IBS WORKING PAPER 03/2018
APRIL 2018

PRE-RELEASE LEAKS AS ONE-TIME INCENTIVES FOR SWITCHING TO UNAUTHORISED SOURCES OF CULTURAL CONTENT

Wojciech Hardy

ibs working paper 03/2018
april 2018

PRE-RELEASE LEAKS AS ONE-TIME INCENTIVES FOR SWITCHING TO UNAUTHORISED SOURCES OF CULTURAL CONTENT[•]

Wojciech Hardy[♦]

Abstract

Pre-release leaks of cultural content incentivise consumers to look for unauthorised sources. I find that such events may induce some television viewers to switch to unauthorised sources to gain access even to content that had not been leaked. To demonstrate that this is the case, I use a unique dataset on a sample of TV shows aired around the time of a pre-release leak of a very popular TV show (Game of Thrones). The results of a difference-in-differences analysis indicate that the leaked TV show lost viewership for both the leaked episodes and those that followed. Moreover, the event also had negative effects for other TV shows that may share an audience with the leaked show. Finally, my results for the shows with a shared audience are corroborated by evidence of an increase in Google searches for phrases including the show names and the words "*watch online*", after the leak. I argue that the one-time incentive to use unauthorised sources caused some viewers to engage in unauthorised consumption even of shows not affected directly by the leak. These conclusions are consistent with the existence of one-time costs of switching channels of content acquisition.

Keywords: File-sharing, copyright, intellectual property rights, TV, piracy, game of thrones

JEL: D12, K42, L82, O34, Z11

[•] I would like to thank Iga Magda of IBS, the participants of the DELAB 2017 and WIEM 2017 conferences, as well as of the Faculty of Economic Sciences, University of Warsaw, doctoral seminar and IBS internal seminar for their valuable comments. I would also like to thank the IMDb team for pointing me to the IMDb public use files. This study has been conducted as part of a research project financed by the National Science Centre (grant UMO-2016/21/N/HS4/01803). The usual disclaimers apply. All errors are mine.

[♦] Institute for Structural Research (IBS), Warsaw, and Faculty of Economic Sciences, University of Warsaw. E-mail: wojciech.hardy@ibs.org.pl.

1. Introduction

As digital markets for cultural goods continue to evolve, file-sharing remains a hot topic for content providers. Since the famous case of Napster at the beginning of the 21st century, the entertainment industries have largely tackled the problem of unauthorised distribution by introducing digital distribution of their own. Indeed, subscription-based services and low-cost digital files have levelled the field and allowed consumers to easily access content from authorised channels. However, the unauthorised sharing of content has continued to evolve, and has shifted towards mobile consumption and streaming. This trend is reinforced by increasing internet connection speeds, which allow for the easier distribution of high-definition content. With the continuously changing digital market and the entrance of new players (such as digital distribution companies), the growing challenge for entertainment industries is to incentivise non-paying consumers to switch to authorised sources, as well as to retain current consumers.

The lack of an authorised alternative can, however, be a strong incentive for consumers to switch to the unauthorised channels. A few previous studies have shown that a pre-release leak of content can significantly reduce levels of authorised consumption in the days immediately after the official premiere. Similar situation also takes place is often exacerbated by due to non-synchronised release dates across the world, as the content may be shared on the internet before it is officially released in some countries. To address this problem, publishers have sometimes withdrawn their content from the digital providers, which usually leads them to incur large losses. In such a situation, the lack of an authorised alternative may cause the would-be purchasers to turn to file-sharing networks to gain access to the content. In this study, I seek to contribute to the current state of knowledge on the relationship between file-sharing and sales by analysing the potential of temporary events to incentivise consumers to switch to unauthorised sources.

A one-time incentive to download from the internet might have long-term effects. Danaher et al. (2010) observed that a temporary removal of NBC station programmes from iTunes may have permanently caused the network's viewers to turn to "piracy" networks to gain access to the programmes. The potential reasons for this effect include one-time technology costs (e.g., learning how to use the P2P software), know-how acquisition (e.g., learning where and how to search for free content), and one-time moral costs (e.g., feeling less morally repelled by the act after doing it the first time). In such cases, the content of other providers may also be negatively affected, since these networks share some of their viewers with those of the downloaded NBC programmes. Unfortunately, while the findings of Danaher et al. (2010) partially support this hypothesis, the authors admit that their data do not allow for a comprehensive analysis of the effect.

An opportunity to observe such an effect presented itself with the pre-release leak of episodes of one of the highest-grossing serials: Game of Thrones [GoT]. On 11 April 2015, four episodes of the hit TV show were leaked to the web a day before the season's official premiere. As during the season just one episode was released per week, the leak allowed consumers to access the four episodes from one day to three weeks before their TV premieres. This leak therefore created a strong incentive for viewers to watch the episodes early, rather than to wait for the broadcast.

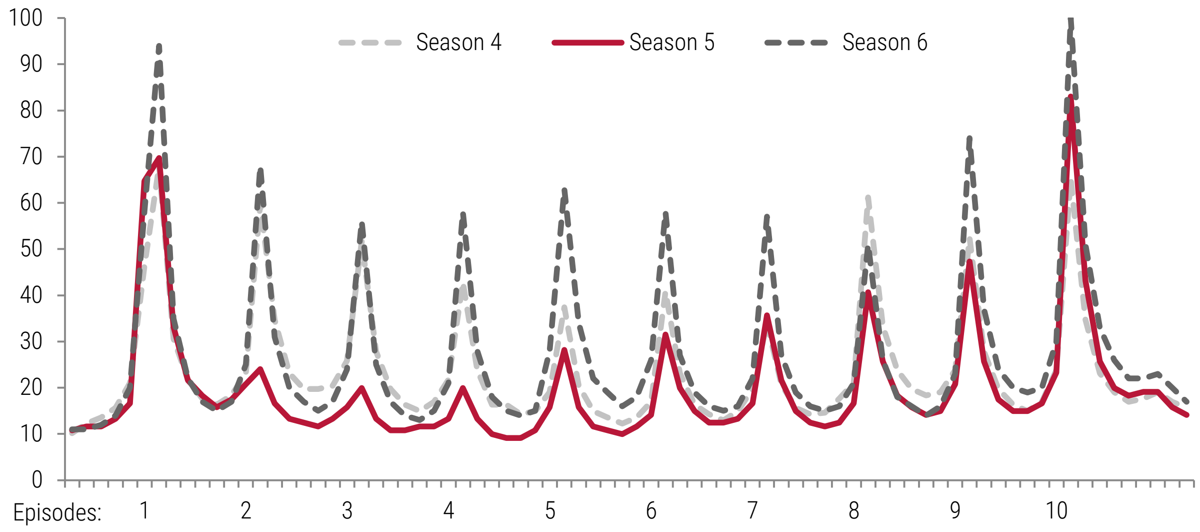
It indeed appears that the attention of GoT fans who were searching for information about the show on the internet largely shifted from the air dates of episodes 2-4 to the time of the leak. Google Trends[1] data suggest that relatively few people looked for any GoT-related terms on the web at the time episodes 2-4 of season 5 were being released, but that a relatively large number of searches took place on the day of the leak and on

[1] The Google Trends application allows users to track the Google search popularity of specified phrases. For more on the popularity index, see Section 4.

the following day. This pattern is very different from that of the GoT web searches that were conducted during the previous (fourth) and the following (sixth) season (see Figure 1).

TV shows offer us a unique opportunity to study the effects of shocks in unauthorised availability. A TV show is a good that its experienced over time. As such, by looking at the authorised and the unauthorised popularity of TV shows, we are able to observe roughly the same audience over time, and to observe the changes in their viewing patterns due to exogenous events. The viewers of popular ongoing serials can also be expected to be more susceptible than other kinds of consumers to incentives such as pre-release leaks, since episodes often end with so-called "cliff-hangers" aimed at increasing anticipation of the follow-up. In addition, analysing individual TV shows allows us to control for their characteristics.

Figure 1. Google Trends data for the Google search popularity of the phrase "Game of Thrones"



Note: The data for the three-season period are displayed with weekly values in the Google Trends application. I exported the daily data for each season separately, and then rescaled them using whole-period weekly data so that the popularity levels of seasons 4-6 are comparable (Google Trends sets the largest value at 100 for each separate search). There was a two-week gap between the seventh and eighth episodes of season 4 – the graph presents two-day averages for this season in this period. The start dates of seasons 4, 5, and 6 were 2014-04-06, 2015-04-12, and 2016-04-24, respectively (synchronised here for convenient comparison).

Source: Own elaboration and calculations based on Google Trends data.

In this article, I examine how the leak of GoT episodes affected the show's authorised and unauthorised viewership, as well as the authorised and unauthorised viewership of other TV series, and whether these potential effects persisted after the incident. I contribute to the current state of knowledge by exploiting the continuous nature of TV shows as consumption goods. I extend the current findings on the effects of pre-release leaks to the case of individual TV shows. Moreover, I am able to analyse whether the GoT leak affected the viewership and unauthorised consumption of TV shows that share an audience with GoT. This setting allows me to study in depth whether the move to unauthorised sources led to a persistent shift in viewers' behaviour, as suggested in Danaher et al. (2010).

In the second section of the article, I review the most relevant literature on the topic of pre-release leaks and TV shows in the digital age. In the third section, I provide a brief description of the GoT TV show and outline the methodology. In the fourth section, I provide a description of the data collected for this study. In the fifth section, I present the results of the analyses. I then close the article with a discussion of the results.

2. Literature review

Internet file-sharing and its effects on sales have been studied in various contexts, with the typical finding being that file-sharing hurts sales. Negative effects of file-sharing have, for example, been found by Danaher & Smith (2014) on digital movie sales, by Barker & by Maloney (2012) on music CD sales, and by Ma et al. (2014) on box office revenues. The reasons for these effects seem intuitive: unauthorised channels allow consumers to acquire their content free of charge, and often in ways that are more convenient than using the official distribution channels. Thus, file-sharing is clearly a valid concern for most entertainment industries.

One specifically harmful form of this activity is file-sharing that occurs in the absence of a legal alternative. In many of these cases, the unauthorised channels make content available even before it gets to the official channels. Such events can often be attributed to leaks of review-only copies of the content or to leaks in the factories that produce the physical copies (see Witt, 2015). The pre-release leaks tend to cannibalise potential sales, as the most eager consumers turn to file-sharing rather than waiting for the official channels to release the content. Ma et al. (2014) observed such effects for movies coming to cinemas.[2] Moreover, file-sharing may occur when a producer pulls content from a popular distributor. For example, Danaher et al. (2010) found that levels of file-sharing increased when NBC removed content that had been hosted on iTunes, and Hiller (2016) showed that a wave of file-sharing occurred when Warner Music removed its content from YouTube.

The same mechanisms can operate when the release dates are diversified across countries. In such cases, consumers in some countries might be able to download content from the internet before it hits the local official channels, which creates incentives for file-sharing. McKenzie & Walls (2013) showed that a release gap can contribute to "piracy" levels early in a film's theatrical life, and thus lower its revenues. Danaher & Waldfogel (2012) also found that the delayed legal availability of movies can harm their box office sales due to file-sharing.

The lack of an alternative is often cited as the main justification for file-sharing. This view is quite popular in the online debate, with many commenters citing poor distribution methods, prices, lack of access, or release lags as reasons why people share files rather than buying content (see, e.g., a compilation by Hart, 2012). However, previous research has found that the actual effects of the legal availability of content on the ethical judgements of consumers are not obvious, or vary between types of content. Krawczyk et al. (2017) showed that the offline unauthorised sharing of sports broadcasts is seen as more ethically acceptable if there is no free TV alternative available, but found the opposite relationship in the online context. Moreover, Krawczyk et al. (2015) showed that the availability of an easily accessible alternative does not influence people's ethical judgements of the unauthorised acquisition of a TV show. Finally, Veitch & Constantiou (2012) found that survey respondents were more likely to report watching films in the cinema and purchasing or streaming music (rather than downloading it) if the content was legally available in their preferred way. However, they found no effect of the legal availability of the content on the likelihood of purchasing movies.

The lack of an alternative works both ways, with more consumers switching to authorised channels when file-sharing is not available or becomes less attractive. Some studies have examined the effects of the implementation of stricter copyright laws regarding online infringements. Adermon & Liang (2014) showed that when Sweden implemented a harsher law regarding file-sharing, the internet traffic dropped and music sales went up. Danaher et al. (2014b) uncovered a similar effect for digital music sales after France adopted

[2] Admittedly, Hammond (2012) concluded that pre-release file-sharing does not reduce music sales, although his results appear to describe the effect of the magnitude of pre-release file-sharing, and not the existence of pre-release file-sharing *per se* (as noted by Danaher et al., 2014a).

the HADOPI graduated response law. Other authors considered the effects of a direct reduction in access to file-sharing sources. Danaher & Smith (2014) and Peukert et al. (2017) showed that the Megaupload shutdown had small but positive effects on major studios and blockbuster movies. Danaher et al. (2015) found that blocking The Pirate Bay website alone did not affect file-sharing levels, but that blocking 19 major piracy websites – and, later, 53 additional piracy websites (Danaher et al., 2016) – successfully decreased 'piracy' levels and increased legal consumption.

Whether those who move from one channel to another switch back or persist in their new behaviour has not been fully established. This is a very important issue, as the incentives for the switch are often only temporary, but the consequences of the switch might be far-reaching. For example, Danaher et al. (2010) hinted at a permanent shift of NBC viewers from iTunes to file-sharing networks, but – as they themselves admitted – they lacked a proper counterfactual. By contrast, Adermon & Liang (2014) showed that while a stricter law in Sweden decreased file-sharing and increased sales, the use of both channels soon reverted to the previous patterns.

3. The case of the Game of Thrones leak and methodology outline

GoT is a highly popular, award-winning TV show produced by HBO. It follows a serial structure – i.e., each episode is a direct follow-up to the previous one, and each episode typically ends with a cliff-hanger. At the time of this writing, the show has been nominated for over 460 awards and has won over 270, including a Golden Globe [as of April 2018]. It is listed as the most popular TV show currently and the fourth most highly rated TV show ever[3] on the Internet Movie Database (IMDb) [as of April 2018]. From its start in 2011 through 2016, one 10-episode season of the show was released each year, with weekly episodes spanning April-June[4]. The seventh season was released in July 2017, and comprised seven episodes, while the eighth and final season is scheduled for release in 2019. The premiere of each episode of the first season attracted an average of 2.5 million U.S. viewers. Since then, the audience has grown to an average of 7.7 million U.S. viewers for the episode premieres of season 6, and of 10.3 million viewers for the episode premieres of season 7 (see section 4 for a description of the data sources).

On 11 April 2015, four episodes of the fifth season were leaked online, just a day before the official premiere of the season. The video files were not of the highest quality, though Figure 1 shows that they attracted significant attention on the web. The fifth season also happened to be the first in the show's history not to record a meaningful increase in the TV audience, while the sixth and seventh seasons continued the trend observed for seasons 1-4. This pattern coincided with a small drop in viewers' ratings of the show for the initial episodes of season 5. The drop alone does not, however, explain the change in the viewership of GoT (see Figure 2).

The results of a Zivot-Andrews test for structural breaks confirm that there was a break in the viewership of GoT at the second episode of season 5. While the test cannot explain the effects of the leak on TV viewership levels, its findings suggest that the potential effects on viewership were not immediate. This could be because the leak occurred only a day before the airing of the first episode. Thus, it is possible that the viewers did not learn of the leak in time to take advantage of it, or that they decided to wait one day to watch the

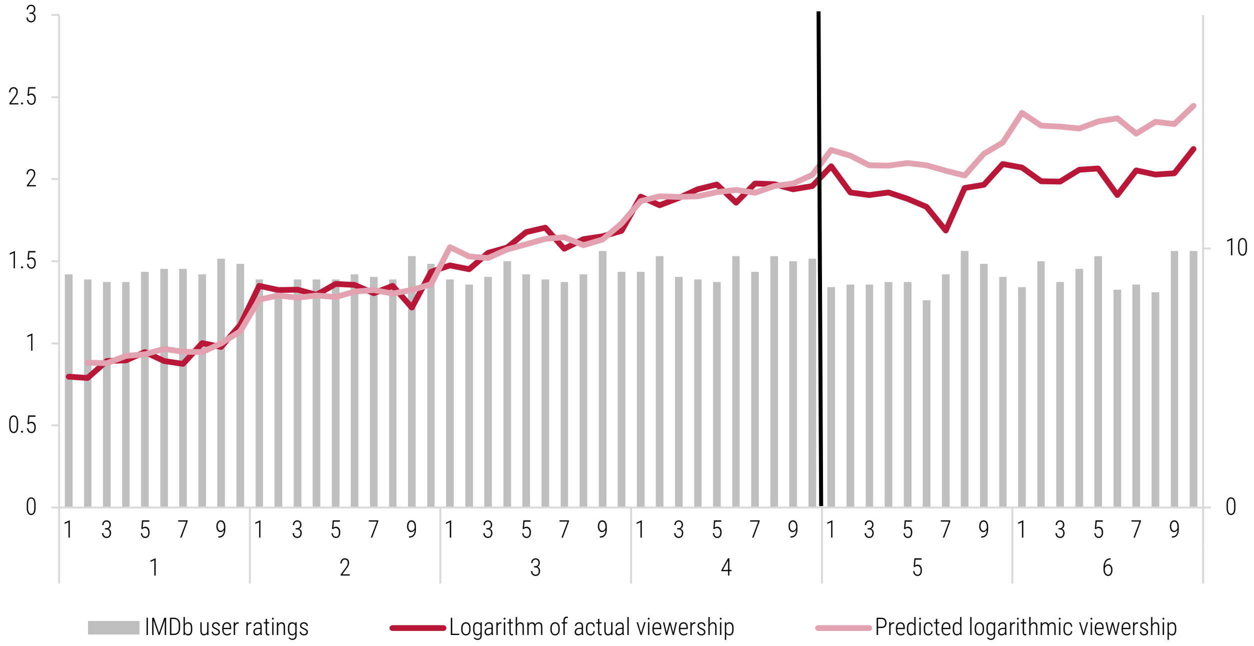
[3] It was outperformed only by shows of different formats: two documentaries (Planet Earth I & II) and a TV mini-series (Band of Brothers) from 2001.

[4] First episode of season 3 was aired on 31 March.

episode in high quality. It appears that the incentives for the unauthorised viewing of the episodes 2-4 were considerably larger.

The aim of this study is to measure whether the leak of GoT episodes contributed to the drop in the viewership levels of GoT, and whether it had spillover effects on the viewership levels of other TV shows in that period. To address these issues, I have collected data from several online sources that I use to construct a database on TV shows that were aired around the time of the GoT leak. I use a difference-in-differences approach to analyse whether the shift in GoT viewership levels was indeed characteristic of GoT viewership patterns, and whether the shift was temporary or persisted for more than four episodes. I also identify the TV shows that may share an audience with GoT, and run a difference-in-differences analysis to assess whether these shows experienced unusual viewership patterns after the leak of the GoT episodes. Finally, I corroborate these results with an analysis of the Google search popularity of phrases related to the TV shows in the sample and unauthorised sources. The following sections describe these processes in detail.

Figure 2. Game of Thrones ratings, viewership, and predicted viewership



Note: The horizontal axis shows episode numbers (1-10; row 1) in each of the seasons (1-6; row 2). The vertical line marks the leak, which occurred just before the first episode of season 5. The prediction is based on an OLS model on the sample of seasons 1-4, with the logarithm of viewership as an explaining variable; and the previous episode's viewership, the previous episode's ratings, the season number, and the dummies for season premieres and finales as explaining variables. The seventh season had the highest viewership numbers of all the seasons, but is not included here because it had a different number of episodes (seven) and a very different (summer) air date schedule (July-August).

Source: Own elaboration and calculations on IMDb and Nielsen (as reported on Wikipedia) data. The IMDb ratings come from IMDb public use files, downloaded on 23 June 2017.

4. Data

4.1. TV show data

I use a unique dataset of information collected from several websites. The purpose of gathering the data was to describe both GoT and a set of TV shows that were being aired at the time of the GoT leak. The sources of the data include: Nielsen (as reported by TV By The Numbers, and aggregated by Wikipedia), IMDb, Wikipedia, Google Trends, and several websites with movie recommendation algorithms (e.g., Jinni, TasteDive). For some of the sources, I have made use of the available data dumps (Wikipedia) or personal-use files (IMDb, after having ascertained that the manner in which I intended to use the data was permitted). Importantly, I only used automated data mining tools for the previously downloaded data files in order to avoid generating unwarranted traffic to the websites. The listed sources have previously been used for research by other authors; see, e.g., Sood and Drèze (2006) for IMDb, Cadario (2015) for Nielsen/TV By The Numbers, Choi and Varian (2012) for Google Trends or Lee et al. (2014) for coverage of the movie recommendation systems.

As a first step, I searched through the TV ratings uploaded daily on the TV By The Numbers website for the one week before and the one week after the airing of the first episode of GoT season 5. I identified all of the TV shows aired in these periods. However, I decided not to include cartoons (e.g., The Simpsons) as they seemed non-comparable to GoT and other series. I also cleared the data of any programmes and shows without a plot, such as sports broadcasts and late-night shows. I then removed movies as well as TV shows with episodes with a runtime shorter than 40 minutes. I chose to exclude this last group of shows because shorter episodes require less time to consume. Moreover, unlike shows with a longer runtime, these shows are typically comedies with no larger plot requiring the viewer to follow each episode. Finally, I cleared the data of TV series with episodes that were not aired both before and after the leak of the GoT episodes. My final dataset contains information on 52 TV shows of varying popularity. For each of the shows (or episodes, when appropriate), I then collected the data from the other previously mentioned data sources (see Table 1 for the list of sources and data).

Table 1. Sources of the data for the TV shows in the sample

Source	Data (episode-level when applicable)
Wikipedia	Genres, Typical runtime, Number of seasons, Numbers of episodes, Episode names, Directors, Screenwriters, Episode air dates, Number of first-day viewers in US*.
IMDb	TV Rating (i.e., MA/14/PG), Typical runtime, Genres, User rating, Number of votes on ratings
Google Trends	Google search popularity of phrases: "*show-name torrent*", "*show-name download*", "*show-name watch online*"
Jinni, IMDb, Netflix, TasteDive, Trakt.tv	Show recommendations for GoT fans

*Note: *Wikipedia pages contain compilations of ratings from Nielsen (as reported in TV By The Numbers website posts). These compilations in turn report partial data (for most popular TV shows) from the Nielsen ratings.*

Source: Own elaboration.

Some of the shows in the sample date back as far as 1999. Of the shows in the sample, 13 had started airing by 2010, and 42 had started airing by 2014. By 2016, the number of the shows in the sample that were still on the air had fallen to 39. Of the shows recommended to GoT fans [hence referred to as the "GoT-like" shows], only four had started by 2012, but 10 had started by 2013 and 12 had started by 2014.

Table 2. Number of shows in the sample that had seasons in specific years

	1999-2002	2003-2004	2005-2006	2007-2008	2009	2010	2011	2012	2013	2014	2015	2016
All shows	1	2	6	7	10	13	17	23	32	42	52	39
GoT-like shows	0	0	1	1	2	2	3	4	10	12	12	11

Note: Game of Thrones is not included. It has been aired since 2011 and will finish in 2019.

Source: Own elaboration.

Table 3 contains summary statistics on the variables from the final dataset. The data contain information on all seasons of the shows in the sample. However, for the analysis I drop the seasons that were aired in 2017 because not all of the shows had finished their seasons at the time of data collection. To avoid bias towards shows that had aired their seasons earlier, I decided not to include the 2017 seasons for any of the shows. Table 3 reflects this choice by presenting the statistics and numbers without the 2017 seasons. Without the seasons aired in 2017, the dataset comprises 52 shows and 4,624 episodes.

Table 3. Summary statistics of the sample and sample description, without the seasons aired in 2017

Variable name	Mean	Min	Max	Std. Dev.	Note
			Show-level		
Genre:					Comprises:
- *Action*	35%	0	1	-	Action
- *Comedy*	15%	0	1	-	Comedy, Sitcom
- *Crime*	34%	0	1	-	Crime, Thriller, Mystery
- *Drama*	51%	0	1	-	Drama
- *Romance*	15%	0	1	-	Romance
- *Supernatural*	38%	0	1	-	Sci-fi, Fantasy, Supernatural, Horror
Typical runtime	45.4 min	40 min	60 min	5.7	As listed by IMDb
Number of seasons	4.5	1	17	3.5	Aired (or being aired) at the time of data collection
Episodes per season	18.3	10	24	5.1	The numbers are for the show averages (e.g., the mean of the show averages)
Episodes per show	88.9	10	389	83.2	-
First day viewers in US (in millions)	5.9	0.76	17.1	3.8	The numbers are for the show averages (e.g., the mean of the show averages)
TV rating:					
- *PG*	15.4%	0	1	-	Parental Guidance Suggested
- *MA*	5.8%	0	1	-	Mature Audience Only
- *G*	1.9%	0	1	-	General Audience
- *14*	76.9%	0	1	-	Parents Strongly Cautioned
IMDb user rating	7.8	5.4	9.5	0.6	Average of episode ratings
Network					
- *CBS*	21.2%	-	-	-	Some of the shows changed networks at some point. For this reason the shares do not sum up to 100%.
- *ABC*	19.2%	-	-	-	
- *NBC*	15.4%	-	-	-	
- *CW*	15.4%	-	-	-	
- *Fox*	7.7%	-	-	-	
- *Other*	28.8%	-	-	-	
			Season-level		
Month of first episode:					
- *September*	56%	-	-	-	-
- *October*	18%	-	-	-	
- *March*	9%	-	-	-	
- *Other*	17%	-	-	-	
Month of last episode:					
- *May*	78%	-	-	-	-
- *April*	8%	-	-	-	
- *June*	7%	-	-	-	
- *Other*	7%	-	-	-	
			Episode-level		
First-day viewers in US	8.0	0.3	37.9	5.0	-
IMDb user rating	8.2	3	9.9	0.6	-

Source: Own elaboration and calculations based on IMDb and Wikipedia data.

The genre categories were first taken from the Wikipedia and IMDb websites. The Wikipedia classification contained 44 distinct genres and the IMDb classification contained 14 distinct genres, with the two groups overlapping in some cases. In principle, the Wikipedia genres were much more specific (e.g., *neowestern*, *occult*) or lacked consistency across the TV show webpages (e.g., *comedydrama/dramedy* or *soap opera/telenovela*). I aggregated the genres into six broader non-exclusive categories, with each representing a significant part of the sample. I left out the genres that did not fit into the broader categories (e.g., *neowestern*). It should be noted that since each show could have more than one genre in IMDb and Wikipedia, all of the shows had some genres after this procedure. The detailed genres are listed in Table 3.

The IMDb data contain user ratings for each episode of each of the shows. However, one concern is that the IMDb user database might be affected by internet file-sharing, as viewers of the unauthorized copies might be less willing to share their views and opinions on IMDb. To check the reliability of the data, I also collected the GoT episode ratings from Rotten Tomatoes, which is a review-aggregating website. Thus, the review-based scores are not subject to changes among the users. However, these scores are limited to the shows that receive sufficiently large reviewer coverage. Only two of the shows in the sample had per-episode ratings on Rotten Tomatoes, with one of them having ratings only for one season. For Game of Thrones, the correlation between the IMDb user ratings and the Rotten Tomatoes reviewer ratings was 77%. Importantly, for seasons 5 and 6 the correlation was 86% and 90%, respectively, which suggests that the correlation did not break at the time of the leak. I conclude that the IMDb data are consistent measures of the quality of the episodes.

For the show recommendations, I have collected the recommendations from several websites for TV and film enthusiasts (see Table 4 for the list of the recommendation websites). Unless otherwise noted, the recommendations were collected in June 2017. In total, 12 shows in the sample were marked as recommended for GoT fans.

Table 4. TV shows recommended for fans of Game of Thrones

Source	Shows in the sample	Source description
Jinni	Vikings	Jinni was a publicly available website with a recommendation engine. Since 2015, it has been a paid service for larger companies in need of movie recommendation engines. However, archived copies of the recommendations for GoT viewers can be found dating back to March and April 2015 (the approximate time of the leak).[5] Only one recommendation coincides with the shows in my sample.
IMDb	Arrow, Flash, Gotham, Supernatural, Vikings	IMDb lists 12 non-personalised recommendations per show. I browsed the archived copies of the GoT page, between 9 April and 15 June 2015. Of all of the recommendations posted over this period, six coincided with the shows in my sample.
Netflix	Arrow, Blacklist, Flash, Gotham, Grimm, Originals, Reign, Vampire Diaries, Vikings	Netflix is a video streaming service with a marketing focus on audience discovery. It does not contain GoT in its catalogue, but lists recommendations of similar TV shows for those who look for it in the Netflix search engine. Of the 42 Netflix recommendations, nine coincided with the shows in my sample.
TasteDive	Vikings, Gotham, Arrow, Americans	TasteDive (former TasteKid) presents numerous recommendations based on a specified TV series. Of the 50 first recommendations for those who like GoT, four coincided with the shows in my sample.
Trakt.tv	Agents of S.H.I.E.L.D, Flash	Trakt.tv offers six recommendations for a specified show. Two recommendations for those who like Game of Thrones coincide with shows in my sample.

Source: Own elaboration.

4.2. File-sharing popularity data

I collected weekly US web popularity data for phrases related to file-sharing of the seasons in the sample, which were aired around the time of the GoT leak. I used the Google Trends service to collect the data,

[5] Wayback Machine Archive: http://web.archive.org/web/20150409011637/http://www.jinni.com/tv/game-of-thrones/

starting from the time of the first episode until one week after the last episode (i.e., capturing the popularity of the search phrases one week after the last episode's initial airing).

Internet search data are viable proxies for file-sharing traffic, and especially for that of the new users of unauthorised sources. According to MUSO (2017), internet searches account for approximately 35% of traffic leading to file-sharing sites, after direct traffic (42%) and before referrals (20%). However, it is likely that among the new downloaders who are the focus of this study, the share who accessed the content through internet searches was higher than the share who accessed the content through direct searches. Indeed, according to a report by Millward Brown Digital (2013), search engines play an important role in accessing unauthorised online content. In 2010-2012, the first-timers (people who were accessing the unauthorised sources for the first time) were almost twice as likely as the repeat viewers to have found content by using search engines. At the same time, the first-timers were half as likely as repeat viewers to have used direct entry to websites with unauthorised content. Finally, Sivan et al. (2014) showed that the search engine results influence the choices users make about whether to access authorised or unauthorised sources, which proves that search traffic can be translated into further consumption behaviour.

Google Trends constitute an accurate measure of internet searches. The Google Trends popularity index measures the popularity of specific search phrases in Google. Google is the most popular search engine in the USA and globally (see, e.g., Statista, comScore, or NetMarketShare[6]). The Millward Brown Digital (2013) report shows that approximately 82% of people who conducted a search before viewing infringing content conducted the search via Google. Hence, Google Trends data for internet searches of unauthorised content are a viable proxy for the file-sharing traffic.

I collected the Google Trends data for each show separately, and separately for three file-sharing-related phrases. The phrases were: "torrent", "watch online", and "download". Regular downloaders might use more refined search phrases that are more effective for finding unauthorised sources, specific episodes, specific formats, or certain levels of quality (see, e.g., WareZ naming guidelines). However, as I am mainly interested in the viewers who have only recently turned to unauthorised sources, I have chosen to focus on general and simple search phrases. My choice of search phrases is backed by the Millward Brown Digital (2013) report, which assigns internet film/TV-related search phrases to one of three categories: "piracy domain", which indicates the target domain; "title", which is the title of the searched film/TV show; and "generic", which are phrases related to watching and movies, like "watch TV online", and "free movies". According to the report, approximately 23% of the searches include both a generic word and the title. However, the first-timers were less likely to have used domain names and more likely to have searched for titles and generic terms. I use the combination of titles and generic terms because it appears that both types of phrases were popular among new users of unauthorised sources, and because they allow me to identify the searches on a TV show level.

The Google Trends popularity index measures how popular specific search phrases are, and allows me to filter the results by country of origin and time. Importantly, the Google Trends result for a search phrase like "Game of Thrones download" takes into account search phrases that contain additional words like "s05e05", "The Pirate Bay", or "HD". The data are standardised, with the highest value in the range scaled to 100 and the other values scaled accordingly. Thus, the resulting data for each of the search phrases contain values ranging from zero to 100. For some of the shows, the Google Trends app reports daily data instead of weekly

---

[6] Statista: https://www.statista.com/statistics/267161/market-share-of-search-engines-in-the-united-states/; comScore: http://www.comscore.com/Insights/Rankings/comScore-Releases-February-2016-US-Desktop-Search-Engine-Rankings; NetMarketShare: https://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=1. All accessed on 2017-09-13.

intervals. I aggregated the daily numbers to weekly numbers by summing the daily values, and then rescaling all of the numbers again. The data collected in this manner do not allow for meaningful comparisons of popularity across shows, but they do allow me to study trends in the Google search popularity, as well as potential breaks in these trends.

The Google Trends data allow me to study the effect of the GoT leak on the internet popularity of GoT-like TV shows. The data for GoT suffer from a lack of traffic directly before the leak, since the preceding episodes were first aired a year earlier. Moreover, the web popularity volume tends to change between seasons, so it is impossible to compare the level of popularity after the leak with the level of popularity of the previous season aired before the leak. For these reasons, I restrict the Google popularity data to only the seasons aired around the time of the leak and remove GoT from the sample. The data are therefore used solely to compare the popularity of GoT-like shows to other TV shows.

5. Results

5.1. TV viewership

I run a difference-in-differences regression on GoT and other TV shows. I analyse the trends in viewership of a sample of TV shows, checking whether there were changes following the leak, and whether these changes were different for GoT, for shows that share an audience with GoT, and for other shows. I use episode numbers as a measure of time. Though there are (typically yearly) time gaps between each show's seasons, this approach has been previously applied by, e.g., Cadario (2015). Moreover, I cannot restrict the sample to the seasons aired during the time of the leak, as the leak occurred before the airing of the GoT fifth season (i.e., there is no GoT season with episodes aired both before and after the leak). Finally, I mark the first episode after the leak as episode 1 for all shows.

The base model can be described as follows:

$$
\begin{aligned}
&\left.\begin{aligned}
\log(Viewers) = \beta_{\mathrm{r}} * \log(Rating_{t-1}) + \beta_{\mathrm{v}} * \log(Viewers_{t-1}) + \\
+ \beta_{\mathrm{SP}} * SeasonPremiere + \beta_{\mathrm{SF}} * SeasonFinale + \beta_{\mathrm{SB}} * SuperBowl + \\
+ \beta_{\mathrm{Ep}} * EpNumber + \beta_{\mathrm{GTEp}} * GoTEpNumber + \beta_{\mathrm{GTLEp}} * GoTLikeEpNumber +
\end{aligned}\right\} \text{Control variables} \\
&\left.\begin{aligned}
+ \beta_{\mathrm{AL}} * AfterLeak + \beta_{\mathrm{GTAL}} * GoTAfterLeak + \beta_{\mathrm{GTLAL}} * GoTLikeAfterLeak +
\end{aligned}\right\} \text{Shifts after leak} \\
&\left.\begin{aligned}
+ \beta_{\mathrm{EpAL}} * EpNumberAfterLeak + \beta_{\mathrm{GTEpAL}} * GoTEpNumberAfterLeak + \\
+ \beta_{\mathrm{GTLEpAL}} * GoTLikeEpNumberAfterLeak
\end{aligned}\right\} \text{Trend breaks after leak}
\end{aligned}
$$

, where:

$\log(Viewers_t)$ – is the logarithm of the number of US viewers on the day of the premiere of episode *t*.

$\log(Rating_{t-1})$ – is the logarithm of the IMDb rating of the previous episode.

$\log(Viewers_{t-1})$ – is the logarithm of the number of US viewers on the day of the premiere of the previous episode.

$SeasonPremiere$ – is a dummy indicating the first episode in a season.

$SeasonFinale$ – is a dummy indicating the last episode in a season.

$SuperBowl$ – is a dummy indicating that the Super Bowl took place on the same day.

$EpNumber$ – is the episode number within the show, interpreted as a general trend.

$GoTEpNumber$ – is the episode number within the show for GoT, interpreted as a GoT-specific trend.

$GoTLikeEpNumber$ – is the episode number within the show for GoT-like TV shows, interpreted as a GoT-like shows-specific trend.

$AfterLeak$ – is a dummy for episodes aired after the leak of GoT.

$GoTAfterLeak$ – is a dummy for Game of Thrones episodes after the leak.

$GoTLikeAfterLeak$ – is a dummy for GoT-like episodes after the leak.

$EpNumberAfterLeak$ – is the episode number within the show after the GoT leak, interpreted as the change in the general trend.

$GoTEpNumberAfterLeak$ – is the episode number within the GoT show after the GoT leak, interpreted as the GoT-specific change in the general trend.

$GoTLikeEpNumberAfterLeak$ – is the episode number within the GoT-like shows after the GoT leak, interpreted as the GoT-like shows-specific change in the general trend.

The key control variables are $GoTAfterLeak$, $GoTLikeAfterLeak$ (for shifts in the viewership of GoT and GoT-like shows, respectively, after the leak) and $GoTEpNumberAfterLeak$, $GoTLikeEpNumberAfterLeak$ (for changes in the trends in the viewership of GoT and GoT-like shows, respectively, after the leak). Specifically, if GoT or the GoT-like shows displayed an unusual drop in their viewership at the time of season five, we would expect the β_{GTAL} or β_{GTLAL}, respectively, to be significant and negative. Moreover, the β_{GTEpAL} will reflect the GoT-specific change in the trend after the leak, and β_{GTLEpAL} will reflect an analogous change for the GoT-like shows. If the drop was not immediate and occurred over the period of the leaked episodes, we would expect β_{GTEpAL} and β_{GTLEpAL} to be negative. On the other hand, the coefficients may also reflect the viewership returning to its initial levels. The remaining variables control for the changes occurring globally for all TV shows aired at the time of the leak. A significant β_{AL} would mean that all of the shows experienced changes in their viewership trends after the leak, though not necessarily because of it. A significant β_{EpAL} would mean that all of the shows experienced changes in their viewership trends after the leak, though not necessarily because of it. The lagged viewership and ratings variables allow me to track most of the dynamics in a show's viewership. Finally, the control variables: $SeasonPremiere$, $SeasonFinale$, and $SuperBowl$ have been shown to explain much of the TV viewership of the episodes they apply to (Cadario, 2015).

Table 5 reports the results of the model estimation using a panel OLS regression with fixed effects. Columns (1) and (3) apply the base model described above, while columns (2) and (4) include show-specific trends in both linear and squared forms, instead of the aggregate trends from the base model. Columns (3) and (4) include calculations on a sample reduced to episodes aired in 2013 and later. This reduction makes the panel more balanced, and 2013 is the first year in the data that had seasons of the majority of the TV shows in the sample, including most of the GoT-like shows (10 out of 12, compared to only four in 2012; see Table 2).

All of the models are well fitted and explain most of the variance in viewership. The between-show R-squared is larger than 80% (except in the model with both sample reduction and trend variables), and the within-show R-squared is at the levels of 83% and 87% for the whole sample specifications and 76% and 82% for the reduced period sample. As expected, the lagged viewership dummies for season premieres and finales, as well as the dummy for Super Bowl, are found to be highly significant and positive.

Table 5. Difference-in-differences estimation of the effect of the leak on viewership. Panel OLS regressions with fixed effects

	Whole sample				Only years since 2013			
	(1) Base model		(2) Show-specific trends		(3) Base model		(4) Show-specific trends	
Log(Rating$_{t-1}$)	0.100*	(0.055)	0.084	(0.063)	0.023	(0.058)	0.085**	(0.037)
Log(Viewers$_{t-1}$)	0.803***	(0.034)	0.494***	(0.038)	0.723***	(0.025)	0.362***	(0.031)
After Leak	-0.013	(0.010)	0.005	(0.015)	-0.017**	(0.007)	-0.015	(0.012)
GoT After Leak	-0.048***	(0.012)	-0.098***	(0.015)	-0.042***	(0.011)	-0.085***	(0.014)
GoT-Like After Leak	-0.025*	(0.014)	-0.052**	(0.023)	-0.021	(0.013)	-0.026	(0.023)
Ep. No. After Leak	-0.001	(0.001)	0.001	(0.001)	-0.001	(0.001)	0.001	(0.001)
GoT Ep. No. After Leak	-0.000	(0.001)	-0.004***	(0.001)	0.000	(0.001)	-0.011***	(0.002)
GoT-like Ep. No. After Leak	0.001	(0.001)	-0.003	(0.002)	0.001	(0.001)	0.001	(0.004)
Ep. No.	-0.000*	(0.000)	-		-0.001**	(0.000)	-	
GoT Ep. No.	0.006***	(0.001)			0.006***	(0.001)		
GoT-like Ep. No.	-0.000	(0.000)			-0.001	(0.001)		
Season premiere	0.067***	(0.014)	0.074***	(0.015)	0.072***	(0.018)	0.080***	(0.018)
Season finale	0.059***	(0.009)	0.051***	(0.009)	0.059***	(0.010)	0.054***	(0.010)
Super Bowl	0.854***	(0.104)	0.865***	(0.122)	0.865***	(0.106)	0.870***	(0.125)
Show-specific trends (linear and squared)	No		Yes		No		Yes	
Episodes	4,538		4,538		2,915		2,915	
Shows	52		52		52		52	
R-sq: within	0.833		0.865		0.760		0.817	
R-sq: between	0.998		0.858		0.991		0.154	

*Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at the show level in parenthesis. The show-specific trends are in both the linear and squared forms.*

Source: Own calculations based on IMDb, Wikipedia data and information compiled in Table 4.

For all of the specifications, the coefficient on GoT after the leak is negative and significant. The average estimated effect ranges from 4% to 10%. Moreover, there is no indication of a positive change in the trend in GoT viewership after the leak. This finding suggests that the viewership did not recover from the shift – i.e., that the viewers did not return to the authorised channel.

Moreover, in all of the specifications the coefficient on GoT-like shows after the leak is negative, although it is insignificant in the reduced period sample. The estimated effects range from 2% to 6%. The trends in the viewership of the GoT-like shows did not change after the leak, which again suggests that there was no recovery in the viewership.

The pre-leak trend in the viewership of GoT differs significantly from those of other TV shows. Indeed, GoT is a unique case in the sample, as its viewership across seasons was non-declining. This poses a risk of violating the common-trends assumption of the difference-in-differences model. However, the model explains more than 80% of the within-show variation, which might be sufficient to account for this bias.

As a robustness check of the effect on the GoT-like TV shows, I replicate the regression without GoT in the sample (see regression (1) in Table 6). Excluding GoT allows me to make further adjustments to the model. I only include the seasons that were aired around the time of the GoT leak. This eliminates the risk of shifts in the viewership between seasons, as well as of other factors, like changes in the TV network or in the broadcast times, from season to season.

The regression confirms that the GoT-like TV shows experienced a drop in viewership that was not characteristic of other TV shows in this period. It also shows that the viewership might have recovered thereafter, albeit slowly. This finding is based on a change in the trend in the viewership of GoT-like TV shows towards a more positive one, following the leak. According to the estimates, the viewership returned to its pre-leak levels in 5-6 episodes after the leak. Note, however, that this finding reflects the viewership of episodes that were not available through unauthorized channels prior to their premieres. Thus, the 5-6 episodes might be translated to a period of 5-6 weeks or 1-2 months.

Table 6. Difference-in-differences estimation of the effect of the leak on the viewership of GoT-like shows. Panel OLS regressions with fixed effects

	(1) Base model		Placebo leak year earlier (2014)				Placebo leak year later (2016)			
			(2) exactly year earlier		(3) day before the GoT season started		(2) exactly year later		(3) day before the GoT season started	
Log(Rating$_{t-1}$)	0.009	(0.055)	0.149	(0.095)	0.148	(0.092)	-0.001	(0.072)	0.002	(0.074)
Log(Viewers$_{t-1}$)	0.382***	(0.045)	0.313***	(0.064)	0.300***	(0.065)	0.416***	(0.066)	0.442***	(0.063)
After Leak	0.036**	(0.015)	-0.015	(0.014)	0.004	(0.014)	-0.002	(0.014)	0.004	(0.016)
GoT-Like After Leak	-0.104***	(0.035)	-0.017	(0.035)	-0.067**	(0.031)	-0.035	(0.026)	-0.023	(0.029)
Ep. No. After Leak	-0.019***	(0.004)	-0.008	(0.007)	-0.011**	(0.005)	-0.004	(0.005)	-0.007	(0.006)
GoT-like Ep. No. After Leak	0.021***	(0.007)	0.003	(0.011)	0.013	(0.008)	0.002	(0.006)	-0.000	(0.008)
Ep. No.	-0.004***	(0.001)	-0.003**	(0.001)	-0.003**	(0.001)	-0.002**	(0.001)	-0.003**	(0.001)
GoT-like Ep. No.	0.001	(0.002)	-0.005*	(0.002)	-0.003	(0.002)	-0.000	(0.002)	-0.001	(0.002)
Season premiere	0.099***	(0.027)	0.117***	(0.036)	0.119***	(0.034)	0.047**	(0.017)	0.040**	(0.019)
Season finale	0.077***	(0.020)	0.077***	(0.017)	0.075***	(0.016)	0.061***	(0.015)	0.069***	(0.015)
Super Bowl	1.056***	(0.006)	-	-	-	-	-	-	-	-
Episodes	942		598		611		682		653	
Shows	51		29		30		32		30	
GoT-like shows	12		10		10		11		11	
R-sq: within	0.538		0.403		0.408		0.367		0.382	
R-sq: between	0.272		0.229		0.171		0.623		0.625	

*Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at show-level in parenthesis.*

Source: Own calculations based on IMDb, Wikipedia data and information compiled in Table 4.

I also conducted four placebo tests to further support my results (see regressions (2)-(5) in Table 6). To do so, I replicate the base model, but set a hypothetical leak event at four different dates: a day before the first episode of a GoT season in 2014, a day before the first episode of a GoT season in 2016, 11 April 2014, and 11 April 2016. The dates are set in order to check both whether the previously observed effects are characteristic of the time of the leak (11 April), or whether they were related to the time of a GoT season premiere. The latter would be the case if, for example, some people had only had enough time to watch one TV show in a week, and thus chose to switch from other shows to GoT when it premiered.

The samples for the placebo tests are somewhat smaller, as not all of the shows in the sample had episodes aired before and after the hypothetical leaks. For the placebo leaks on 11 April 2014, 29 shows were included in the regressions (incl. 10 GoT-like shows). For the placebo leaks on 11 April 2016, 32 shows were included in the regressions (incl. 11 GoT-like shows). For the placebo leaks a day before seasons 4 or 6 of GoT, 30 shows were included in the regressions (incl. 10 GoT-like shows in 2014 and 11 GoT-like shows in 2016).

In only one of the four cases is the coefficient for the GoT-like shows after the leak negative – for the hypothetical leak a day before the premiere of the GoT season in 2014. Notably, the size of the coefficient is smaller and less significant than it is for the 2015 season. This outcome suggests that the leak event in 2015 was indeed unique, and at the very least contributed to the drop in viewership in this period.

5.2. Google trends for file-sharing

Google Trends data were collected for three search phrases per each show: "*show_name* download", "*show_name* torrent", and "*show_name* watch online". The data were collected for each show from the beginning to the end of the season during which the leak of GoT occurred. There were two exceptions to this rule: for the two-part seasons of the Finding Carter and the Mad Men shows, only the parts aired around the time of the leak were considered. Moreover, the data are available only when the searched phrases were popular enough. For this reason, the sample of TV shows for this analysis is somewhat smaller, ranging from 45 to 48 depending on the search phrase. In each specification, there are 12 GoT-like shows.

The popularity data were aggregated to weekly levels by averaging seven consecutive days and rescaling so that the most popular week has the value of 100. Each of the shows had observations with weeks after the leak (starting 12-19 April 2015) and before the leak. For each of the shows, the first week was the one during which the first episode of the season was aired, and the last week was the one when the last episode of the season was aired. The regressions presented in Table 7 follow closely the model for TV viewership in Table 6, but also include the lags of Google Trends popularity.

The coefficients on the GoT-like shows after the leak are insignificant in all of their specifications. However, for the phrases including "watch online", the trends changed to more positive ones for the GoT-like shows. This suggests that a growing number of people were searching for the GoT-like shows to watch online after the leak of GoT.

As in the viewership analysis, I run additional regressions with placebo leaks in analogous periods of 2014 and 2016. The sample sizes again become smaller. For the placebo leaks on 11 April 2014, 27 shows are included in the regressions (incl. 10 GoT-like). For the placebo leaks on 11 April 2016, 30 shows are included in the regressions (incl. 11 GoT-like). For the placebo leaks a day before seasons 4 and 6, 28 or 29 shows are included in the regressions (incl. 10 GoT-like shows before season 4 and 11 before season 6). In all cases, the third model ("watch online") has one more show in the sample.

Table 7. Difference-in-differences estimation of the effect of the leak on Google popularity of the GoT-like TV shows. Panel OLS regressions with fixed effects

	show_name + *"download"*		*show_name* + *"torrent"*		*show_name* + *"watch online"*	
Google Popularity(t-1)	0.316***	(0.047)	0.360***	(0.080)	0.379***	(0.064)
Rating(t-1)	0.260	(0.200)	0.273	(0.189)	0.149	(0.141)
Viewers(t-1)	-0.109	(0.108)	0.101	(0.074)	-0.077	(0.078)
After Leak	-0.002	(0.026)	-0.050	(0.035)	-0.061*	(0.032)
GoT-like After Leak	-0.048	(0.051)	0.053	(0.061)	-0.039	(0.045)
Ep. No. After Leak	-0.003	(0.009)	0.013	(0.009)	0.007	(0.010)
GoT-like Ep. No. After Leak	0.021	(0.016)	-0.001	(0.020)	0.028**	(0.012)
Ep. No.	-0.001	(0.002)	-0.000	(0.003)	0.002	(0.004)
GoT-like Ep. No.	0.003	(0.005)	0.004	(0.006)	0.003	(0.005)
Season finale	0.034	(0.043)	-0.016	(0.044)	-0.018	(0.043)
Super Bowl	0.252***	(0.017)	0.420***	(0.032)	1.003***	(0.044)
Episodes	797		833		855	
Shows	45		47		48	
R-sq: within	0.113		0.156		0.166	
R-sq: between	0.350		0.359		0.151	

*Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at show-level in parenthesis.*

Source: Own calculations based on Google Trends, IMDb, Wikipedia data and information compiled in Table 4.

The coefficients on shifts and trends in all placebo treatments are insignificant, except for two specifications with the placebo leak in 2014, in which the coefficients are significant at a 10% level but negative. This finding confirms that the positive change in the trend for phrases including words "watch online" was unique to the year 2015.

Table 8. Coefficients on the difference-in-differences variables from a placebo effect estimations from models analogous to those in Table 7

	show_name + "download"		*show_name + "torrent"*		*show_name + "watch online"*	
Day before GoT season 4 (GoT-like After Leak)	-0.053	(0.059)	0.022	(0.065)	-0.070	(0.044)
Day before GoT season 4 (GoT-like Ep. No. After Leak)	-0.004	(0.016)	-0.039*	(0.022)	-0.013	(0.013)
Day before GoT season 6 (GoT-like After Leak)	0.061	(0.042)	-0.020	(0.047)	0.045	(0.061)
Day before GoT season 6 (GoT-like Ep. No. After Leak)	-0.015	(0.014)	0.012	(0.011)	0.014	(0.023)
11 April 2014 (GoT-like After Leak)	0.008	(0.050)	0.072	(0.048)	0.013	(0.044)
11 April 2014 (GoT-like Ep. No. After Leak)	-0.010	(0.018)	-0.050*	(0.028)	-0.024	(0.017)
11 April 2016 (GoT-like After Leak)	-0.017	(0.056)	-0.000	(0.061)	-0.046	(0.060)
11 April 2016 (GoT-like Ep. No. After Leak)	-0.006	(0.015)	0.002	(0.014)	0.004	(0.014)

*Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at the show level. The coefficients are for the variables in parentheses, in specifications analogous to those in Table 7. Full regression results for the placebo leaks are available upon request.*

Source: Own calculations based on Google Trends, IMDb, Wikipedia data, and information compiled in Table 4.

6. Discussion

The results of my analyses suggest that the leak of GoT episodes caused some of the show's viewers to shift to unauthorised channels. To demonstrate that this is the case, I used a unique dataset containing information on the TV viewership of a sample of TV shows, as well as on the popularity of these shows in Google search traffic. Importantly, a negative shift in viewership was found, evidenced both by a drop in the viewership of GoT and by a decline in the viewership of TV shows that share an audience with GoT. I also observed an increase in the search traffic for unauthorised sources of shows that share an audience with GoT. While the viewership levels of the GoT-like TV shows slowly returned to their previous levels, it is likely that some viewers continued watching some of these TV shows through unauthorised channels. It is possible that viewers used their newly acquired knowledge about unauthorised sources to start watching shows that they had not watched before through unauthorised channels. As the increased search traffic pertained only to the phrase "watch online", it appears that the leak did not have adverse effects on non-video types of content. However, it is unclear whether a similar pattern would emerge if the leaked content was a video game or music.

Overall, the estimated decrease in viewership was not very large. For Game of Thrones, viewership likely declined between 6% and 10%. This finding cannot be extended to other such events. For one thing, Game of Thrones is in no way representative of TV shows. In addition, GoT is scheduled during a specific time of year and follows a different format than many other shows. The effects of a given leak might also depend on how many episodes were leaked and how early. In the case of GoT, the leaked episodes represented almost half of the season, with the fourth episode having been leaked more than three weeks before its official premiere. The effects of this leak would probably have been smaller if fewer episodes had been leaked, and if the episodes had been leaked closer to their official premieres. Moreover, the effects might depend on the network the show is aired on. GoT is aired on HBO, which is a subscription-based TV network. One could argue that since TV viewers of GoT are already paying for the subscription, they might be less willing to switch sources. The effects might therefore have been different if the leaked episodes had been of a show on another network, and might have been even more different if the episodes had belonged to an online distributor and producer.[7] Finally, it has previously been shown that sci-fi shows and movies are more prone to file-sharing than other genres (see, e.g., Danaher & Waldfogel, 2012), a pattern that might extend to fantasy shows like GoT. This tendency could also mean that the number of first-time switchers among these sci-fi/fantasy viewers would be relatively low.

The decline in the viewership of GoT-like shows ranged from around 2% to 6% in the joint, whole period model to around 10% in the model with only one season per show. Moreover, the latter model also showed that viewership levels recovered slowly after the leak. Again, these estimates are not very large, and are similar in magnitude to the effects on GoT. However, it is important to remember that the audiences of the GoT-like shows are much smaller. Therefore, a small shift in the viewership of GoT could have an observable impact on the viewership of GoT-like shows, even if only a small share of the GoT viewers were also viewers of other GoT-like shows. Moreover, my data were on the first-day viewers only. It is therefore likely that my estimates cover only the effects on the most eager part of the audience, while those who were willing to wait might not have been affected by the leak.

This study has a few limitations, which I now address. First, before the start of the fifth season of GoT, HBO launched its video-on-demand service HBO NOW. The previous HBO VOD service required viewers to be TV subscribers. HBO NOW lifted that requirement, and was the first attempt by HBO at targeting audiences who did not want to subscribe to a TV service – the so-called “cord-cutters” or “cord-nevers”. The launch of HBO NOW could have affected the results if some of the previous TV subscribers had switched to HBO NOW instead of watching GoT on TV. However, according to the New York Times (Steel, 2016), HBO NOW had attracted around 800,000 subscribers by February 2016. Even if some of these subscribers were previous TV viewers of GoT, the number is too low to explain the decrease in GoT viewership following the leak.

Another potential concern is that the difference-in-differences analysis is biased by the lack of trend commonality before the leak. Indeed, GoT is the only show in the sample that displays a continuous increase in viewership across seasons (with the exception of season 5). I try to control for this exceptional pattern by including trend controls in some of the model specifications and adding a lagged dependent variable as a control variable. My model also manages to explain a large majority of the variance within each show in the sample. Still, it is possible that the results are biased due to the lack of common trends. However, the shows identified as sharing an audience with GoT did not differ significantly in their trends relative to other shows in the sample (as evidenced by the regressions with group-specific trends). Still, the GoT-like shows also experienced decreases in viewership that did not occur for other TV shows, which supports the conclusion that the leak had negative effects on TV viewership.

The results of the Google Trends analysis support my findings on the effects of the leak, although they indicate that the effects were pronounced for one search phrase only. However, I argue that the phrase “watch online” is actually the term that was most likely to reflect the effects of the leak. First, the word “torrent” implies some level of knowledge about how to download from the internet. Using torrents also requires the viewer to install specific software, which might be less attractive to new unauthorised viewers. For these reasons, it is unlikely that people looking for unauthorised sources of TV shows for the first time would start

[7] In an interesting case from 2017, a hacker or a group of hackers identifying themselves as TheDarkOverlord threatened Netflix that they would leak 10 out of 13 episodes of the new season of Orange is the New Black around two months ahead of the official schedule. Netflix refused to respond to the blackmail attempt, and the leaked episodes attracted many viewers. It is unclear whether this leak affected the Netflix viewership of the series, but Netflix’s refusal to acquiesce to the hackers’ demands shows that the leak might not have been harmful for this type of network.

with torrents. Second, the word “download” implies downloading the file directly to the computer. New unauthorised viewers could be especially concerned about saving files from suspicious, previously unknown sources. Moreover, downloading video files might require viewers to install an additional video player. For these reasons, it is also less likely that the people looking for unauthorised sources of TV shows for the first time would start by looking for ways to download the episodes to the hard drive. Finally, in 2015, streaming was by far the main source of unauthorised TV and film content, with streaming sites accounting for approximately 74% of visits to all sites with unauthorised TV and film content (MUSO, 2016). It therefore seems likely that the leak would have affected the popularity of the “watch online” search phrases only.

Interestingly, the effect in Google Trends was not immediate. Contrary to the regressions on viewership, I found a positive change in the trend for phrases with the search word “watch online”, but not a direct shift in the Google popularity of these phrases. One potential explanation for this finding is that the viewers no longer had to conform to the weekly schedule of the show’s episodes. Having learned how to watch the shows online, they might have decided to wait some time before watching the next episodes at a chosen time, instead of watching them immediately after the premiere. Moreover, having learned how to access almost any TV content online, some viewers might have decided to gradually explore TV shows they had not watched before, and especially the GoT-like shows. These two scenarios would help to explain why the change in the Google search happened over time, even though the viewership levels of GoT-like shows slowly returned to their initial levels.

The placebo tests for viewership support the notion that the decreases were characteristic of the year 2015 only. Just one of the four placebo tests showed a decrease in the viewership of GoT-like TV shows, but the decline was smaller than the decline in 2015, and was less significant in statistical terms. Moreover, the placebo tests for the Google trends analysis confirm that the simultaneous change in the Google traffic for the GoT-like shows occurred in 2015 only. Thus, I argue that the GoT leak indeed contributed to the declines in the TV viewership levels of both GoT and the GoT-like TV shows.

In summary, my results suggest that the leak of GoT caused some of the show’s viewers to switch to unauthorized sources for the first time, and to use these sources to watch other TV shows as well. These findings are in line with the observation made by Danaher et al. (2010) that a one-time incentive to switch to file-sharing might cause some TV viewers to make a permanent shift to unauthorised sources. My results are supported by evidence of a simultaneous change in the Google Traffic for the GoT-like TV shows. On the one hand, these outcomes suggest that these industries should focus their efforts on preventing the emergence of even short-term incentives to switch to unauthorised sources, by, for example, ensuring that no pre-premiere leaks happen in the future, and synchronizing release dates across different countries. Taking such steps is important because even one-time incidents could lead to a decrease in their future sales to consumers who have discovered how to use unauthorized channels. On the other hand, it is also possible that providing one-time incentives to use convenient and user-friendly authorized channels could have long-term effects. This issue should be explored in future studies on file-sharing. Moreover, taking down or blocking websites sharing unauthorised content could make some of their users revert to the authorised channels as they would again face the one-time costs associated with finding a reliable unauthorised source. Finally, the simultaneous recovery of the TV viewership of the GoT-like shows and the increase in search traffic suggest that the shift to unauthorised channels might have had promotional effects shows. However, how this promotion mechanism could be monetised is an open question as the findings point to the possibility that these viewers preferred to search for these episodes on the web.

References

Adermon, A., Liang, C-Y. (2014). *Piracy and music sales: The effects of an anti-piracy law*. Journal of Economic Behavior & Organization 105, pp. 90-106.

Cadario, R. (2015). *The impact of online word-of-mouth on television show viewership: An inverted U-shaped temporal dynamic*. Marketing Letters 26 (4), pp. 411-422.

Choi H., Varian, H. (2012). *Predicting the Present with Google Trends.* Economic Record 88 (Issue Supplement s1), pp. 2-9.

Danaher, B., Dhanasobhon, S., Smith, M.D., Telang, R. (2010). *Converting Pirates Without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy*. Marketing Science 29 (6), pp. 1138-1151.

Danaher, B., Smith, M.D. (2014). *Gone in 60 Seconds: The Impact of the Megaupload Shutdown on Movie Sales*. International Journal of Industrial Organization 33, pp. 1-8.

Danaher, B., Smith, M.D., Telang, R. (2014a). *Piracy and Copyright Enforcement Mechanisms*. Innovation Policy and the Economy 14 (1), pp. 25-61.

Danaher, B., Smith, M.D., Telang, R., Chen, S. (2014b). *The Effect of Graduated Response Anti-Piracy Laws on Music Sales: Evidence from an Event Study in France.* Journal of Industrial Economics 62 (3), pp. 541-553.

Danaher, B., Smith, M.D., Telang, R. (2015). *The Effect of Piracy Website Blocking on Consumer Behavior*. Working Paper, SSRN-ID 2612063.

Danaher, B., Smith, M.D., Telang, R. (2016). *Website Blocking Revisited: The Effect of the UK November 2014 Blocks on Consumer Behavior*. Working Paper, SSRN-ID 2766795.

Danaher, B. Waldfogel, J. (2012). *Reel Piracy: The Effect of Online Film Piracy on International Box Office Sales*. Working Paper, SSRN-ID 1986299.

Hammond, R.G. (2014). *Profit Leak? Pre-Release File Sharing and the Music Industry*. Southern Economic Journal 81 (2), pp. 387-408.

Hart, R. (2012). *Top 10 Reasons People Use To Justify Pirating Digital Content (And Why They're Wrong)*. LitReactor. URL: https://litreactor.com/columns/top-10-reasons-people-use-to-justify-pirating-digital-content-and-why-theyre-wrong [accessed: 2017-08-31].

Hiller, R.S. (2016). *Sales displacement and streaming music: Evidence from YouTube*. Information Economics and Policy 34, pp. 16-26.

Krawczyk, M., Tyrowicz, J., Kukla-Gryz, A., Hardy, W. (2017). *Do pirates play fair? Testing copyright awareness of sports viewers*. Journal of Behaviour & Information Technology 36 (6), pp. 650-661.

Krawczyk, M., Tyrowicz, J., Kukla-Gryz, A., Hardy, W. (2015). *„Piracy is not theft!" Is it just students who think so?*. Journal of Behavioral and Experimental Economics 54, pp. 32-39.

Lee, N., Jung, J.J., Selamat, A., Hwang, D. (2014). *Black-Box Testing of Practical Movie Recommendation Systems: a Comparative Study*. Computer Science and Information Systems 11 (1), pp. 241-249.

Ma, L., Montgomery, A.L., Singh, P.V., Smith M.S. (2014). *An Empirical Analysis of the Impact of Pre-Release Movie Piracy on Box Office Revenue*. Information Systems Research 25 (3), pp. 590-603.

McKenzie, J., Walls, W.D. (2013). *File-Sharing and Film Revenues: An Empirical Analysis*. University of Sidney Economics Working Paper Series 2013-14.

Millward Brown Digital. (2013). *Understanding the Role of Search in Online Piracy*. Report for the MPAA.

MUSO. (2016). *MUSO's Global Film & TV Piracy Market Insight Report 2016.*

MUSO. (2017). *Sample MUSO Global Piracy Report 2017*.

Peukert, C., Claussen, J., Kretschmer, T. (2017). *Piracy and Box Office Movie Revenues: Evidence from Megaupload*. International Journal of Industrial Organization 52, pp. 188-215.

Sivan, L., Smith, M.D., Telang, R. (2014). *Do Search Engines Influence Media Piracy? Evidence from a Randomized Field Study*. Working Paper, SSRN-ID 2495591.

Sood, S., Drèze, X. (2006). *Brand Extensions of Experiential Goods: Movie Sequel Evaluations*. Journal of Consumer Research 33 (3), pp. 352-360.

Steel, E. (2016, February 10). HBO Now Has 800,000 Paid Streaming Subscribers, Time Warner Says. *The Nw York Times*. Retrieved from: https://www.nytimes.com/2016/02/11/business/media/hbo-now-has-800000-paid-streaming-subscribers-time-warner-says.html

Witt, S. (2015). *How Music Got Free. The end of an industry, the turn of the century, and the patient zero of piracy.* Viking, United States of America.

Veitch, R., Constantiou, I. (2012). *User decisions in a (partly) digital world: Comparing digital piracy to legal alternatives for film and music.* European Conference on Information Systems (ECIS) 2012 Proceedings 174.

instytut
badań
strukturalnych

www.ibs.org.pl