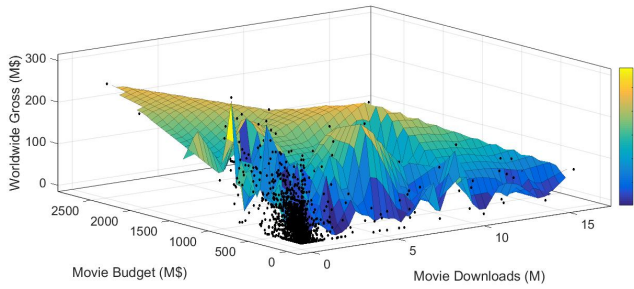


Fig. 2. WW Gross vs. budget and #downloads for all categories

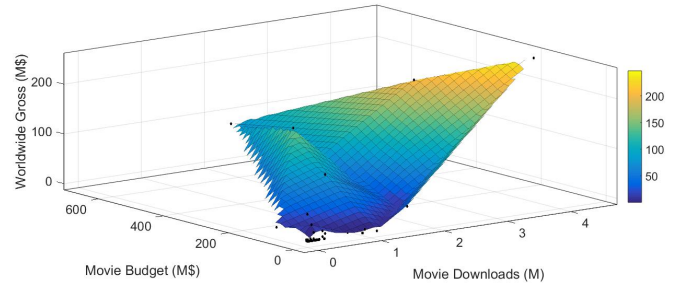


Fig. 3. WW Gross vs. budget and #downloads for *During* category

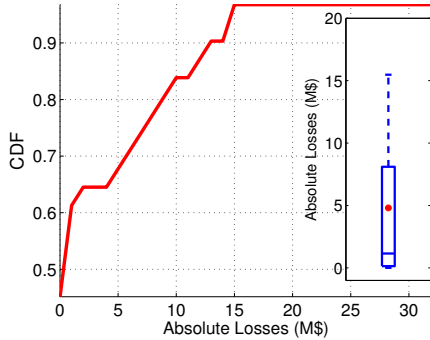


Fig. 4. Distribution of Movie's Viewer Losses

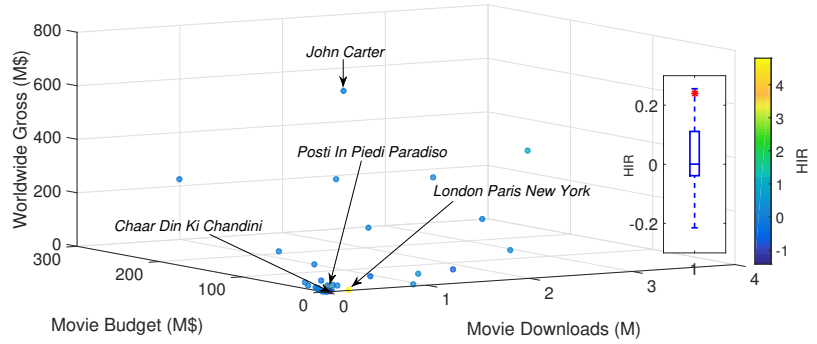


Fig. 5. Harmful Income Ratio (HIR) vs (a)#downloads, (b)Movie budget (M\$), (c)Worldwide gross (M\$)

variables. There were few movies with negative *HIR* and positive *HIR*. However on average, *HIR* in many movies in the *During* category is almost zero (mean: 0.24). We also can identify the maximum *HIR* value (4.8) having a huge absolute loss and very low actual gross belongs to “*London Paris New York*” movie. This movie had 1.7M\$ worldwide gross during the crawling period and 1.3M\$ budget, and only 0.25M number of downloads. Even though this movie had number of downloads than the average number of downloads in the *During* category (0.32M), absolute loss is higher than the actual gross. The next largest positive *HIR* belongs to “*Posti In Piedi In Paradiso*” movie followed by “*London Paris New York*” movie.

On the other hand, the most negative *HIR* (-1.43, the darkest blue marker in Figure 5) represents “*Chaar Din Ki Chandni*” movie. This can be considered as an unpopular movie. The worldwide gross of this movie is less than the budget and shown 68,142 number of downloads. The “*John Carter*” movie indicates its *HIR* almost equal to 1. The highest number of downloads (4M), the largest budget (250M\$) and around 283M\$ worldwide gross is from this movie. As its *HIR* is equal to 1 its absolute loss and actual gross are almost equal.

In summary, the mean *HIR* for many movies from the *During* category is 0.24 indicating that they have very high actual gross compared to absolute loss. Movies those that released before crawling period had obtained considerable amount of revenues regardless of the number of downloads. In our future works, we will crawl more recent data to analyze how these movies revenue changed during these years.

## VI. QUESTIONNAIRE AND CONCLUDING REMARKS

To better understand user opinions and experiences about online downloads, we conducted a survey (<https://goo.gl/Mk6iiZ>) and asked people about their download activities. This section briefly presents some of these findings and results.

### A. Findings from the people feedbacks to a Questionnaire

The survey had received around 500 replies from people across different; countries, age groups, and internet expertise levels. The main objective of this survey was to: (i) understand user download patterns from online sources (ii) understand user interests on visiting to cinema or willingness to pay for a video content instead of downloading illegally, (iii) and finally have a vision on the awareness of people about anti-piracy events related to illegal content availability and sharing.

The main findings from this survey are as follow. First of all, we observed that people had very high awareness among about the antipiracy activities conducted to limit available unauthorized content. This observation might came due to the awareness of Megaupload closure (on 19 January 2012). Megaupload was one of the largest online file sharing application which caused to reduce revenue of the entertainment industries about US\$500 millions. According to the statistics, the shutdown of Megaupload and its associated sites caused digital revenues for major motion picture studios to increase by 6.5-8.5% [10]. Our survey findings shown that most of the users with age < 50 years (around 84%) had experience of using the Internet and were aware about Megaupload closure. Among all the people who attended our

TABLE IV  
SURVEY RESULTS OF NUMBER OF CINEMA VIEWERS FOR DIFFERENT  
TICKET PRICES

Price (\$)	0	1	2	3	4	5	6	7
#viewers	35	22	39	67	69	116	32	76
Probability	1	0.92	0.88	0.79	0.64	0.49	0.24	0.17

survey in the age group <20 years were also aware about the Megaupload closure. Moreover, 97% of the surveyed people are sensible on this closure and they have used Megaupload to download multimedia content many times. Also, all French nationalities take part in this study had the knowledge on the anti-piracy law introduced by the French Parliament namely HADOPI. In summary, these findings indicate that enforcing this type of law/rules for pirated contents is positive effect to limit non-paid consumption.

Based on the results obtained from this survey, many people were refused to download media content illegally from the Internet due to different laws established in their countries. More than 53% of the advanced and expert Internet users were not downloaded illegitimate content due to the punishments undertaken (e.g. Internet disconnection, high fines, sending to the prisons, etc.). This indicates that introducing punishment rules in country level can increase people concern and help to reduce number of downloads as detailed in [2]. Almost half of the participants were not considered that the Internet disconnection was as a major reason for not downloading from online portals and 25% of them claimed that, nothing can prevent their downloads.

In summary, people willing to download online multimedia content but they refuse to undertake any punishment. One approach is to allow consumer to download/view multimedia content by paying less amount of money than the actual prices i.e. Netflix providing benefit for both content provider and consumer. Content provider can deliver services to the registered users (who paid membership fee) and they can also limit download capacity of the files or number of downloads per some period (one month) for each user. Moreover, ad-supported streaming and downloads are able to monetize free consumption behaviors.

#### B. How to increase cinema viewers?

One possible solution to attract cinema viewers is to observe their preferred ticket prices to pay in cinema. We have collected these information from our survey and Table IV presents its summary including number of movie viewers for different ticket prices. Among them, 25% of the users were preferred to visit cinema if the ticket price is 5\$ (ticket price range: 0\$-7\$) without downloading the movie. We use following formula to analyze how to increase movie revenues by increasing number of cinema viewers and reducing online downloads in *During* category.

$$\text{Movie Revenue} = \text{Gross} + X\% \times \#\text{downloads} \times \text{Ticket price}$$

Where  $X$  represents cumulative probability of number of cinema viewers for different ticket prices where we assumed that if a user is willing to pay 7\$ then she visit cinema for

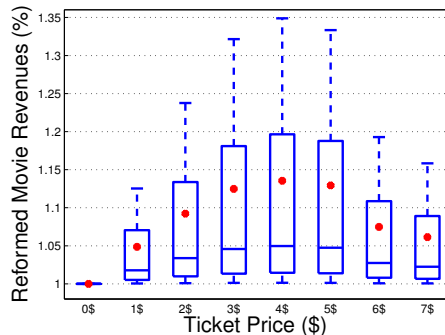


Fig. 6. Overall movie analysis of the *During* category for different cinema ticket prices

lower prices as well. With this assumption, we calculated total movie revenues ranging the ticket price from 0\$ to 7\$ and results are shown in Figure 6. It demonstrates the fluctuation of the revenue for all movies in *during* category with reference to different ticket prices. The highest ticket price and maximum movie revenue can be observed when the ticket price equals to 4\$ and movies' income can be increased by 13.56%. Similarly, a small variation to this result can be observed when a ticket price is 3\$ (by 12.48% increase of the movie revenue) and 5\$ (by 12.94% increment movie revenue). Thus, we can conclude that movie income decreases if the cinema ticket price is very low and very high and keeping a middle range value (5\$) can increase the income of movies by 13.56%.

In summary, there are different ways to increase the revenue of movies Box Office and one possible way is to reduce ticket price and keep in an average price not too high or too low. This will encourage at least a few people to visit cinema.

## VII. CONCLUSION

This paper examine the impact of online movie downloads to the motion picture sales using a collected dataset from BitTorrent and from other portals (e.g. TorrentZ, IMDB portal, and Worldwide Box Office portal). The analyses indicated that soon after releasing a movie many torrent files are uploaded to online portals and many people download it from these portals, indicating that recently released movies are popular among movie viewers and publishers. We also explored that low budget movies were downloaded more than the others and had a very low worldwide gross. Therefore, low budget movies had loss much of their revenues due to online portals such as BitTorrent. In addition, based on a community feedback to a questionnaire, we concluded that although there are solutions such as shutting down a portal or enforcing anti-piracy laws that can reduce number of downloads, but people are willing to support other solutions such as low price online possibility to watch a movie instead of downloading it and also willing to pay for promotional tickets to visit cinema.



## REFERENCES

- [1] R. Farahbakhsh, N. Crespi, Á. Cuevas, R. Cuevas, R. González, Understanding the evolution of multimedia content in the internet through bittorrent glasses, *IEEE Network* 27 (6) (2013).
- [2] R. Farahbakhsh, A. Cuevas, R. Cuevas, R. Rejaie, M. Kryczka, R. Gonzalez, N. Crespi, Investigating the reaction of bittorrent content publishers to antipiracy actions, in: *Peer-to-Peer Computing (P2P)*, 2013 IEEE Thirteenth International Conference on, IEEE, 2013 (2013).
- [3] B. Cammaerts, B. Meng, R. Mansell, *Copyright and creation: a case for promoting inclusive online sharing* (2013).
- [4] B. Gunter, *Predicting movie success at the box office*, Springer, 2018 (2018).
- [5] L. Ma, A. L. Montgomery, P. V. Singh, M. D. Smith, The dual impact of movie piracy on box-office revenue: Cannibalization and promotion, Available at SSRN 2736946 (2016).
- [6] B. Danaher, J. Waldfogel, *Reel piracy: The effect of online film piracy on international box office sales* (2012).
- [7] D. Ibosiola, B. Steer, A. Garcia-Recuero, G. Stringhini, S. Uhlig, G. Tyson, *Movie pirates of the caribbean: Exploring illegal streaming cyberlockers*, in: *Twelfth International AAAI Conference on Web and Social Media*, 2018 (2018).
- [8] M. R. Milot, *Testing the lost sale concept in the context of unauthorized bittorrent downloads of cam copies of theatrical releases*, *Browser Download This Paper* (2014).
- [9] K. R. Apala, M. Jose, S. Motnam, C.-C. Chan, K. J. Liszka, F. de Gregorio, *Prediction of movies box office performance using social media*, in: *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ACM, 2013, pp. 1209–1214 (2013).
- [10] B. Danaher, M. D. Smith, *Gone in 60 seconds: The impact of the megaupload shutdown on movie sales*, *International Journal of Industrial Organization* 33 (2014) 1–8 (2014).