Asymmetric content moderation in search markets: The case of adult websites*

Leonardo Madio† Matthew Mitchell‡
Martin Quinn§ Carlo Reggiani¶

This version: December 2024

We study the competitive impact of content moderation by a leading online platform. We exploit an exogenous shock that led the largest online platform in the adult industry to remove 80% of its video library. Using a difference-in-differences approach, we find that the content moderation of the platform led to a 37-51% drop in visits in one month, indicating strong user preferences for removed content. However, the traffic diverted was largely captured by other firms, including unregulated and often illegal fringe websites as well as mainstream competitors, which increased their traffic by 1.7% per week. Ultimately, the fringe sites saw an increase of 68% visits over 6 months. Search engines facilitated this migration - fringe websites experienced a surge in traffic from search channels, including aggregators, as users sought alternative content sources. The results highlight how asymmetric regulation can reshape competition dynamics by pushing users towards unregulated spaces, and how consumers substitute across platforms.

Keywords: Content Moderation, Platforms, Digital Services Act, Adult Websites, Search.

*We thank Anatole Cheysson, Sarah De Nigris, Néstor Duch-Brown, Chiara Farronato, Álvaro Gómez-Losada, Jack Peiyao Ma, Martin Peitz, Michelangelo Rossi, Lorien Sabatino and seminar participants at the JRC Digital Workshop (Sevilla, 2024), 8TH UNIBG Industrial Organization Winter Symposium, Padova Cybersecurity and Internet Governance workshop, and seminar in Télécom Paris for helpful discussion and comments. The views and opinions expressed in this paper are the authors' and do not necessarily reflect those of the Joint Research Centre or the European Commission. Leonardo acknowledges financial support for the "Cyber resilience: markets, investments and regulation" project - funded by European Union - Next Generation EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Universitá e della Ricerca).

† University of Padua, Department of Economics and Management, Via del Santo, 33, 35123 Padova, Italy. E-mail: leonardo.madio@unipd.it.

‡ University of Toronto, Canada. E-mail: matthew.mitchell@utoronto.ca.

§ Erasmus University Rotterdam, Netherlands. E-mail: quinn@rsm.nl.

¶ European Commission's Joint Research Centre, Seville, Spain and Department of Economics, University of Manchester, Manchester, UK. E-mail: carlo.reggiani@ec.europa.eu.

1. Introduction

Digital platforms that thrive on user-generated content are central players on the Internet. Millions of users contribute content daily to sites that include social media giants like Facebook and YouTube as well as specialized platforms like Reddit and Etsy. This widespread generation of content has raised questions about the regulation and moderation of online materials. As platforms grow, they must balance the tension between expanding content, which in turn may drive traffic, and protecting users from harmful material—ranging from hate speech and misinformation to intellectual property violations. Increasingly, laws such as the European Union's Digital Services Act (DSA) impose new standards on content moderation, particularly for platforms with large numbers of users (e.g., Very Large Online Platforms, or VLOPs).

Adult content platforms, which are among the most visited sites worldwide[1] face unique challenges due to the sensitive and potentially harmful nature of their content, including non-consensual pornography and child sexual abuse material (CSAM). This broader challenge of moderating harmful content extends to other types of platforms and is a critical issue for governments, regulators, and platform operators alike.[2]

This paper studies the impact of content moderation in a leading collection of platforms owned (at the time) by MindGeek hosting adult content. Following an exogenous change in content rules that reduced content up to 80%, we find that traffic declined suddenly at MindGeek sites by approximately 37-51% within one month, depending on the model specification. In the eight countries we considered, this amounted to a drop of around 25-28 million visits per day. Traffic diversion towards alternative sites was not immediate, and it grew at a rate of 1.7% per week. Within 6 months, traffic at smaller, less regulated sites had grown by 68%, and larger sites by 10%. Our results suggest that regulating only the largest platforms pushes problematic content to smaller, less-regulated spaces, raising concerns about the effectiveness of such policies in the broader digital ecosystem.

Our analysis also sheds light on the ability and willingness of consumers to switch between

[1] As of November 2023, two of the top five most-visited websites worldwide were adult platforms that we study—Pornhub (owned by MindGeek) and XVideos—ranking alongside Google, Facebook, and YouTube. See https://www.statista.com/statistics/1201880/most-visited-websites-worldwide/.

[2] As mentioned above, under the DSA the adult content VLOPs are subject to various obligations aimed at addressing issues such as child sexual abuse, deepfake porn, and the protection of users and creators. Additionally, they are required to empower and protect users online, including minors, as well as to assess and mitigate any systemic risks arising from their services.

platforms, which is central to questions of competition between platforms. Even after such a large change in content, it takes many months for traffic to move from one platform to another, suggesting substantial search costs for consumers. In this context, those costs might be socially beneficial, as it helps to regulate harmful content. But to the extent that these platforms have features in common with YouTube and other non-adult sites, the results suggest that there is substantial "lock-in" to sites that takes time to overcome for even large changes in the underlying site content.

MindGeek used a user-generated content model to become the leader in this space. However, this model also opened the door to significant issues with content moderation. The content change at MindGeek, the operator of Pornhub—the most visited adult tube—and several other major sites, was driven by scrutiny due to allegations of hosting illegal and non-consensual content. In December 2020, this culminated in a major scandal when The New York Times published a column titled "The Children of Pornhub," highlighting serious allegations that MindGeek was knowingly monetizing non-consensual and illegal content. The fallout was swift: Visa and MasterCard, two of the largest payment processors in the world, halted their services for all MindGeek platforms that same month.

MindGeek responded to this crisis by implementing a drastic content moderation policy, deleting 80% of its video library—specifically, all non-verified content where the identity of the uploader was essentially anonymous, i.e., around 10 million videos were deleted overnight on Pornhub.[3] Because this change was implemented at MindGeek's sites and not others, it had major implications for both the platform and the broader adult content industry. A large share of users, who had previously visited non-verified content, began searching for alternatives, sparking significant traffic redistribution across competing platforms.

This exogenous shock provides a unique opportunity to analyze how user behavior shifts in response to content moderation policies, and more generally in response to changes in users' match to the sites' content. Moreover, it allows us to investigate the broader competitive effects of such *asymmetric* moderation policies, which are implemented by only a subset of websites. First and foremost, are there preferences for non-verified content? How does removing vast amounts of content from a dominant platform affect competition in a market where search frictions may limit the ability of users to find similar content elsewhere? And to what extent do fringe websites,

[3]This change meant that the remaining content, if it violated the law, could be connected to the responsible uploader.

which are often less regulated and more permissive, benefit from such interventions?

To examine these questions, we employ a difference-in-differences approach, leveraging the plausibly exogenous nature of MindGeek's content removal as a quasi-natural experiment. We compare traffic patterns before and after the intervention, using non-adult websites as a control group to isolate the effect of the content removal. Additionally, we distinguish between traffic diverted to mainstream competitors (which may have some oversight but which face less external regulatory pressure, both public and private) and fringe websites, many of which host content with minimal oversight and often present online safety concerns (e.g., malwares, GDPR non-compliance, no age verification, copyrighted-infringing material).

We use data from Similarweb, which provided detailed traffic and user engagement metrics across 59 adult websites—including free streaming, aggregators, webcams, and subscription sites—as well as a control sample of non-adult websites. This dataset allows us to track user behavior, including visit duration, page views, bounce rates, and the origin of traffic (e.g., direct access, search engines, or referrals), enabling us to understand not only where users went but how they accessed alternative platforms.

Our main results show that MindGeek's content removal had a significant and immediate impact on the free-streaming market. There is evidence that some users had strong preferences towards non-verified content, and therefore substitute away from MindGeek entirely when the content is removed. We estimate that traffic to MindGeek-owned websites decreased by approximately 37-51% in the days after the intervention and remained almost constant over the following months.

In addition to an initial spike in traffic, the traffic diversion towards rivals increased at a rate of 1,7% per week. Competing mainstream websites saw only modest gains in traffic in the short-run and reached a gain of approximately 10% only after six months from the shock. The main beneficiaries were fringe websites, which had gained approximately 68% within six months, suggesting that users actively sought out unverified content on less regulated platforms. We also show that market concentration declined, although with the important caveat that increased competition ultimately benefited smaller but potentially more problematic rivals.

Our back-of-the-envelope calculations show that, as a result of the shock, MindGeek sites lost 25-28 million visits a day within a month in the eight countries we consider. In the subsequent period, rivals gained traffic by 1.6-1.7%.[4] Within 15 weeks, mainstream sites had experienced a

[4]Note that each fringe site receives on average 2.7% of the traffic of a mainstream one. Our back-of-the-envelope

daily gain of 7-24 million visits, whereas fringe sites saw an increase of 2.5-4 million, suggesting a substantial reallocation of traffic within the non-MindGeek sample. Within six months, competitor gains increased further as user searches expanded from within a single platform to multiple platforms, driving additional traffic to MindGeek's competitors.

Our analysis provides evidence about the role of search engines and referral-based aggregators in driving traffic to these alternative platforms. Specifically, we show that the increase in traffic for the fringe mirrors the increase in traffic generated to aggregators, smaller sites that provide links to videos at other sites. Traffic to these aggregators increases by 25% over six months. This shift highlights the critical role that search plays and that search is not solely in the hands of the main search engines.

To provide further evidence on the role of aggregators and search engines, we study traffic origins for adult tubes. We find that access through search engines and referral websites also increased, particularly for fringe sites. This increase in traffic was steady and consistent over time, reflecting a shift in search behavior from within a major platform, Pornhub, to fringe sites through both referrals and search engines. These results suggest that search channels can be more diverse than just the main search engines, and understanding these channels might be essential to understanding substitution across platforms.

We also show that search engines became a battleground for MindGeek, which sought to demote rivals' content by issuing Digital Millennium Copyright Act (DMCA) takedown requests to Google to remove URLs linked to allegedly copyrighted material. We find an increase in the number of URL requests that were unfounded or targeted URLs that were not yet indexed by Google, suggesting that MindGeek may have shifted its strategy to proactively target newly uploaded content on rival websites.

We also present additional results and implications. For example, we show that, following the shock, not only did visits change but also average visit duration, the number of pages viewed, and the bounce rate, i.e., the percentage of visitors that leave a webpage without taking an action (e.g., no further clicks) were also affected. The data suggest that the remaining MindGeek viewers increased the average time spent on the platform per visit and that the bounce rate decreased, meaning that they became more engaged with the platform after the massive content moderation. Instead, the effects on rival websites are more nuanced.

calculation should be seen as a conservative estimate, since we track in our dataset only a subset of fringe sites.

Contribution to the Literature. This paper contributes to the growing body of research on content moderation by online platforms. Much of the existing literature has concentrated on hate speech due to its well-documented impact on user engagement and platform participation (Beknazar-Yuzbashev et al., 2022). Several studies have explored the public enforcement of content moderation policies. For example, Andres and Slivko (2021) and Jiménez Durán et al. (2022) examine how Germany's NetzDG regulation has affected both offline and online hatred targeting minorities.

Other studies have focused on platform self-regulation. Müller and Schwarz (2022); Agarwal et al. (2022); Kumarswamy et al. (2023) and Rizzi (2024) examine different aspects of this issue. Müller and Schwarz (2022) investigate the impact of Twitter's deletion of President Donald Trump's account in January 2021, finding that the level of toxic content among Trump's followers significantly decreased post-deletion. Agarwal et al. (2022) explore the unintended consequences of Parler's deplatforming after the Capitol Hill events of 2021, showing that users migrated to Telegram, where they were exposed to increased levels of misinformation and partisan content. Kumarswamy et al. (2023) focused on Parler's self-regulation after it returned to the Internet and the App Store, finding that stricter content moderation reduced severe toxic content. Rizzi (2024) investigates the effects of Twitter's policy against racist hate speech, introduced in December 2020. She found a significant reduction in such content on Twitter, particularly among users previously identified as racist. Additionally and of particular relevance to our study, she explores spillover effects onto Parler, a less-regulated platform, finding an increase in hate speech among users active on both platforms during the policy's implementation.

While much of the existing literature focuses on content moderation strategies in relation to concerns over the safety versus freedom of speech possible trade-off, our study emphasizes the competitive impact of private enforcement in a market for user attention. Specifically, this paper is the first to provide systematic empirical evidence of traffic diversion from regulated platforms to alternative, less regulated sites and explore the dynamics of competition that arise from this. Lefouili and Madio (2022) suggest the idea that asymmetric content moderation can divert traffic from well-regulated large platforms to niche platforms less capable of identifying and removing illegal or harmful content.[5] We thoroughly explore these competitive dynamics and the consequences for both platform regulation and market outcomes.

[5]For a discussion on platform liability, see also Buiten et al. (2020) and Van Loo and Spier (2024).

Additionally, Madio and Quinn (2024) study theoretically the incentives of a hosting platform to moderate content when advertisers suffer from exposure to toxic or unsafe content but users may have a preference for them. Rendo (2024) extends such a framework to account for platform competition between a regulated platform competes against a fringe of unregulated providers. In the two cases, the theory identifies conditions under which the regulated platform either over-moderates or under-moderates content relative to the socially desirable level. Our empirical findings complement this theoretical framework by offering real-world evidence of how moderation policies influence platform competition and traffic diversion.

Recent research has also examined content moderation in the adult-only industry, although from different perspectives. Zeng et al. (2022) study the removal of two major commercial sex advertising platforms, Craigslist and Backpage.com, in the United States in 2018. Using this event as a natural experiment, they find no causal effect on prostitution arrests or violence against women, attributing this to the agility of advertisers in migrating to alternative websites. Similarly, Chan and Ghose (2014) study the effects of Craigslist's entry into the market, showing that it increased the prevalence of HIV across 33 US states over a 10-year period from 1999 to 2008. On a somewhat different note, Cunningham et al. (2024) find that Craigslist's entry into commercial sex advertisements reduced the risk faced by sex workers, specifically lowering female homicides by 10 to 17 percent. Abrardi et al. (2023) analyze traffic on adult websites before and after the implementation of the EU GDPR, finding an increase in traffic for the fringe post-implementation due to reduced privacy concerns.

Our study also contributes to the broader literature on supply versus demand side enforcement, particularly in the context of online piracy (e.g., Peukert et al. 2017; Aguiar et al. 2018; Dey et al. 2019; Danaher et al. 2020; Wlömert et al. 2023). A key novelty here is that the platform's content moderation is targeting a subset of content providers, and we can track the substitutability of that content across multiple platforms and the mechanism that makes traffic more elastic.

The theory literature on platform choice focuses on network effects, following from Farrell and Saloner (1985).[6] Although users in our context value content, in our case content is largely fixed, except for the almost instant removal event, so the main force in those models is not the primary driver of gradual migration. Instead, the gradual adoption suggests that search is a key driver.

We also offer a model in the spirit of Biglaiser et al. (2022) but where leaving one platform leads

[6]See Biglaiser et al. (2022) for a discussion of these papers.

to search, rather than directly landing onto another platform, and matching rates depend on content (which we take to change once, at the content moderation shock).[7] This reduced form is similar to a model of directed search from the labor literature (see, e.g., Moen 1997 and Burdett et al. 2001); in those models, matching rates are driven by wages, whereas in our framework, they are driven by content.

Roadmap. The structure of the paper is as follows. Section 2 provides information on the industry structure and the exogenous shock that affected MindGeek's business model. Section 3 introduces the data and provides the sample descriptives. Section 4 explains the empirical approach and identification strategy. Section 5 presents the main results. Section 6 identifies the main mechanism at play, with a particular emphasis on the role of search frictions. Section 7 presents concluding remarks.

2. Background

Overview of the market. Much of the Internet traffic comes from adult websites. The global porn industry generates significant revenue, currently estimated at $97 billion.[8] The US market alone contributes $13 billion annually. The porn industry's revenue streams are primarily driven by advertisements, subscription services like premium memberships and live webcams, and online porn games. The industry is a highly concentrated and lucrative market, dominated by a few platforms such as Pornhub, XVideos, XNXX, and xHamster, that mostly rely on user-generated and user-uploaded content to attract traffic.

The largest site, Pornhub, is part of a conglomerate—MindGeek (now Aylo) funded in 2004 and worth more than $740 million—controlling other prominent tube sites such as Redtube, Youporn, and XTube, as well as adult studios.[9]

[7] See Appendix A.

[8] See https://worldmetrics.org/porn-industry-statistics/.

[9] Since 2014, there has been considerable debate about the monopoly power of MindGeek and its acquisition strategies in the industry https://web.archive.org/web/20141219160919/http://www.slate.com/articles/technology/technology/2014/10/MindGeek_porn_monopoly_its_dominance_is_a_cautionary_tale_for_other_industries.html. In March 2023, MindGeek was acquired by Ethical Capital Partners, a private equity firm based in Ottawa, shortly after the release of Netflix's documentary *"Money Shot: The Pornhub Story"*, which focused on the controversies surrounding Pornhub, especially those highlighted by the "Children of Pornhub" column. The acquisition came just a day after the release of Netflix's documentary, which details the fallout from the content takedown and payment restrictions. Following the acquisition, MindGeek rebranded as Aylo, attempting to shift its public image in response to years of controversy.

The company's websites make revenue through ads and subscription services, such as Pornhub Premium, which provides ad-free access to exclusive content from creators and studios, as well as high definition content and the ability to download.[10] The free services monetize the videos by placing advertisements alongside it, using its own vertically integrated advertising network (that is, TrafficJunky). Users, often without professional affiliations, upload videos directly to the platforms. This open-access model has been central to their growth, enabling a large volume of diverse, often amateur, content that appeals to a broad audience. In addition to amateur contributions, professional studios and content creators also upload videos to these sites, especially where content is monetized through subscription services or ad revenue sharing (e.g., Pornhub and its premium version). At the same time, relying on user uploads, a portion of the content comes from less legitimate sources, including revenge porn, child sexual abuse material, and copyright-infringing videos.[11]

Pornhub's main rivals generate predominantly revenue through advertising.[12] Also these platforms are heavily dependent on user uploaded content.[13] In addition to xHamster.com, all of these domains have been designated by the European Commission (EC) as "Very Large Online Platforms" under the DSA.[14]

In addition to the major websites mentioned, the market is also populated by smaller niche sites and cyberlockers—online storage services—that allow users to share files and videos, which are often referred to as "the fringe". With minimal oversight, they allow anonymous uploads, making them havens for both piracy and, in some cases, harmful and non-consensual content. Users who visit fringe sites may also face risks related to data theft and privacy breaches.[15] These sites can be less secure than larger and more reputable platforms, increasing the likelihood of phishing attacks or data leaks. In many cases, these sites also host copyrighted material from mainstream sites (e.g. Pornhub) and escape liability because they are located in jurisdictions

[10] As this is content offered by creators and studios, such material is verified and moderated by definition.

[11] The 2021 Pornhub Transparency report reveals that approximately 500,000 content were removed for DMCA violations.

[12] The rivals rely on either integrated or third party ad networks. For example, xvideos.com uses TrafficFactory whereas xHamster engages exclusively with TrafficStars' services.

[13] Only recently xvideos.com and xHamster.com started to shift their business models toward a freemium model, allowing subscriptions to premium services. Additionally, they now enable creators to become verified, post content, and monetize their work. Moreover, these platforms have now developed their own payment network.

[14] This designation follows EC investigations concluding that these services surpassed the threshold of 45 million average monthly users in the EU in 2023. In December 2023, the EC designated Pornhub, XVideos (WGCZ), and Stripchat. In July 2024, the EC also designated xnxx.com (WGCZ), which is also part of the WGCZ group. See: https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6763.

[15] The fringe websites have also been known to expose users to malware, posing threats to privacy and safety. See: https://www.kaspersky.com.au/blog/porn-themed-threats-report/19477/.

with limited law enforcement.[16]

Emerging adult websites include OnlyFans and Stripchat.com. OnlyFans is a membership-based platform that allows creators to monetize their content directly, while Stripchat is the largest cam website, also designated in 2023 as VLOP by the EC due to its size. The content on these platforms is likely to be consensual, making them less likely to generate harm compared to the tube sites, where users can freely upload content, often without external verification.

"The Children of Pornhub" allegations. In December 2020, Nicholas Kristof published a column in the *New York Times*, titled "The Children of Pornhub", that uncovered allegations against MindGeek.[17] This investigative report made allegations that MindGeek knowingly allowed and monetized illegal and non-consensual content, including videos of child abuse, revenge pornography, and other illegal material.

Although MindGeek denied wrongdoing and claimed a zero tolerance approach toward illegal content, the article prompted MasterCard to investigate and subsequently block credit card transactions on Pornhub's premium service. Visa followed suit by suspending transactions on all MindGeek-owned sites on December 10th. Visa and MasterCard's actions were driven, in part, by the risk of being sued in abuse claims.[18] This move significantly disrupted MindGeek's cash flow, as its revenue heavily depended on premium subscriptions paid via credit cards, in addition to advertising revenues that also relied on credit card transactions. All major rivals were not affected by the intervention.

The purge. The intervention by Visa and MasterCard had a profound impact on MindGeek and the industry as a whole. MasterCard's decision came after an investigation into unlawful content, and Visa's subsequent suspension of transactions across all MindGeek sites further compounded the financial strain on the company. This meant that premium subscribers could

[16] See, e.g., lawsuit between MindGeek and the owner of Daftsex.com's. In 2022, MG Premium, a MindGeek subsidiary, won a legal case against the operator of Daftsex and other piracy sites, receiving an award of $32 million in damages along with the authorization to seize domain names. However, despite this ruling, no payments were made and the infringing sites remain operational. In 2023, in response to continued noncompliance, a US judge transferred ownership of the pirate domains to MG Premium. See https://torrentfreak.com/judge-gives-mindgeek-domains-after-porn-pirate-violates-injunction-230808/.

[17] See: https://www.nytimes.com/2020/12/04/opinion/sunday/pornhub-rape-trafficking.html.

[18] In 2022, a judge ruled that Visa could be held liable in such cases, stating that Visa "knowingly provided the tool used to complete a crime." See: https://www.bbc.com/news/technology-62372964. In 2022, Visa and Mastercard again suspended card payments, this time for advertising, on Pornhub and its parent company MindGeek. See https://www.cnbc.com/2022/08/04/visa-suspends-card-payments-for-ad-purchases-on-pornhub-and-mindgeek-amid-controversy.html.

not renew their subscriptions. Moreover, as much of the revenue of free porn sites comes from advertising, the intervention by payment networks had cascading effects.[19] Advertisers, fearing potential legal and reputational risks, withdrew their spending on MindGeek sites, causing a reduction in cash flow and putting MindGeek on the verge of default.

To address concerns about harmful content and mitigate the fallout, Pornhub—MindGeek's largest platform—enacted a massive content deletion on December 14th, reducing its video library from 13.3 million to 2.9 million videos by removing all non-verified uploads.[20] While the purge was aimed at improving trust and safety on the platform, not all the removed content was illegal.[21] The removed content also included memes and copyright infringing material. The first Pornhub transparency report was released in 2021 and acknowledged the removal of 653,465 items for infringements of terms and services.[22]

Date	pornhub.com	xnxx.com	xvideos.com
03/07/20	12,421,848	8,502,428	9,287,882
03/10/20	13,235,519	8,603,945	9,454,872
09/12/20	13,770,758	8,806,304	9,751,026
16/12/20	2,901,218	8,829,659	9,743,823
21/02/21	3,046,924	8,696,321	9,728,227
20/06/21	3,284,569	8,666,671	9,753,277

Table 1: Website content supply (number of videos available) for the top three adult websites by number of visits worldwide. Data on content supply was retrieved from `https://web.archive.org` using the URLs of the three websites for six different dates in our sample period, three before and three after the considered shock.

After the first radical removal, Pornhub's video library eventually stabilized at approximately 3 million videos. MindGeek's other tube sites also carried out comparable deletions of non-verified content. This was not the case for the exclusive content available to Pornhub Premium or associated studios, since all of these contents were verified.

Table 1 presents evidence on the evolution of content supply on the main platforms of the sector. More precisely, using data retrieved from the Internet Archive's Wayback Machine

[19]The Financial Times podcast *"Hot Money: Episode 5"* discusses how Visa and MasterCard's actions led to a financial collapse for MindGeek, disrupting both of these revenue channels. For further reference, see https://www.ft.com/content/1add56d6-82d9-4d83-a6a6-a5cdff70def5.

[20]This response was stricter than content moderation policies on other major platforms like YouTube, Facebook, and Twitter, the company claimed. See https://www.cnbc.com/2022/08/04/visa-suspends-card-payments-for-ad-purchases-on-pornhub-and-mindgeek-amid-controversy.html.

[21]This was clearly stated also in the testimony of MindGeek's officers to the Canadian Parliament, https://www.ourcommons.ca/DocumentViewer/en/43-2/ETHI/meeting-19/evidence.

[22]Following the incident, the platform engaged in a series of activities and took measures against illegal material, ranging from strict content moderation, mandatory uploader verification, and deploying a variety of tools scan incoming images and videos against hash-lists provided by NGOs (e.g., CSAI Match, PhotoDNA, etc.)

(https://web.archive.org), we gathered information on the number of videos available on pornhub.com and its two largest competitors, XNXX and XVideos.[23] The table illustrates the magnitude of content reduction in Pornhub, which experienced a 80% decrease in its video library, dropping from 13.7 million to less than 3 million videos. Meanwhile, MindGeek's largest competitors were not subjected to the same restrictions imposed by Visa and MasterCard, and, in turn, their content remained stable.

3. Data

The main data source for our analysis is Similarweb (similarweb.com), a data aggregation and software development company specializing in web analytics, web traffic and performance. Similarweb collects information on website traffic volume, as the total number of visits to each domain, as well as several measures of user online behavior. Their data are used by a large number of companies and institutions, as well as researchers (see, e.g., Lu et al. 2020, Schmitt et al. 2020, Calzada et al. 2023, Congiu et al. 2022, Abrardi et al. 2023, Sabatino and Sapi 2023).

The provider collects daily data along a number of relevant metrics, including:[24]

- *Visits:* total visits to a domain through all traffic channels. A visit is the access of a user to one or more pages of a website. All subsequent page views belong to the same visit, until the visitor is inactive for 30 minutes.
- *Average visit duration:* the average duration of the visit measured in seconds, calculated as the time elapsed between the first and last page view of a visit.
- *Bounce rate:* the share of users who close the website after landing on it without loading other pages.
- *Average pages per visit:* the average number of pages viewed per visit.
- *Page views:* the estimated number of page views for a domain.

A website is defined by its domain, so we use both terms interchangeably. *Visits* are the main outcome of interest. In line with the previous literature (Congiu et al., 2022), we refer to the

[23] We could not retrieve similar information for the fourth largest platform, xHamster.com.

[24] For more information about the variables available through Similarweb, one can refer to the API documentation at https://developers.similarweb.com/docs/similarweb-web-traffic-api.

other outcomes as *engagement measures*, as they provide information on how users interact with the website. We also construct a measure of the *average total time* spent on the website by multiplying the number of visits by their average duration. This new composite measure can be considered to be a measure of the attention generated by a given website.

Our data set also includes monthly information on how traffic is generated, that is, its origin. A domain receives traffic through different channels. The traffic originates via the *direct* channel if the user reaches it by typing the website address into the URL bar in the browser.[25] Other channels include *search* (i.e., when the traffic comes from a search engine such as Google or Bing), *referral* (i.e., generated by third-party links on websites other than social networks and search engines), *social*, or *email*. Finally, in Section 6.4, we also use data from the Google Transparency Report on DMCA takedown requests for copyright infringement.[26]

Table 2 provides the descriptive statistics of the traffic and engagement variables of our sample. The sample is made up of both adult and non-adult websites. Adult websites were chosen starting from the top 4 adult websites per traffic using the *Similarweb Similar Sites* tool, which returns a list of the most similar websites to a domain and a similarity score.[27] After cleaning and removing duplicates, a list of 59 domains was obtained. Non-adult websites, which will contribute to the construction of a control group for our main empirical analysis, were chosen between the most popular domains in the News & Media, Social Media, Movies and Entertainment and Search Engines in the countries of our sample, resulting in a list of 251 domains.

Among adult websites, we distinguish between *tube sites*—free streaming services such as Pornhub, XVideos, among others—*subscription* and membership services offering premium content (e.g., Pornhub Premium, OnlyFans), as well as *cam services* (e.g., Stripchat) and *aggregators*. Aggregators are defined as websites that redirect traffic to other sources, including adult forums and communities, porn search engines, or websites that embed videos hosted on other tubes in their sites. Our dataset comprises a total of 26 tubes, 11 aggregators, 8 webcam sites, and 14 subscription and membership sites.

The table presents an overview of traffic patterns across adult and non-adult websites before and after MindGeek's content moderation initiative. Here, the unit of observation is a domain in a given country. The table consists of two panels. Panel A compares key traffic metrics

[25] Direct traffic also covers clicking on a bookmark or links from outside the browser (e.g., a desktop shortcut).
[26] See: https://transparencyreport.google.com/copyright/overview.
[27] For more information, refer to https://developers.similarweb.com/reference/similar-sites.

Panel A: Adult vs non-adult websites				
	Adult		Non-adult	
Metrics	Before	After	Before	After
Visits	418,934	409,172	1,449,946	1,426,855
Pages views	3,686,801	3,609,975	11,494,923	11,633,863
Bounce rate	0.40	0.40	0.55	0.55
Pages per visit	7.14	7.19	4.09	4.11
Duration	342.22	340.48	275.02	275.58
Total time (×100)	2,533,076	2,401,776	8,311,461	8,518,290
N. domains	59		251	
Panel B: Adult websites				
	MindGeek		Non-MindGeek	
Metrics	Before	After	Before	After
Visits	739,385	582,389	330,385	363,353
Pages views	5,686,286	4,531,230	3,134,289	3,366,288
Bounce rate	0.47	0.46	0.38	0.39
Pages per visit	5	4.99	7.73	7.78
Duration	256.29	252.03	365.97	363.88
Total time (×100)	3,496,203	2,571,098	2,266,938	2,356,987
N. domains	15		44	
Domain type:				
Tubes	4		22	
Aggregators	0		11	
Webcams	0		8	
Subscription	11		3	

Table 2: Weekly aggregated average daily statistics by country-domain. Panel A reports statistics for adult and non-adult domains in our dataset. Panel B reports statistics for adult tube domains, further separating MindGeek-owned and non-MindGeek-owned domains. The sample period spans from July 1st, 2020, to June 20th, 2021. *Before* and *After* refer to the periods surrounding MindGeek's content removal on December 14th, 2020. Total time and duration are expressed in seconds.

between adult and non-adult websites. Before to the shock, on average an adult website in a given country averaged 418,934 visits, which slightly decreased to 409,172 after the content purge by MindGeek. Similarly, total page views for adult websites decreased from 3.69 million to 3.61 million. The bounce rate for adult websites remained steady at 0.40, while pages per visit showed a slight increase. The average session duration slightly declined. By comparison, non-adult websites exhibited minimal variation across these metrics.

Panel B focuses on adult websites, distinguishing between MindGeek-owned and non-MindGeek-owned domains. MindGeek's websites saw a considerable decline in visits, dropping from 740 thousand before the shock to 582 thousands after. Page views similarly fell from 5.7 million to 4.5 million. However, MindGeek's bounce rate remained almost stable (i.e., 0.46-0.47), as well as pages per visit, whereas the average session duration decreased modestly. The total time spent on MindGeek domains decreased from 3.5 billion to 2.5 billion seconds.

In contrast, non-MindGeek websites followed a different trajectory. The average visit per country-domain increased from 330 thousands to 363 thousands, and page views also grew from 3.1 million to 3.3 million. The bounce rate remained almost constant at 0.38-0.39, and pages per visit and average session duration also showed minimal changes. The total time spent on non-MindGeek websites rose from 2.27 billion to 2.36 billion seconds.

In Appendix B we provide additional descriptive statistics on tubes only, distinguishing between those belonging to MindGeek and those classified as fringe tubes and mainstream tubes (XVideos, xHmaster, and XNXX).

4. Empirical strategy

Our empirical strategy relies on a difference-in-differences (DiD) approach to estimate the effect of the radical content moderation pursued by Pornhub's parent company on traffic to adult websites. Specifically, we compare traffic patterns for adult websites (treated group) to non-adult websites (control group) before and after MindGeek's intervention, with a particular focus on distinguishing between websites owned by MindGeek and those owned by other adult-content providers. By removing about 10 million videos overnight, the shock to its library can be seen as exogenous, occurring in response to the decision by credit card providers (i.e., Visa and Mastercard) to stop processing payment transactions. This setup allows us to capture the

heterogeneous effects of the intervention in the different segments within the adult industry.

We start by using all non-adult websites as the control group, but we then implement two additional methods to select alternative suitable control groups in order to test the robustness of our results. Furthermore, in our main analysis the treated group consists of free tube adult sites, specifically those that allow users to upload content freely. Within this group, we distinguish between mainstream and fringe tube sites. Subsequently, we extend our analysis to encompass other types of adult websites, such as aggregators (which link to other sites), webcam sites, and subscription and membership services.

Difference-in-Differences Design. We start our exploration with a DiD analysis where all non-adult websites serve as the control group. The underlying assumption is that non-adult websites were unaffected by the radical content moderation initiative. Thus, if they followed a similar trajectory to the treated group before the shock and if there were no other major shocks that only affected one of the two groups, they could represent a valid counterfactual for what would have happened to adult websites in the absence of the regulatory shock. We estimate the following DiD regression:

$$\begin{aligned} Y_{ict} = \alpha &+ \beta\, After_{ict} + \gamma_1\, AdultMindGeek_{ict} + \gamma_2\, AdultnonMindGeek_{ict} \\ &+ \delta_1(After_{ict} \times AdultMindGeek_{ict}) + \delta_2(After_{ict} \times AdultnonMindGeek_{ict}) \\ &+ \mathbf{\Psi}_{ict} + \varepsilon_{ict}, \end{aligned} \tag{1}$$

where Y_{ict} represents the outcome variable for domain i in country c during week t, measured by various website activity metrics, such as visits in our main specification.[28] $After_{ict}$ is a binary variable equal to 1 if the observation refers to a period after December 14th, 2020, when Mindgeek's content removal took place (our shock). $AdultMindGeek_{ict}$ is a binary indicator for MindGeek-owned adult websites, while $AdultnonMindGeek_{ict}$ is an indicator for non-MindGeek adult websites. $\mathbf{\Psi}_{ict}$ is a vector of fixed effects, including fixed effects at country, website, country-website, and week levels. In some specifications, we also allow for a website-specific trend and exclude Thanksgiving and Christmas holidays (i.e., weeks 42-53 of the year 2020). Finally, ε_{ict} is the error term.

The coefficients of interest, δ_1 and δ_2, measure the treatment effects of MindGeek's radical content

[28] We use pages views, bounce rate, average visit duration, or monetization (total time) in Section 6.5.

removal on traffic to MindGeek-owned and non-MindGeek adult websites, respectively, relative to the control group of non-adult websites. Under the assumptions briefly anticipated above, the coefficients of the interaction terms δ_1 and δ_2 capture the average treatment effects (ATT) on these groups after the shock. In further analyses, we also interact $After_{ict}$ and $AdultnonMindGeek_{ict}$ with $fringe_{ict}$, an indicator for non-MindGeek fringe websites, to decompose traffic diversion between the fringe and the top 3 mainstream non-MindGeek sites.

In more detail, the identification of the ATT from equation (1) is based on several key assumptions. First, MindGeek's decision to swiftly remove content in response to VISA and Mastercard's withdrawal of payment processing was the only major exogenous shock asymmetrically affecting the adult industry, while non-adult websites were not directly impacted. It is likely that there were no significant anticipation effects, as the intervention was quickly implemented. MindGeek was pressured to delete non-verified content following the media coverage of the issue and the subsequent decision by VISA and Mastercard to stop processing payments.

Furthermore, we assume that the parallel trends assumption holds, meaning that adult and non-adult websites followed similar traffic trajectories before the drastic content removal, and would have continued to do so in the absence of the content shock. We provide evidence to support this assumption using an event study specification that extends equation (1) by including time-period interactions.

One potential threat to identification is the possibility of spillover effects, whereby, for example, non-adult websites may experience increased traffic due to diversion from adult websites. If such spillovers exist, they would violate the stable unit treatment value assumption (SUTVA), compromising the interpretation of δ_1 and δ_2 as the ATT for the treated groups. However, Figure 1 suggests that non-adult websites exhibited stable traffic patterns throughout the study period, suggesting that spillover effects, if any, were probably minimal.

Finally, to avoid confounding factors, such as MindGeek's temporary provision of free access to its premium services during the early COVID-19 pandemic period, we restrict our analysis to the period between July 1st, 2020, and June 20th, 2021. This reduces the risk that our estimated treatment effects are biased by unrelated shocks that affected the adult industry.

Although the baseline specification assumes that non-adult websites are a valid control group, there may be systematic differences between adult and non-adult websites that could bias the

results. To address this concern, we replicate our main analysis with two alternative estimation techniques. First, we employ Propensity Score Matching (PSM) before estimating (1). PSM is a statistical technique that matches treated and control units based on observable characteristics, thereby reducing selection bias (Heckman et al., 1998; Caliendo and Kopeinig, 2008). Second, we also apply the Matrix Completion - Nuclear Norm (MC-NN) method, which hinges upon machine learning techniques to predict the potential outcomes of treated units, and it is particularly useful in cases where there is unobserved heterogeneity or selection bias that is difficult to account for using traditional matching techniques (Athey et al., 2021). More details about these techniques, as well as the results of these analyses can be found in Appendix D.

5. Main Results

Descriptive evidence. Before presenting our main results, we provide some descriptive evidence. Given that the shock affected the content volume available across the major tube sites (i.e., streaming hosting platforms) under MindGeek's ownership, this part of our analysis focuses solely on this type of websites. In Section 6.3, we present results that also consider websites featuring professional and moderated content, such as cam websites and subscription or membership services.

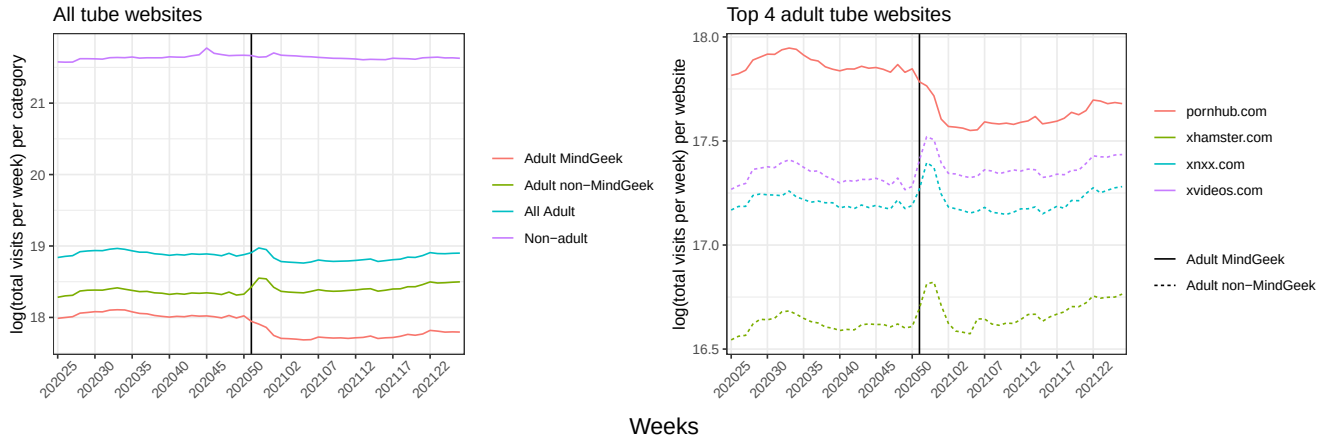


Figure 1: Visits per week in log scale between 1st of July 2020 and 20 of June 2021. The left panel plots the logarithm of the weekly sum of visits across countries and domains per group: non-adult domains in our sample (violet), adult domains (light blue), MindGeek domains (red) and non-MindGeek domains (green). Among the adult websites, we only focus on free tubes. The right panel plots the logarithm of the weekly sum of visits across countries for the four largest adult websites. The vertical dark line indicates the week in which MindGeek removed all non-verified content.

Figure 1 presents model-free evidence on the average weekly activity on adult and non-adult

tube websites.[29] The left panel includes all tube websites and plots the logarithm of the number of weekly visits across countries for non-adult domains in our sample (violet) and adult domains (light blue). While visits to non-adult domains remain relatively stable over time, visits to adult domains exhibit an initial increase immediately after MindGeek's intervention (vertical line), followed by a reduction and a subsequent rise towards pre-intervention levels. Visits to MindGeek websites (red) show a drastic and persistent decrease after intervention. This suggests that the radical removal of non-verified content led to a large number of users leaving the platform, indicating a strong preference for such content, non-verified although not necessarily unsafe or illegal. In contrast, non-MindGeek sites (green) experienced a surge in visits immediately after the content moderation implementation, which not only coincided with the holiday period but may also indicate users' increased exploration of alternative websites.The highlighted initial bump in visits, that mostly characterizes non-MindGeek sites plays a significant role and it will be discussed in detail in our ensuing analysis. Importantly, the trends in visits prior to the shock are quite similar across both adult and non-adult websites.

The right panel of Figure 1 focuses on the four main websites, three of which have been later designated by the European Commission as large platforms with moderation obligations. This panel shows the daily average visits per week and country, and confirms the substantial drop for Pornhub, the surge in traffic for rivals immediately after the MindGeek's purge, followed by a reduction towards pre-intervention levels and a subsequent increase over time.

Difference-in-Differences. We now present the initial set of findings from regression analysis. Table 3 displays the results of the DiD analysis, estimating the log of daily visits per domain for MindGeek and non-MindGeek adult websites, *vis-à-vis* the control group of non-adult websites.

The first two columns focus on the effect of the content removal by MindGeek on overall traffic towards adult websites relative to non-adult websites. Column (1) reports the estimate of a DiD specification that includes all fixed effects. Column (2) adds controls for domain-specific trends, accounting for the fact that websites may grow or decline at different rates over time (e.g., due to varying popularity or internal changes), independent of the treatment. Whereas Column (1) indicates that on average traffic to adult sites significantly increased relative to the control group after the shock, the coefficient of interest becomes insignificant when controlling

[29] Figure C.1 in Appendix C presents model-free evidence on the average weekly activity on adult and non-adult tube websites in levels.

	Dependent variable:							
	log(Daily Visit per Domain-Country)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
After × Adult	0.134***	−0.021						
	(0.040)	(0.021)						
After × Adult MindGeek			−0.647***	−0.466***	−0.718***	−0.570***	−0.460***	−0.684***
			(0.065)	(0.049)	(0.052)	(0.046)	(0.053)	(0.058)
After × Adult non-MindGeek			0.277***	0.060***	−0.049	0.084***	0.088***	0.004
			(0.036)	(0.017)	(0.030)	(0.027)	(0.017)	(0.029)
After × Adult MindGeek × t						−0.005***	0.003	0.008***
						(0.002)	(0.002)	(0.002)
After × Adult non-MindGeek × t						0.014***	0.014***	0.013***
						(0.001)	(0.002)	(0.002)
Domain FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Domain-Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Week FE	✓	✓	✓	✓	✓	✓	✓	✓
Domain Trends		✓		✓	✓		✓	✓
Holidays Included	✓	✓	✓	✓		✓	✓	
Observations	107,111	107,111	107,111	107,111	85,595	107,111	107,111	85,595

Table 3: Estimates of the traffic diversion between MindGeek and non-MindGeek's domains under different model specifications. Columns (1) to (4) consider the whole sample period, and column (5) excludes weeks 42 to 53 of the year 2020. Columns (6), (7) and (8) add the interaction term t which represents the weekly trend. Among the adult websites, we only focus on free tubes. Standard errors clustered at the country-domain level. Note: *p<0.1; **p<0.05; ***p<0.01.

for website trends. The previously estimated average treatment effect is now actually absorbed by the trend terms capturing differences in website growth over time. This suggests that the impact on the adult industry as a whole is limited and not statistically significant. Moreover, the finding also hints at the fact that time trends may play an important role in our data.

In Columns (3) and (4), we separately examine the effects on MindGeek and non-MindGeek adult websites. Column (4) adds to Column (3) website-specific trends. Both model specifications show that, following the removal of non-verified content, MindGeek's websites experienced on average a sharp reduction in traffic (i.e., a lower bound of 37%[30] approximately, according to Column 4) in visits to MindGeek's websites post-intervention.

However, the weeks surrounding the shock pose challenges to our identification as a result of various events and activities. For example, as remarked in the descriptive analysis, the control group experienced a short but sharp spike in visits during the week of Thanksgiving, whereas the surge in demand for adult content during the Christmas Holiday period is also well documented. To address these challenges, in Column (5) we exclude a period that includes both Thanksgiving and the Christmas holidays. This corresponds to a jump from week 42 of 2020 to the first week

[30]Note that $\exp(-0.46) - 1 = -0.37$. We use similar computations to derive percentage changes for subsequent results.

of 2021. In this case, the average reduction in traffic after content purge is estimated to be around 51%. We consider this estimate to be the most reliable measure of the immediate effect, as it corresponds to the period when all the anomalies in the data have faded.

According to Column (3), rival websites increase their daily number of visits on average by around 24% in the six months following the content purge by MindGeek. However, when accounting for domain-specific trends (Column 4), the average increase, which has to be interpreted as the immediate effect, is lower and equal to 6% on average. The effect becomes not statistically significant when excluding the Thanksgiving and Christmas holidays (Column 5).

In recognition of the importance of time patterns, highlighted by both descriptives and the DiD specifications with domain-specific trends, in Columns (6), (7) and (8), we add an interaction term t that measures the average change in visits per week for both MindGeek and non-MindGeek sites after the shock.

Here we find a difference between model specifications that account for the presence of holidays for Mindgeek sites, whereas the results are remarkably stable for its rivals. In particular, when holidays are included, the effect is negative and small in magnitude (Column 6) or not significantly different from zero (Column 7) for MindGeek sites, which seems to suggest that the drop in visits was immediate and persistent. However, when holidays are excluded, MindGeek experiences an estimated 0.8% growth rate per week.

As we shall see in what follows, this variation relates to the drop in visits that occurs within a single month of the shock, which then stabilizes and shows a slight recovery in the following months. When holidays are excluded, in light of the previously discussed abnormal patterns, the average effect captured by our variable reflects only the period during which the drop stabilizes (e.g., from week 2 of 2021) and, subsequently, the positive recovery.

Meanwhile, the non-MindGeek sites unambiguously grew at a weekly rate of 1.3-1.4%. The difference here is in the estimate of the immediate effect: when estimated including the holiday period, it amounts to 8.4-8.8% (Columns 6 and 7), whereas it is very small and non-statistically significant when the weeks around the shock are omitted (Column 8).

Furthermore, qualitatively similar results are also found when using DiD-PSM estimates and MC-NN (see Table D.2 in Appendix D), and similar patterns emerge when using visits in levels rather than logarithms (see Table C.2 in Appendix C). These analyses further corroborate the

validity of the empirical strategy and its results.

We posit that after the content moderation by MindGeek, users started searching for non-verified content across other websites, entering in a search pool. In Appendix A we present a theoretical model that rationalizes the almost immediate drop in MindGeek visits and the increasing trend for rivals.

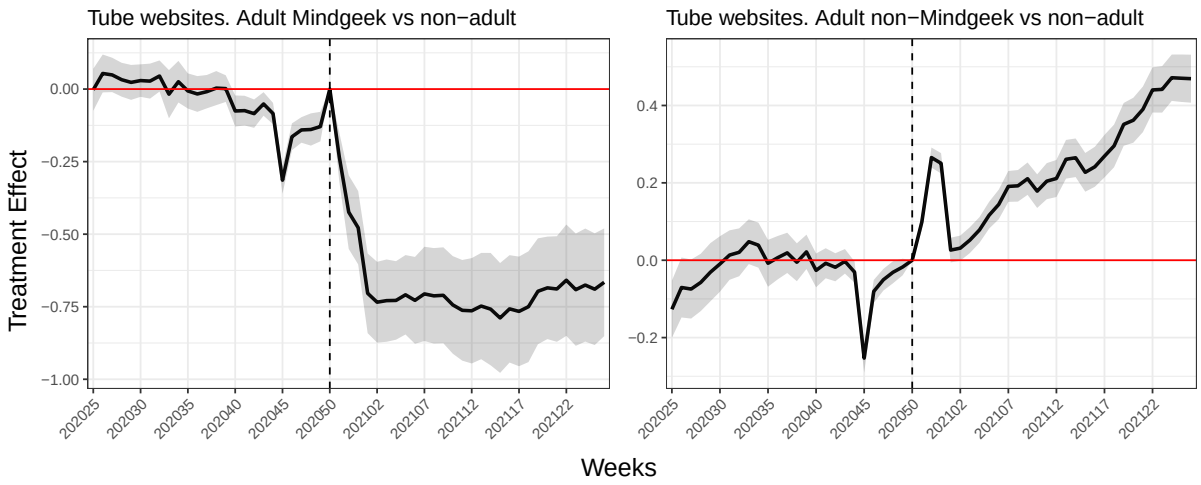


Figure 2: Event study analyses for the daily average number of weekly visits in log scale. The x-axes show the week. The y-axes report the estimated event study coefficients. Each observation is a country-domain observation in a given week. Standard errors are clustered at the country-domain level. The gray area shows 95% confidence intervals. The left panel plots the average treatment effect of domains belonging to the MindGeek group compared to non-adult domains. The right panel plots the average treatment effect of domains not belonging to the MindGeek group compared to non-adult domains. Among the adult websites, we only focus on free tubes.

The relevance of the growing increase in visits to non-MindGeek sites is further clarified in Figure 2, which presents the results of the event study and illustrates the dynamic treatment effects on adult websites relative to non-adult websites. The figure consists of two panels: the left panel displays the treatment effects for MindGeek-owned adult tube sites, while the right panel shows the equivalent for non-MindGeek adult tubes. Non-adult websites serve as the control group in both panels.

In particular, both panels show a significant drop in visits around week 45 in 2020, which is due to a change in average traffic in the control group (see, e.g., the bump in traffic for non-adult websites around week 45 in Figure 1). However, this is an outlier, potentially attributable to the Thanksgiving holiday, and it does not affect the validity of our analysis. Indeed, the event study moderates possible concerns regarding the validity of the parallel trends assumption: any significant difference in the traffic patterns prior to the content purge of both non-MindGeek

and MindGeek domains from non-adult websites appears to be non-systematic and potentially attributable to outliers (e.g., Thanksgiving).

The left panel of Figure 2 shows a significant and immediate negative treatment effect after the intervention for adult websites owned by MindGeek. The effect of treatment changes markedly negative immediately after the intervention (one-month period), consistent with the sharp decline in visits documented in the earlier descriptives and regressions. The effect stabilizes at less than half of the pre-shock traffic, and remains almost stable throughout the subsequent months, with a slight recovery towards the end of the sample period. This result suggests that MindGeek websites never recovered from the content removal shock, which led to a lasting impact on their user base. Within 6 months, MindGeek tubes had lost approximately 50% of their pre-purge daily traffic. Indeed, the event study shows the enduring impact of MindGeek's drastic content moderation, after which the platform hardly regained a substantial share of visitors.

In contrast, the right panel shows a markedly different trajectory for non-MindGeek adult websites. Immediately after the intervention, there is a spike in the treatment effect, reaching around 28% of their pre-shock traffic. However, as anticipated above, adult non-MindGeek websites exhibit a distinct pattern: after the initial and temporary surge in traffic, there is a gradual but constant increase throughout the post-intervention period.[31] This explains why, when accounting for website-specific trends, the immediate average increase in traffic for non-MindGeek tube sites is modest, but then it keeps increasing, and within six months these websites have gained approximately 60% more visits than their initial level.

Coming back to the temporary peak, it is also important to elaborate on what may have contributed to the bump in traffic for non-MindGeek websites immediately after the content purge. Two main factors concur to explain this. First, the Christmas period, which typically sees an increase in adult website consumption and a reduction in traffic to other types of sites.[32] Second, the likely increase in multihoming activity, where users searching for non-verified material no longer available on MindGeek's websites began exploring multiple platforms, each benefiting from increased visits. The subsequent reduction in visits could be due to users finding their preferred alternatives, while the longer-term upward trend can be explained by traffic substitution driven by search engines and referrals, which we analyze later.

[31] When removing holidays, our analysis basically cuts-out the spike and focuses on the effect starting from the second week of 2021.

[32] See, e.g., https://charismanews.com/culture/5-reasons-porn-usage-spikes-during-the-holidays/.

The event study results presented here are based on non-adult websites without any matching. Similar results using the PSM and MC-NN methods are provided in the Appendix D, confirming the robustness of the findings.

These results suggest that imposing moderation on some platforms while leaving others untouched leads to user migration, potentially undermining the intended policy goals. The experience of MindGeek provides a stark illustration: the immediate drop in traffic, coupled with the growth of rivals, explains why, in a *laissez-faire* regulatory environment, platforms have limited incentives to rigorously moderate unsafe or non-verified content, as content moderation encourages viewers to search for content on other platforms.[33]

5.1. Decomposing Traffic Diversion

As discussed in the background section (Section 2), non-MindGeek domains are highly heterogeneous. Some are established mainstream websites with a legal entity registered in specific jurisdictions, and since 2024, are subject to the EU DSA provisions. Others, however, are small websites, often malicious, relying on user-uploaded content with limited control over what is being uploaded. In other words, due to the size variability across non-MindGeek sites, the estimate of the ATT in our baseline model may fail to accurately convey the heterogeneity of the effects. For example, before the shock, visits at each fringe site were around 2.7% of the visits to a mainstream non-MindGeek site. Descriptive statistics are available in Table B.1 in Appendix B.[34]

In this section, we provide a decomposition of traffic diversion from MindGeek to fringe and top 3 rivals. We first present results using an event study approach that distinguishes between fringe and mainstream non-MindGeek adult tubes, i.e., the top 3 mainstream domains. The control group, as in the previous analyses, is composed of non-adult websites. Figure 3 presents the treatment effects for fringe websites and the top 3 mainstream non-MindGeek adult websites after the intervention. It is worth noting that before the intervention, all non-MindGeek sites followed a similar trend as the non-adult websites in the control group. As illustrated by the red curve, which shows that the non-MindGeek adult websites' treatment effect before the intervention had no systematic statistical difference with the control group. However, after the

[33] For a further discussion of these mechanisms, see also the theoretical background in Appendix A.

[34] Note that another source of heterogeneity may emerge across countries. However, it is hard to imagine a systematic heterogeneity in preferences for non-verified content between countries in our dataset.

intervention, the estimated treatment effects for non-MindGeek websites diverge between the fringe and mainstream tubes.

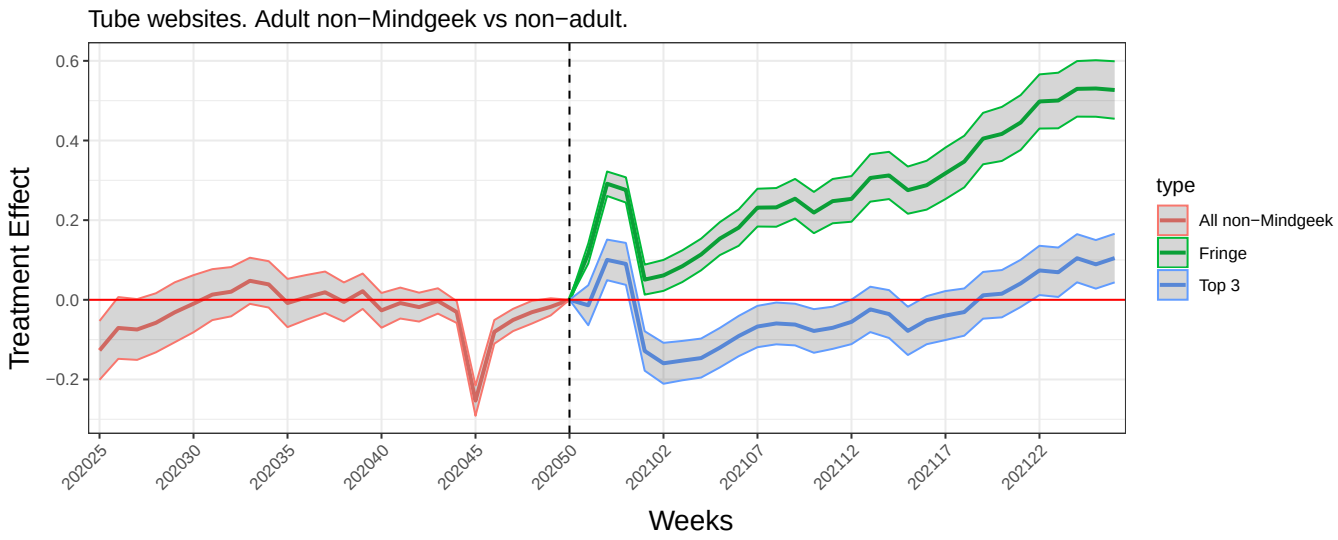


Figure 3: Event study analysis for the decomposition of the traffic diversion to both mainstream (top 3) and fringe websites in visits per week in log scale. The x-axis shows the week. The y-axis reports the estimated event study coefficients. Each observation is a country-domain in a given week. Standard errors are clustered at the domain level. The gray area shows 95% confidence intervals. Among the adult websites, we only focus on free tubes.

The figure shows a significant and sustained positive treatment effect for fringe websites after the intervention. Immediately after the policy shock, fringe websites experience a sharp increase in traffic, which continues to increase gradually over time. This suggests that fringe websites benefited considerably from MindGeek's content moderation measures. The substitution towards fringe websites is slow but persistent, with the positive treatment effects accumulating and remaining substantial even after several months. Within six months, the fringe websites had gained approximately 68% more traffic.

The top three mainstream non-MindGeek adult websites exhibit a different pattern. Here, too, traffic shows an immediate spike. Starting from the fifth week after content moderation (second week of 2021), these sites begin to experience an upward trend. The effect of content moderation on traffic to these sites appears almost negligible until week 20 of 2021. Ultimately, though, they gained approximately 10% more traffic within six months. We note that this increase is substantially smaller than that observed for fringe sites in percentage terms. Nevertheless, fringe sites are both more numerous (and not all are included in our dataset) and smaller compared to the top three mainstream non-MindGeek sites.

Interestingly, these results indicate that the diverted traffic following the shock was in a large percentage eventually redirected to fringe websites, whereas the flows to mainstream platforms are more modest when evaluated in percentage terms. They highlight an unintended consequence of the intervention: fringe websites—often associated with harmful and malicious content, lack of content verification, and non-compliance with privacy and copyright regulations—became significant beneficiaries. This shift poses serious risks to both users and creators. Content creators producing for MindGeek, where they could monetize via ads, lost exposure to alternative tubes that host pirated material and do not compensate creators for their content provision.

In Figure D.3 in the Appendix D, we reproduce the event study analysis presented above, accounting for the heterogeneous sizes of fringe and mainstream tubes by weighting the ATT by the average daily number of weekly visits for each country domain prior to the shock. The size and the dynamics of the estimated effects are comparable to the results presented so far, with the exception of the ATT for non-MindGeek sites as a whole. The six-months estimated ATT is reduced in size compared to Figure 2, attesting the heterogeneity of this group and reflecting the higher weight given to the top 3 mainstream sites.

Next, we quantify the traffic diversion using a triple-difference regression approach. Table 4 presents the estimates of the logarithm of the daily visits for both MindGeek and non-MindGeek adult websites, with an emphasis on the interaction between the intervention and the time trends (t). This interaction term helps capture how traffic patterns evolved over time after the intervention, particularly distinguishing between mainstream (top 3) and fringe non-MindGeek websites.[35]

Column (1) does not include domain-specific trends, whereas Columns (2) and (3) do. They, however, differ on the inclusion of holidays. When excluding holidays (Column 3), the immediate effect is calculated after 4 weeks from content moderation, and consistently with the event study figure, it is negative for both MindGeek and top 3 rivals. All firms face an upward-sloping trend. When holidays are included (Column 2), MindGeek loses approximately 37% of its traffic and never recovers, whereas rivals immediately gain a 6%, and then a further 1.6% per week.

In Appendix E, we provide a back-of-the-envelope calculation for the main effect based on results from Columns (2) and (3) of Table 4. Our estimates represent a lower bound of the effect on

[35]Similar results are provided in Table D.3 where we report DiD-PSM estimates and Matrix Completion (MC-NN).

	Dependent variable:		
	log(Avg Daily Visits per Domain-Country)		
	(1)	(2)	(3)
After × Adult MindGeek	-0.570^{***}	-0.460^{***}	-0.684^{***}
	(0.046)	(0.053)	(0.058)
After × Adult non-MindGeek	-0.072^{***}	0.064^{***}	-0.119^{***}
	(0.013)	(0.011)	(0.018)
After × Adult non-MindGeek × fringe	0.181^{***}	0.028	0.142^{***}
	(0.032)	(0.021)	(0.036)
After × Adult MindGeek × t	-0.005^{***}	0.003	0.008^{***}
	(0.002)	(0.002)	(0.002)
After × Adult non-MindGeek × t	0.005^{***}	0.016^{***}	0.017^{***}
	(0.001)	(0.001)	(0.001)
After × Adult non-MindGeek × fringe × t	0.010^{***}	-0.002	-0.004^{*}
	(0.001)	(0.002)	(0.002)
Domain FE	✓	✓	✓
Country FE	✓	✓	✓
Domain-Country FE	✓	✓	✓
Week FE	✓	✓	✓
Domain Trends		✓	✓
Holidays Included	✓	✓	
Observations	107,111	107,111	85,595

Table 4: Estimates of the traffic diversion to non-MindGeek mainstream (top 3) and fringe websites. Columns (1) and (2) consider the whole sample period, and column (3) excludes weeks 42 to 53 of the year 2020. Among the adult websites, we only focus on free tubes. Standard errors clustered at the domain-country level. Note: *p<0.1; **p<0.05; ***p<0.01.

fringe sites, as we have data for 19 fringe domains, while the long tail of smaller websites is much larger. Moreover, we only focus on 8 countries. Thus, our estimates of the effects on the traffic levels to fringe sites are conservative. Our estimates show that as a result of the content moderation, MindGeek sites lost 25-28 million visits per day within a month.

We also track the consumption diversion over time. Excluding holidays, mainstream sites experienced a growth of 7 million daily visitors by week 15, whereas the figure reached 24 millions when the holiday period is included.[36] Fringe sites, in contrast, saw an increase of 2.5 to 4 million daily visitors within the same period.

Over six months from the shock, the fringe sites in our sample gained approximately 6-7 million daily visits, while mainstream sites experienced a rise in traffic amounting to 27-43 million visits. Thus, within six months, the lost traffic from MindGeek was more than fully compensated for

[36]Note that the 7 to 24 million range is quite large, but this is simply due to the fact that, when holidays are not included, the immediate effect (captured at week 2 of January 2021) on visits is negative and statistically significant.

by the increase in visits to its competitors, both the fringe and the mainstream sites.

In Appendix F, we also show that the shock and the subsequent change in user behavior, particularly among those searching for unverified content, had a pivotal impact on market dynamics. Industry concentration, measured by the Herfindahl-Hirschman Index (HHI) of the visits received, decreased significantly after the shock. MindGeek lost almost 10% of its overall visits' share, while the both the mainstream and the fringe websites in our dataset increased their joint share of visits.

As we shall see, most of the traffic post-shock was diverted via aggregators, and this may have led to multiple counting of visits across sites of what were previously considered visits within site. Whereas our data do not allow to fully evaluate the extent of the phenomenon, in Appendix G we allay concerns about this issue and show that our results are qualitatively unaffected when considering viewership time on the platforms.

One could also argue that indirect network effects may have also played a role. First, content creators did not shift their supply of content to alternative tubes because they were unable to monetize their work on those platforms. If anything, they moved their content to subscription and membership platforms such as OnlyFans, as we discuss in Appendix H, or to cam sites, as we discuss in Section 6.3. Second, while users could have uploaded deleted content elsewhere, we observed no significant change in the content supply of major competitors (see Table 1). It is possible that users posted removed content on fringe tubes, but this seems unlikely since it would have required them to download the content to their own devices prior to the shock, which was unanticipated.

6. Mechanisms

In this section, we provide more insights on the mechanisms that can explain our results. We place particular emphasis on the role of search engines and referral links in shaping user navigation patterns, especially toward fringe websites. We also study traffic diversion towards cam sites (Section 6.3) and membership sites (Appendix H).

6.1. The Role of Aggregators

Adult tube sites receive visits from a wide range of sources, an important one being aggregators. We shall start our analysis of the mechanisms by looking at their role in the aftermath of MindGeek content purge. Aggregators are sites that collect data from other sources across the internet and put the information in one place where users can access it. In this section, we analyze how these aggregators adapted to content moderation, reflecting a shift towards fringe websites. We provide evidence indicating that visits to aggregators lead to increased traffic for these sites, suggesting a funneling effect. We further substantiate this by examining referral traffic from aggregators.

More specifically, we shall now analyze the traffic data of various adult aggregators in our dataset, including specialized search engines, video-hosting sites, platforms embedding external videos, and dedicated community forums. These aggregators, by pooling links and content, significantly influence traffic directed to fringe websites. This analysis may be conservative, as some users might engage directly with content on the aggregators, without visiting the original tube sites.

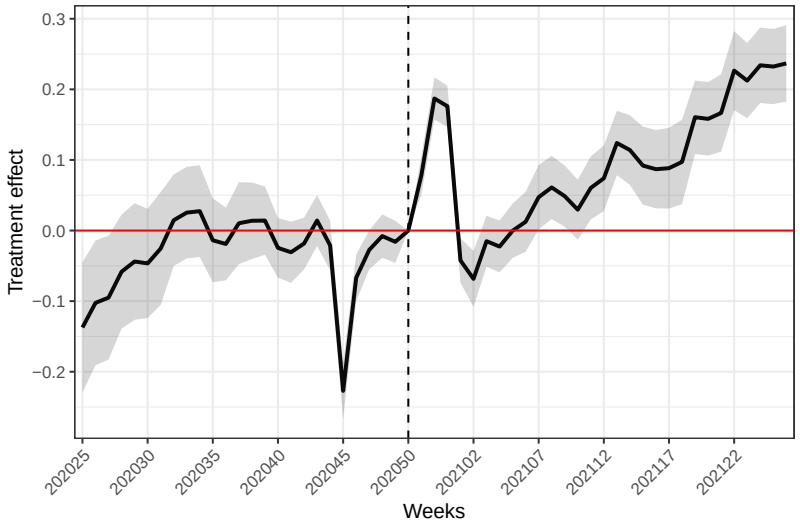


Figure 4: Event study analysis for the daily average number of weekly visits to aggregators relative to non-adult websites in log scale. All aggregators are independent and not belonging to either MindGeek or WGCZ. The x-axis shows the week. The y-axis reports the estimated event study coefficients. Each observation is a country-domain observation in a given week. Standard errors are clustered at the domain level. The gray area shows 95% confidence intervals.

We conduct an event study analysis where the control group still consists of non-adult websites, whereas the treatment group now features the domains of the aggregators in our sample. The unit of observation is still the country-domain, and the dependent variable is the logarithm of

the average daily visits. Figure 4 provides some support for the parallel trends assumption: in the vicinity of the shock and prior to it, there were no significant differences between the two groups, except for week 45 of 2020, when a notable traffic spike was observed for non-adult sites (our control group).

Importantly, the figure also shows that traffic to aggregators increased substantially and steadily after MindGeek's content purge. As a result, the finding suggests that the aggregators became major access points for traffic, facilitating the discovery of fringe content. This effect closely mirrors the findings for fringe websites presented in Figure 3. Over six months, traffic to aggregators rose by approximately 28%, highlighting their role as major drivers of traffic diversion to the unregulated market. Also in this case, the increase is gradual, meaning it takes time for traffic to be reallocated and search costs to be eventually overcome.

6.2. Traffic Origin

To shed further light on the search mechanism that leads to the discovery of alternative websites, including fringe ones, and the role of referrals, we now focus directly on the *traffic channels* that direct visits to the domains in our sample.

One of the key factors contributing to the rise of fringe websites is the changing role of search engines as a gateway for users to discover and access content. Large platforms such as Pornhub traditionally host a wide variety of content, including both verified and non-verified material. Users familiar with these websites typically access them directly by typing the URL of the website in their browser. However, shifts in the availability and distribution of content, particularly following the intervention in MindGeek's content supply, appear to have altered user behavior in terms of how they access adult websites.

We provide descriptive evidence that is suggestive of the fact that, after the shock to content supply on MindGeek's platforms, some users changed their access points, shifting from direct access to search engines and referral-based aggregators. This shift is likely to have been a crucial driver in the discovery of fringe websites, which ultimately captured a significant portion of the traffic diverted from MindGeeek websites.

Figure 5 plots the logarithm of the monthly traffic data for the different free tube domains in our dataset, distinguishing between the channels from which the visits originated: direct access,

referrals and search engine traffic. The top panel displays the traffic origin channel (direct, referrals, search) per each group of websites. The bottom panel shows the visits received by the different groups of domains, per each origin channel. Referrals include traffic driven by aggregators, which increased following the content moderation shock as demonstrated in the previous section.

Let us first consider MindGeek, shown in the top-left panel. We observe a steady decline in traffic originating from direct access, which is gradually overtaken by traffic from search engines around the time of the content purge. Interestingly, referral traffic, which remained constant throughout the pre-treatment period, plunged after content purge and never recovered. This could be due to two reasons: first, many links were no longer working due to content removal, and second, links to other sites may have eventually become more prominent as 80% of Pornhub's content was taken down. This suggests that content moderation hits specific content and is not random, since aggregators were much less likely to find content on MindGeek after content moderation.

Consider then the fringe websites, in the middle-top panel. We observe here an upward trend for both referrals and search engines, and no substantial change for the direct access following the treatment. Interestingly, access by referrals spikes immediately in correspondence of MindGeek's intervention and continues over time. This suggests that the channel became an essential access point for users discovering these fringe platforms. Since it shows an increasing trend in the period after content purge, its role can contribute to explaining the main results of the event study and the difference-in-differences analysis in Section 5 on the trend-driven growth of fringe websites. Moreover, search engines surpassed direct access as a source of traffic precisely in the period surrounding the shock, and then consolidated their role as the main source of traffic to the fringe websites in the following weeks.

Finally, consider the non-MindGeek mainstream websites, in the top-right panel. Here it is interesting to see that there is no significant variation in access via search engines, a trendy decline in direct access that starts before the shock, whereas there is a surge in traffic via referrals just after the shock. Also in this case, referrals seem to have played a critical role in generating traffic towards sites alternatives to MindGeeks' ones, consistently with the fact that users changed their habits: from search within a major platform, Pornhub, to search across platforms through search engines and aggregators. These results are indeed consistent with the patterns observed in Section 6.1 regarding the role of aggregators.

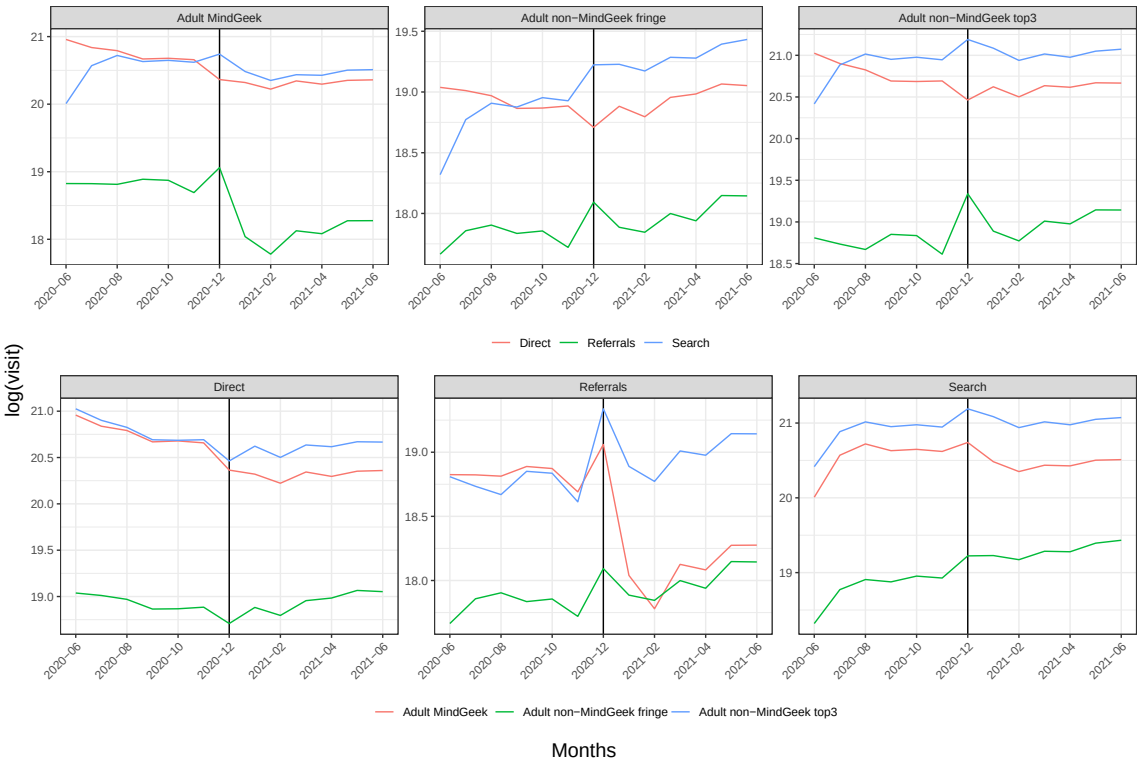


Figure 5: Monthly traffic data in log scale to tube websites by access type aggregated by group: domains belonging to MindGeek, mainstream domains not belonging to MindGeek (i.e., xnxx.com, xvideos.com, xHamster.com), and fringe domains not belonging to MindGeek. The top panel presents, for each group, the log of the visits originated via direct channel, referrals, and search. The bottom panel presents, for each access point, the log of the visits per each group, i.e., MindGeek, fringe and mainstream non-MindGeek domains.

Across all categories of websites, search engine is the leading channel of traffic origin post-intervention, suggesting that users increasingly relied on search engines to find content. This confirms once again the key role of these intermediaries within the digital economy.

Moreover, focusing on the bottom-right panel ('Search'), the trend observed for the fringe (green) is particularly relevant to explain the ascent of this group of small websites in the aftermath of the MindGeek's purge, as search can generate discovery. The bottom-left panel depicts traffic originated via direct access. Both MindGeek and other leading adult websites were characterized by a declining trend in this type of traffic. However, direct-access traffic declined further for MindGeek tubes than for its main rivals in the aftermath of the shock. Traffic through the direct channel, instead, increased slightly for the fringe sites after the shock. Finally, the bottom-middle panel confirms that the sharp drop in referral traffic to MindGeek's tube sites was partially and slowly compensated by an increase in traffic from this source to both groups of rival sites.

Overall, the steady access to fringe content through search engines and referrals emphasizes the

significant role that search mechanisms play in content discovery, extending beyond the control of major search engines. Aggregators and smaller sites that link to external videos increasingly contribute to this process. This diversification of search channels implies that, in the presence of user preferences for unverified information, the ease of overcoming search frictions, whether facilitated by traditional search engines or alternative platforms, can undermine the effectiveness of potential regulation of large platforms.

6.3. Diversion to Cam Sites

The adult entertainment market also includes live streaming websites, such as cam sites like Stripchat.com, which in 2023 was designated as a VLOP by the European Commission under the DSA. For this reason, it is important to examine the search diversion to these platforms. Unlike free tube sites that rely heavily on user-uploaded content, cam sites provide live content that is both verified and monetized.

The event study analysis provided in Figure 6 shows that, relative to non-adult websites, the traffic to cam sites increased substantially and steadily after the content purge. This effect qualitatively mirrors the one identified for fringe websites.

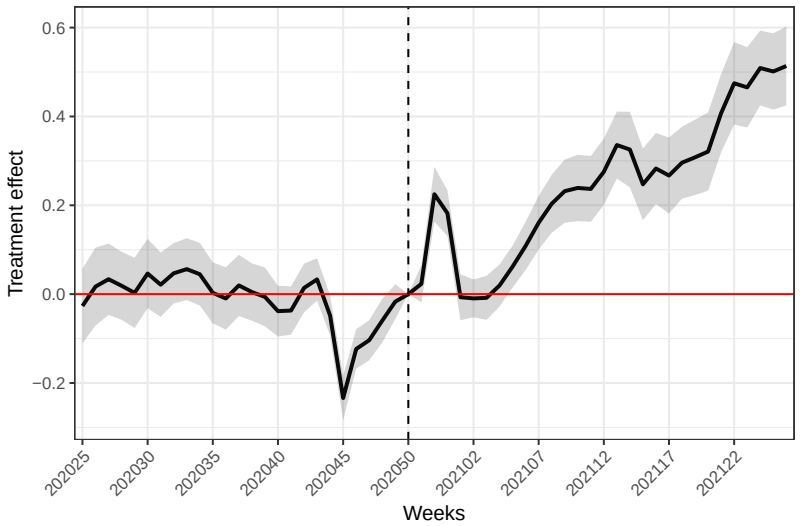


Figure 6: Event study analysis for the daily average number of weekly visits to cam websites relative to non-adult websites in log scale. All cam websites are independent and do not belong to either MindGeek or WGCZ. The x-axis represents the week, while the y-axis reports the estimated event study coefficients. Each observation is a country-domain in a given week. Standard errors are clustered at the country-domain level. The gray area represents 95% confidence intervals.

There are two possible explanations for this result. First, users who previously consumed non-

verified content on MindGeek's platforms may have migrated to these cam sites, seeking similar user-generated content, which was often uploaded without the creators' permission. Second, users who previously consumed verified content on Pornhub Premium, but were unable to renew their subscriptions (see Appendix H), may have moved to cam sites to access comparable content. Regardless of the specific mechanism, this change in cam content has proven particularly relevant, with cam websites experiencing an almost 65% increase in traffic in six months.

This traffic diversion appears less problematic than the diversion to fringe websites for two reasons. First, creators on cam sites can monetize their content, which would not have been the case if their material had been recorded and uploaded elsewhere. Second, the content at the cam sites is typically consensual and safer.

6.4. Strategic Reaction and Competition on Search

In this section, we present evidence suggestive that competition for traffic on search engines intensified significantly, consistent with the findings from the previous section. Furthermore, we show that MindGeek adopted a more strategic approach to targeting its rivals in this context.

Under the DMCA framework, online search engines such as Google Search and Bing can only remove infringing URLs from search results; the content itself remains accessible via direct URLs unless the copyright holder takes further action to notify the rival platform. MindGeek extensively leveraged DMCA requests to address this issue. According to the Google Transparency Report,[37] MindGeek is the largest individual copyright holder that submits DMCA takedown requests and the second largest organization overall, behind only Link-Busters.com, a firm specializing in DMCA services for multiple copyright holders.[38]

Before the shock, MindGeek hosted a diverse portfolio of content that included both copyrighted and infringing/problematic material. Rivals, mostly platforms with user-generated content, primarily hosted videos uploaded by users, some of which included copyrighted material produced or owned by MindGeek's studios or creators on its platforms. Essentially, competitors were offering unauthorized access to MindGeek's copyrighted material. This infringement occurred when users uploaded videos to MindGeek's platform, which were later hacked and reposted on

[37] See https://transparencyreport.google.com/copyright/overview.

[38] Similar patterns can be observed in Microsoft Bing's Transparency Report. See https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW1k79Z.

competing websites. To combat this, MindGeek routinely submitted DMCA takedown requests to Google and Bing to remove links to unauthorized videos from search results.

In this section, we analyze data from Google's DMCA takedown requests, particularly those submitted by MindGeek Premium Ltd., to assess whether the volume and nature of these requests changed after the company removed 80% of its catalog. This removal marked a significant shift in competition as it led to an increased reliance on search engines to access content, thus intensifying competition with its rivals.

The dataset provided by Google distinguished URLs into the following groups:

- URLs that were removed by Google, thus recognized as copyright-infringing;
- URLs for which Google took no action, indicating that Google did not find evidence of copyright infringement or for which MindGeek was not the copyright holder;
- URLs that were not indexed by Google at the time of the request. According to Google, indexing takes 1-2 weeks even when there is a request by the domain hosting the content. Moreover, if a URL is not indexed, it would not appear in Google's search results. Therefore, there is no need to remove it, as it is already absent from the index.[39]

We restrict our analysis to organizations submitting a sufficiently high average number of URLs for take down, i.e., the top 500 reporting organizations, and we construct a balanced panel on that basis. The focus on the top 500 submitters is justified by the existence of a long tail of URL request submitters, some of whom are very small and infrequently reporting organizations. In Appendix I we provide details on the monthly URLs requests by MindGeek and the other reporting organizations (see Table I.1). In our analysis, we then use all other 499 reporting organizations as a control group and focus on monthly submissions. Since we have count data of the number of reports, we estimate the following Poisson regression:

$$\Lambda_{it} = \alpha + \beta\, After + \gamma\, MindGeek_t + \delta After \times MindGeek_t + \mathbf{\Psi}_i + \mathbf{\Omega}_t + \varepsilon_{it},$$

where Λ_{it} stands for the URLs submitted by the reporting organization i at month t per each of the above groups, $\mathbf{\Psi}_i$ is the organization fixed effect, whereas $\mathbf{\Omega}_t$ is the month fixed effect. We focus on the total number of URL requests, those for which there was no action by Google, those removed by Google, and those that were not indexed yet.

[39] See, e.g., https://support.google.com/webmasters/answer/9012289?hl=en.

Table 5 reports our main results. The first column shows no significant effect on the total number of requests to Google, although the coefficient is positive. There is a statistically significant increase in the number of requests for which Google did not intervene, which suggests that MindGeek might have started strategically demoting its rival's content on search engines, even for content that was not infringing or about which there was uncertainty. This may indicate a strategic move by MindGeek to suppress the visibility of competing content.

	Dependent variable: URLs monthly requests			
	Total Requests	No Action	Not Indexed	Removed
After × MindGeek	0.1826	0.4400**	0.5625***	-0.1109
	(0.1440)	(0.2076)	(0.1691)	(0.1473)
Month FE	✓	✓	✓	✓
Organization FE	✓	✓	✓	✓
Observations	6,500	6,500	6,500	6,500

Table 5: Poisson regression on the top-500 reporting organizations. The treatment group is composed by MindGeek Premium Ltd., whereas the control group is composed by the remaining 499 major reporting organizations. Data aggregated at the monthly level between June 2020 and July 2021. Standard errors clustered at the organization level. Note: *p<0.1; **p<0.05; ***p<0.01.

The data also show a significant increase in the number of takedown requests for URLs that were not indexed by Google post-shock, which suggests that MindGeek might have shifted its strategy to proactively target newly uploaded content on rival websites. This may support the idea that MindGeek, recognizing the importance of search engine visibility, began to focus on fresh content, attempting to prevent it from being listed on Google altogether.

Interestingly, while the number of URLs removed by Google decreased over time, the effect is not significantly different from zero. This could indicate that MindGeek was flagging content more quickly after its upload, before it was indexed by Google. The decrease in removed URLs may reflect the fact that MindGeek's new strategy prevented infringing content from ever appearing in Google's search results.

Thus, the number of requests remains stable, suggesting a temporal substitution between the unindexed content and the removed ones. In particular, the significant increase in non-indexed URLs suggests that MindGeek strategically targeted fresh content from rivals, aiming to curtail their ability to gain traffic through search engines.

In addition to the findings from the Poisson regression analysis, Figure 7 presents an event study

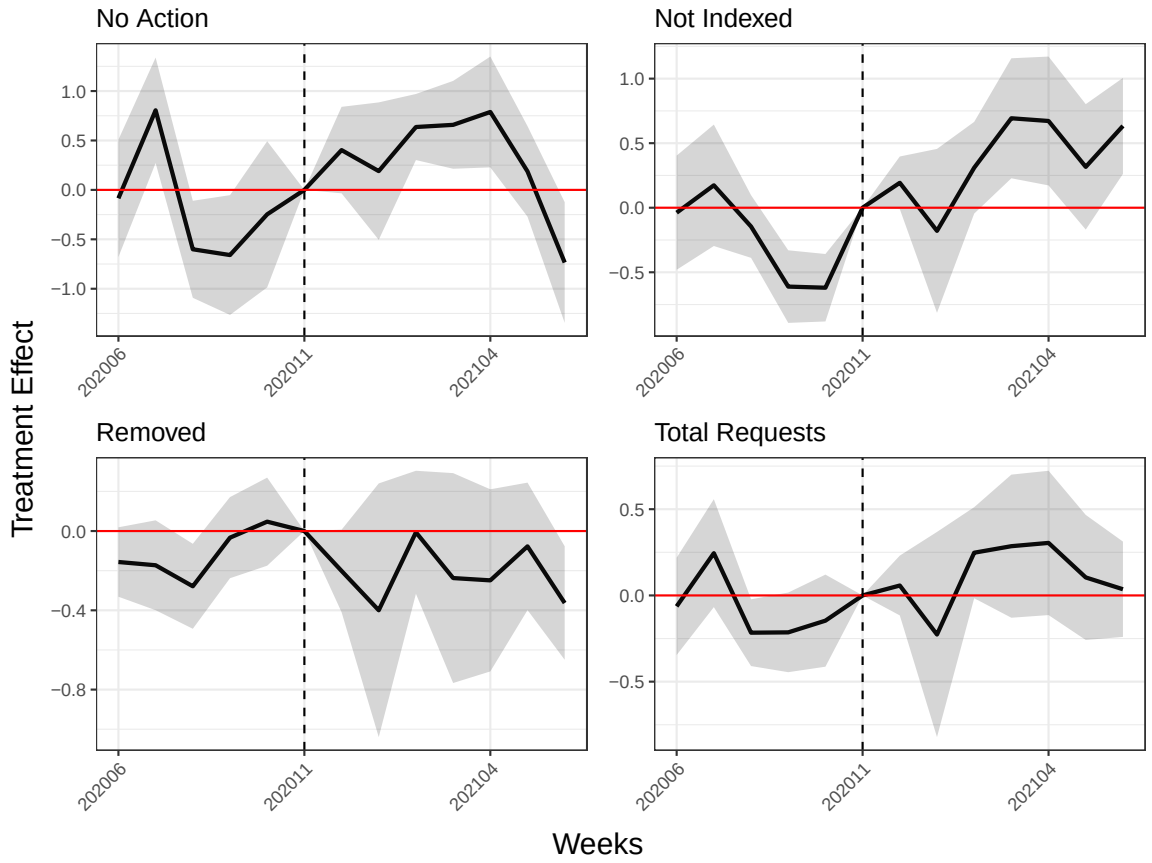


Figure 7: Event study analysis for the monthly average URLs submissions by MindGeek relative to the top-500 organizations. The x-axes show the month. The y-axes report the estimated event study coefficients. The gray area shows 95% confidence intervals.

with additional information on the changes in MindGeek's strategy following its massive content moderation initiative. Although the analysis is noisy due to the large heterogeneity between MindGeek and other organizations, there is evidence of an increase in the number of requests by MindGeek that were followed by no action by Google or whose URLs were not indexed yet.

6.5. Other Metrics of Engagement

Our dataset contains other measures of engagement in our dataset, such as page views, duration, bounce rate, and pages per visit. We can therefore study how the content purge by MindGeek affected these additional metrics.

Figure 8 presents event study plots for all such measures in MindGeek's tube domains, while Figure 9 shows event study plots for tubes not owned by MindGeek. For MindGeek sites, we observe a significant reduction in page views. However, the number of pages per visit and average duration increased significantly, and both displayed a 22% increase after six months, suggesting

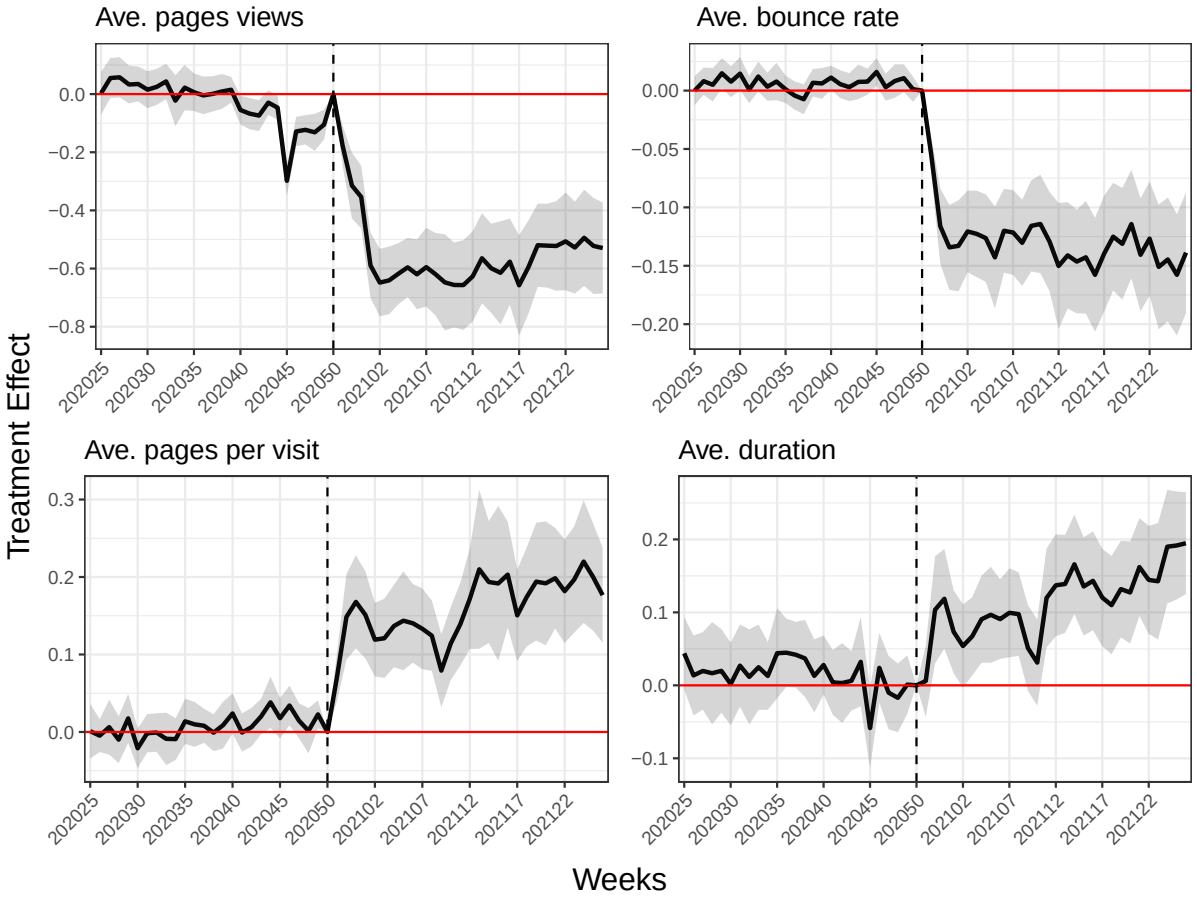


Figure 8: Event study analyses for different measures of engagement for MindGeek's domains using daily average measures per week in log scale. The x-axes show the week. The y-axes report the estimated event study coefficients. Each observation is a country-domain observation in a given week. Standard errors are clustered at the domain-country level. The gray area shows 95% confidence intervals. The control group consists of non-adult websites.

that infra-marginal users (those who remained on the platform) became more engaged with the available content, or were the most engaged users to begin with. Indeed, these users likely represent a more loyal or dedicated subset, as they navigated more deeply through the content despite the overall decline in traffic. Furthermore, the bounce rate decreased significantly and substantially, indicating that users were less likely to leave after viewing only a single page. This may reflect an increased focus on verified content, which could have attracted users specifically seeking higher-quality or safer material, or that the platform recommendation system became more precise because of the more homogeneous user base.

In contrast, non-MindGeek websites experienced a positive shift in the average page views, likely benefiting from the users who migrated away from MindGeek platforms. Unlike MindGeek, non-MindGeek sites did not see significant changes in other metrics, or, in some cases, the event

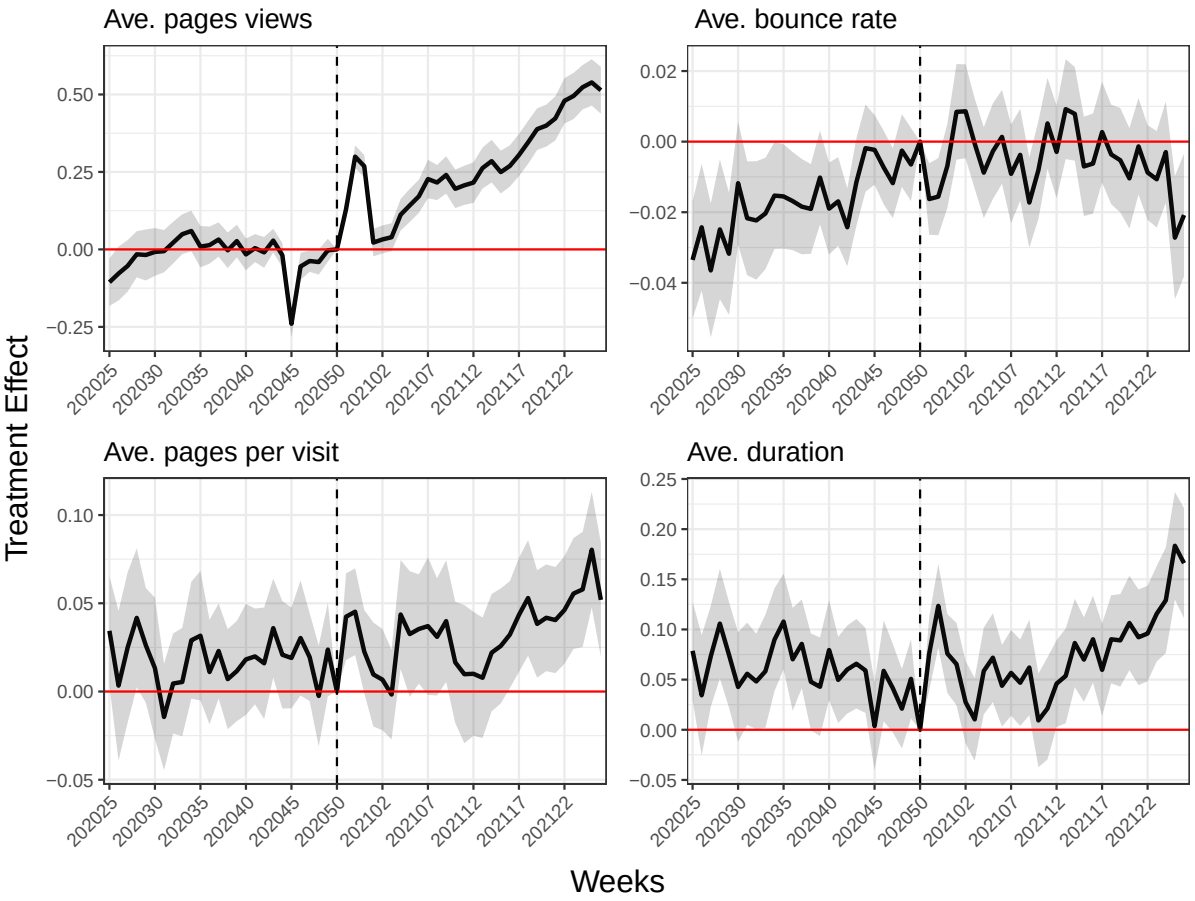


Figure 9: Event study analyses for different measures of engagement for non-MindGeek's domains using daily average measures per week in log scale. The x-axes show the week. The y-axes report the estimated event study coefficients. Each observation is a country-domain observation in a given week. Standard errors are clustered at the domain-country level. The gray area shows 95% confidence intervals. The control group consists of non-adult websites.

study plots cast doubts on the fact that the parallel trends assumption may actually hold.

7. Conclusions

This paper highlights some of the possible unintended consequences of asymmetric content moderation in digital markets, particularly in the context of adult content. Our analysis of MindGeek's removal of non-verified content reveals that, while the action was intended to enhance trust and safety on the platform, it had the side effect of driving a substantial number of users towards unregulated fringe websites and mainstream competitors.

We document a sizeable drop of about 37-51% in visits to MindGeek sites within a month and a steady increase for rival sites, which gained 1.7% traffic per week. Fringe sited had

gained 68% more visits over six months and the top three mainstream competitors of MindGeek approximately 10% each, on average. Our back-of-the-envelope calculations show that MindGeek losses in the eight countries we considered amounted to 25-28 million daily visits and that part of the lost traffic was diverted towards the fringe and mainstream rivals. This results in a gain of at least 2.5-4 million daily visits between fringe sites, which constitute a highly conservative estimate, since we only have information about a small subset of the universe of these sites. Mainstream rivals, on the other hand, saw their traffic increase by 7-24 million visits per day overall. Interestingly, MindGeek content moderation was effective and had a long-lasting effect.

It is interesting to note that while different governments have attempted to regulate this market for many years, it took only five days for two payment giants, Visa and Mastercard, to completely reshape the market and force a dominant platform to clean its video library and transition to a verified system. Substitution to sites that allow anonymous uploads is compelling evidence that anonymity has considerable downsides, as documented in Ederer et al. (2024). On some platforms, like Uber, where harm is mostly between participants on the platform, there is a strong incentive for the platform to insist on transparent identities. However, in cases where harm is to third parties, there may not be enough private incentive to insist on such safeguards.

Our analysis identifies several important aspects. First, it shows that a substantial share of users had a preference for the non-verified content available on MindGeek tubes before the content purge. This may render some regulatory interventions ineffective if similar content is available on non-compliant platforms. Specifically, when traffic is diverted to fringe sites—often located in jurisdictions with less enforcement and serving as havens for harmful and illegal content—a more holistic regulatory approach may be required. The ability of fringe domains to capture this traffic underscores the limitations of asymmetric regulatory interventions, Second, mainstream competitors of MindGeek gained market share, and their growth might have enabled them to surpass the threshold for being designated as VLOPs under the EU DSA, thereby requiring them to undertake content moderation and verification as well. Third, while search frictions exist in these markets, they can be overcome over time due to the roles of aggregators and search engines. Our data show that much of the diverted traffic to fringe sites was facilitated by these platforms, which played a pivotal role in content discovery. Notably, aggregators did most of the work in guiding users toward the less-regulated fringe. This indicates that aggregators, alongside traditional search engines, are crucial in sustaining the demand for non-verified content

and should be monitored accordingly. Fourth, while asymmetric content moderation can level the playing field, a competitive environment in which some platforms remain unregulated may backfire.

Our analysis offers implications for the regulation of online platforms where users can often anonymously post content and materials. The DSA, which imposes content moderation obligations on VLOPs, is an important step toward ensuring online safety. However, our findings indicate that focusing solely on major platforms may leave critical gaps. When users prefer moderated content, they may simply migrate to smaller, unregulated or non-compliant sites that evade these restrictions. In this sense, the DSA that puts in place a framework of layered responsibilities targeted at different types of services (i.e. intermediary services, hosting services, online platform services, and VLOPs) with a set of asymmetric obligations to ensure transparency, accountability, and regulatory oversight of the EU online space is a step in the right direction. However, given the nature of the fringe websites that make up the on-line platform services layer, our research seems to point to the need for enhanced monitoring and obligations for these players if the regulation is to achieve its goals in full. Moreover, MindGeek's intervention was undoubtedly driven by the urgency to regain support from payment systems, leading to the removal of all unverified content. Regulation targeting online platforms may not necessarily produce similar effects if platforms are given specific obligations and sufficient time to comply.

Similar challenges arise with laws imposing age verification requirements on adult sites (e.g., in Texas, Utah, as well as in the EU by the DSA), as smaller or illegal sites operating in other jurisdictions may bypass such rules, especially if search frictions are low. Thus, targeting only large platforms may inadvertently encourage harmful content to spread across less regulated spaces. Addressing these challenges requires regulatory frameworks that extend beyond large platforms (which are more likely to be compliant) to include intermediary actors, particularly search engines and aggregators, which facilitate user migration to fringe sites lacking content verification and legal compliance. Without incorporating these intermediaries into the regulatory scope, content moderation policies may be ineffective, simply shifting harmful material to more permissive environments. Regulators may need to think of various ways, including delisting methods, to control the most important search channels.

References

Abrardi, L., Sabatino, L., and Sapi, G. (2023). Privacy regulations and online safety: Evidence from adult-only websites. *Mimeo*.

Agarwal, S., Ananthakrishnan, U. M., and Tucker, C. E. (2022). Deplatforming and the control of misinformation: Evidence from Parler. *SSRN Working Paper 4232871*.

Aguiar, L., Claussen, J., and Peukert, C. (2018). Catch me if you can: Effectiveness and consequences of online copyright enforcement. *Information Systems Research*, 29(3):656–678.

Andres, R. and Slivko, O. (2021). Combating online hate speech: The impact of legislation on Twitter. *ZEW-Centre for European Economic Research Discussion Paper*, 21-103.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536):1716–1730.

Beknazar-Yuzbashev, G., Jiménez Durán, R., McCrosky, J., and Stalinski, M. (2022). Toxic content and user engagement on social media: Evidence from a field experiment. *SSRN Working Paper 4307346*.

Biglaiser, G., Crémer, J., and Veiga, A. (2022). Should I stay or should I go? Migrating away from an incumbent platform. *RAND Journal of Economics*, 53(3):453–483.

Buiten, M. C., De Streel, A., and Peitz, M. (2020). Rethinking liability rules for online hosting platforms. *International Journal of Law and Information Technology*, 28(2):139–166.

Burdett, K., Shi, S., and Wright, R. (2001). Pricing and matching with frictions. *Journal of Political Economy*, 109(5):1060–1085.

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1):31–72.

Calzada, J., Duch-Brown, N., and Gil, R. (2023). Do search engines increase concentration in media markets? *CESifo Working Paper No. 10671*.

Chan, J. and Ghose, A. (2014). Internet's dirty secret. *MIS Quarterly*, 38(4):955–976.

Congiu, R., Sabatino, L., and Sapi, G. (2022). The impact of privacy regulation on web traffic: Evidence from the GDPR. *Information Economics and Policy*, 61:101003.

Cunningham, S., DeAngelo, G., and Tripp, J. (2024). Did Craigslist's erotic services reduce female homicide and rape? *Journal of Human Resources*, 59(1):280–315.

Danaher, B., Hersh, J. S., Smith, M. D., and Telang, R. (2020). The effect of piracy website blocking on consumer behavior. *MIS Quarterly*, 44:631–659.

Dey, D., Kim, A., and Lahiri, A. (2019). Online piracy and the "longer arm" of enforcement. *Management Science*, 65(3):1173–1190.

Ederer, F., Goldsmith-Pinkham, P., and Jensen, K. (2024). Anonymity and identity online. *arXiv preprint arXiv:2409.15948*.

Farrell, J. and Saloner, G. (1985). Standardization, compatibility and innovation. *RAND Journal of Economics*, 16(1):70–83.

Heckman, J., Ichimura, H., Smith, J., and Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*, 66(5):1017.

Jiménez Durán, R., Müller, K., and Schwarz, C. (2022). The effect of content moderation on online and offline hate: Evidence from Germany's NetzDG. *SSRN Working Paper 4230296*.

Kumarswamy, N., Singhal, M., and Nilizadeh, S. (2023). Impact of stricter content moderation on Parler's users' discourse. *arXiv preprint arXiv:2310.08844*.

Lefouili, Y. and Madio, L. (2022). The economics of platform liability. *European Journal of Law and Economics*, 53(3):319–351.

Lu, S., Wang, X., and Bendle, N. (2020). Does piracy create online word of mouth? An empirical analysis in the movie industry. *Management Science*, 66(5):2140–2162.

Madio, L. and Quinn, M. (2024). Content moderation and advertising in social media platforms. *Journal of Economics & Management Strategy*. Forthcoming.

Moen, E. R. (1997). Competitive search equilibrium. *Journal of Political Economy*, 105(2):385–411.

Müller, K. and Schwarz, C. (2022). The effects of online content moderation: Evidence from president Trump's account deletion. *SSRN Working Paper 4296306*.

Peukert, C., Claussen, J., and Kretschmer, T. (2017). Piracy and box office movie revenues: Evidence from Megaupload. *International Journal of Industrial Organization*, 52:188–215.

Rendo, I. (2024). Content moderation in presence of fringe platforms. *Mimeo*.

Rizzi, M. (2024). Self-regulation of social media and the evolution of content: a cross-platform analysis. *Mimeo*.

Sabatino, L. and Sapi, G. (2023). Privacy regulation and online concentration during demand peaks: Evidence from the e-commerce sector. *Journal of Industrial and Business Economics*, 50(2):265–282.

Schmitt, J., Miller, K. M., and Skiera, B. (2020). The impact of privacy laws on online user behavior. *Mimeo*.

Van Loo, R. and Spier, K. E. (2024). Foundations for platform liability. *Notre Dame Law Review*. Forthcoming.

Wlömert, N., Papies, D., Clement, M., and Spann, M. (2023). The interplay of user-generated content, content industry revenues, and platform regulation: Quasi-experimental evidence from Youtube. *Marketing Science*, 43(1):1–12.

Zeng, H. S., Danaher, B., and Smith, M. D. (2022). Internet governance through site shutdowns: The impact of shutting down two major commercial sex advertising sites. *Management Science*, 68(11).

Appendices

A. Theoretical background

Our setting studies substitution across platforms. Perhaps the closest theoretical work is the migration model of the platform of Biglaiser et al. (2022). In that paper, consumers value scale of the platform and have limited opportunities to switch; switching is based on expected future scale. One can interpret the quality of the MindGeek's platforms to have fallen, in the language of Biglaiser et al. (2022), because of the content moderation. This causes a gradual switch in Biglaiser et al. (2022) as consumers in that model can only contemplate switching periodically.

However, an important difference is that visits to the MindGeek platform drop suddenly, while visits to the rest of the platforms are gradual. In Biglaiser et al. (2022), these two are mirror images. We imagine that this is generated by search; as in many online applications, the content is vast and presumably heterogeneous, so once a consumer decides a platform does not have enough content, it does not stay. This section derives some implications of search, in order to understand the forces that might underlie the measured changes in traffic and referrals that we document.

The basic notation of the model follows Biglaiser et al. (2022). We treat, as they do, one platform as the one that is being migrated "from" in response to the change in content (in their case, they call this the incumbent, but here we model migration away from MindGeek.) Like they do, we model the rest of the platforms as one; it is straightforward to consider more heterogeneous platforms that consumers switch into, but the basic model gets across the key forces we imagine.

Let the number of users (which we take as measuring visits in our data) at time t be $h_i(t)$ for $i \in \{MG, N\}$ for MindGeek (MG) and Non-MindGeek (N). We note three stylized facts about this in the data, mapping h to views. First, we take the values of $h_i(t)$ to be roughly constant before the content moderation date τ. At τ, h_{MG} falls rapidly and then stabilizes, while h_N rises even after h_{MG} stabilizes. We will show how this arises in a model in the spirit of Biglaiser et al. (2022), but with search.

In addition to search, a difference between this model and Biglaiser et al. (2022) is that we do not take migration as impacting the value of consumers directly, since many platforms have

uploads and views from different users.[40] Moreover, these changes appear to be mostly one-time in the data: content does not increase substantially at the non-MG even though views do, while MG content is permanently lower (but not shrinking after content moderation). Let the content on the platform i at time t be $c_i(t)$. One can interpret c_i as, for the representative consumer, the amount of content that fits their tastes at a given platform. Consistent with the data on the pages, we take content as falling discontinuously on MG at the date of content moderation at date τ, which we denote B for $t < \tau$ and A for $t > \tau$, so that content is summarized by $c^A_{MG} < c^B_{MG}$ and c_N, which is constant.[41]

Users are separated from their current platform choice at the rate $\mu_i(c(t))$, which is increasing and can in turn be summarized by μ^B_{MG}, μ^A_{MG} and μ_N. One can interpret this as going to the site and not finding the content the user is looking for, either because the content has changed or their tastes have changed. Since content is reduced after moderation, $\mu^A_{MG} > \mu^B_{MG}$: users do not find the content they are looking for after the content moderation begin to search.[42] We assume that when a consumer considers switching, they do not necessarily find the content they are looking for. Instead, they enter a search pool from which they attempt to find content that fits their tastes. Formally, separations lead the user to enter a search pool of size $S(t)$.[43]

Consumers in the search pool find content at site i at rate $\lambda_i(c_i(t))$, which is increasing and can be summarized by $\lambda^B_{MG}, \lambda^A_{MG}$ and λ_N.[44] If they find content that fits their taste, they become viewers of that platform (until they separate). We allow users to switch any number of times. The evolution of users (and, therefore, views), where dots over variables represent time derivatives, is

$$\dot{h}_i(t) = S(t)\lambda_i(c_i(t)) - h_i(t)\mu(c_i(t))$$

so, in percent,

$$\frac{\dot{h}_i(t)}{h_i(t)} = \frac{S(t)}{h_i(t)}\lambda_i(c_i(t)) - \mu(c_i(t))$$

[40] Biglaiser et al. (2022) envision a platform such as social networks, where posters and viewers are the same pool.

[41] This treatment of search is consistent with a reduced form of directed search as in Moen (1997) and Burdett et al. (2001), but with content playing the role of wages in driving matching rate.

[42] In Biglaiser et al. (2022), migration is driven by a utility calculation, but it is easy to see how such a calculation would map to this environment, since less content means that one should migrate.

[43] One could model, separately, within platform search versus across platform search, but it is not necessary to explain the basic facts.

[44] One could imagine that λ_i depends on the entire vector of content, since limited space at search engines and aggregators might be impacted by total content. But this is also not necessary to explain the main facts.

and for the search pool

$$\begin{aligned}\dot{S}(t) &= \sum_i \left(h_i(t)\mu_i(c_i(t)) - \lambda_i(c_i(t))S(t)\right) \\ &= -\sum_i \dot{h}_i(t)\end{aligned}$$

Since both platforms are taken to be stable before τ,

$$\begin{aligned}\bar{h}^B_{MG} &= S^B\lambda^B_{MG}/\mu^B_{MG} \\ \bar{h}^B_N &= S^B\lambda_N/\mu_N\end{aligned}$$

Normalize $S^B = 1$ to fix market size. Then we can compute

$$\frac{\dot{h}_i(t)}{h(t)}|_{t=\tau} = \mu^B_{MG}\frac{\lambda^A_{MG}}{\lambda^B_{MG}} - \mu^A_{MG} < 0$$

since $\mu^B_{MG} < \mu^A_{MG}$ and $\frac{\lambda^A_{MG}}{\lambda^B_{MG}} < 1$, i.e., MG shrinks at t. MG reaches its new steady state level quickly; note that at the new stable level for MG:

$$\bar{h}^A_{MG} = S^A\lambda^A_{MG}/\mu^A_{MG}$$

Although this is below the old level, since $\lambda^A_{MG}/\mu^A_{MG} < \lambda^B_{MG}/\mu^B_{MG}$, it could be that S^A is below or above $S^B = 1$. Immediately after content moderation, $\dot{h}_{MG} < 0$ but $\dot{h}_N = 0$ since nothing has changed immediately in their evolution, so $\dot{S} = -\left(\dot{h}_{MG} + \dot{h}_N\right) > 0$. If, when MG stabilizes, $S^A > S^B = 1$, then at this new level of

$$\frac{\dot{h}_N(t)}{h_N(t)} = \frac{S^A}{h_N(t)}\lambda_N - \mu_N$$

Since the new search pool is larger, views on non-MG sites grow until $S^A/h^A_N = 1/h^B_N$. Non-MG sites grow even after MG stabilizes. Since the search pool grows, referrals from search also grow for non-MG sites. For MG it is ambiguous: the search pool is larger but $\lambda^A_{MG} < \lambda^B_{MG}$.[45] The substantial drop in referrals in the data for MG suggests a large reduction in λ_{MG}.

[45] Alternatively one could assume that h_{MG} is reduced suddenly at the time of content moderation but that, for remaining viewers, μ_{MG} is unchanged. If so, then the reduction in $\bar{h}^A_{MG}$ relative $\bar{h}^B_{MG}$ to implies that $S^A\lambda^A_{MG}$ is lower than $S^B\lambda^B_{MG}$, i.e., referrals from the search pool are lower than before. Moreover the drop in referrals to MG, given that there has been a sudden increase in search, suggests a substantial decrease in λ_{MG}.

In this environment, one can consider what alternative interpretations of content moderation might be. For instance, suppose that the content moderation simply moved, suddenly, a group of users into the search pool, but did not change the underlying λ_{MG} and μ_{MG}. This would be consistent with the content reduction being simply random, and not targeted at a specific type of content. Then the greater search pool would make MindGeek and non-MindGeek sites grow (after the initial loss for MindGeek). Since we did not observe this growth at MindGeek, it seems clear that the MindGeek policy was not a random removal of content.

B. Additional Descriptive Statistics

In this section, we provide additional descriptive statistics about non-MindGeek domains, distinguishing between the top 3 largest mainstream tubes (e.g., xvideos.com, xHamster.com, xnxx.com) and the fringe tubes. In particular, our dataset comprises 19 fringe sites. Since we focus on 8 countries and all sites are available and visited in all of them, the total of country-domains amounts to 152. Similarly, the number of mainstream country-domains is 24.

Table B.1 shows that both fringe and mainstream sites saw an increase in the average number of daily visits after content moderation. Specifically, fringe sites experienced an increase from 88 thousand to 112 thousand visits, while the top 3 domains rose from approximately 3.3 million to 3.5 million visits. Before the shock, visits to each fringe site were on average approximately 2. 7% of visits to a mainstream site, but increased to approximately 3. 2% after MindGeek content removal.

	MindGeek		Non-MindGeek			
			Fringe		Top 3	
Metrics	Before	After	Before	After	Before	After
Visits	2,131,211	1,607,478	88,091	112,221	3,323,777	3,447,942
Pages views	16,691,923	12,840,403	652,893	835,465	33,831,160	34,854,432
Bounce rate	0.39	0.33	0.44	0.44	0.20	0.20
Pages per visit	6.32	7.15	6.23	6.30	12.44	12.40
Duration	423.68	452.75	316.89	319.61	778.85	773.24
Total time (×100)	10,363,028	7,378,330	329,408	421,706	26,921,596	27,308,495
N. domains	4		19		3	
N. country-domains	32		152		24	

Table B.1: Average daily statistics per country-domains aggregated weekly. The table presents descriptive statistics for adult tube websites, distinguishing MindGeek's ones and non-MindGeek ones fringe and top 3 mainstream domains (i.e., xvideos.com, xhamster.com, xnxx.com). The sample period spans from July 1st, 2020, to June 20th, 2021. *Before* and *After* refer to the periods surrounding MindGeek's content removal on December 14th, 2020. Total time and duration are expressed in seconds.

Page views also increased, while the bounce rate remained constant for both types of domains. In particular, fringe sites have a much higher bounce rate than mainstream ones. Pages per visit and duration increased slightly for fringe sites, while they decreased slightly for mainstream sites.

C. Analysis in Levels

We also present evidence based on the average number of daily visits in levels, rather than logarithms. Table C.1 reports the summary descriptives.

Figure C.1 presents model-free evidence of the shock on the average number of visits for MindGeek tubes and non-MindGeek tubes (left panel) and the four largest sites (right panel).

Group	Mean	Median	S.D.	Min	Max
Non-Adult Control	1,437,717	9,220	17,512,114	38	685,061,969
MindGeek	658,208	16,980	3,084,072	96	33,702,182
Non-MindGeek	347,791	28,533	1,670,619	195	26,494,066

Table C.1: Additional descriptive statistics on free adult tubes (MindGeek and non-MindGeek tubes) and non-adult websites. Average daily visits per country-domain, aggregated at the weekly level. The sample period spans from July 1st, 2020, to June 20th, 2021.

The event study graphs relating to the estimated specifications in the levels are in Figure C.2, illustrating the impact of content purge on the average number of daily visits on MindGeek and non-MindGeek domains *vis-à-vis* the control group, Figure C.3, focusing on the heterogenous impact on the fringe and the non-MindGeek leading domains. Finally, Figure C.4 illustrates the results of the same analysis performed through MC-NN. Table C.2 and C.3 presents our main results from the difference-in-differences approach using levels rather than the log scale.

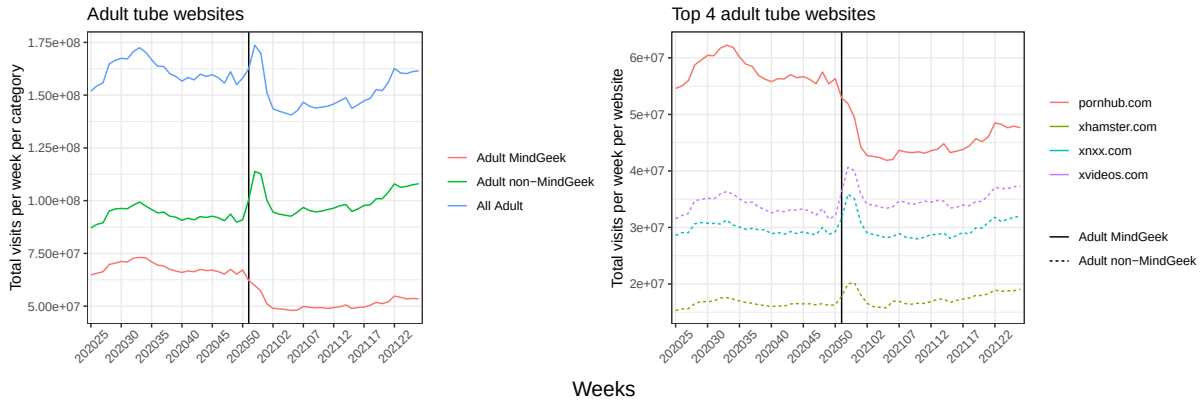


Figure C.1: Visits per week in levels scale between July 1st, 2020 and June 20th, 2021. The left panel plots the number of the weekly sum of visits across countries and domains per group: adult domains (light blue), MindGeek domains (red) and non-MindGeek domains (green). Among the adult websites, we only focus on free tubes. The right panel plots the number of the weekly sum of visits across countries for the four largest websites. The vertical dark line indicates the week in which MindGeek removed all non-verified content.

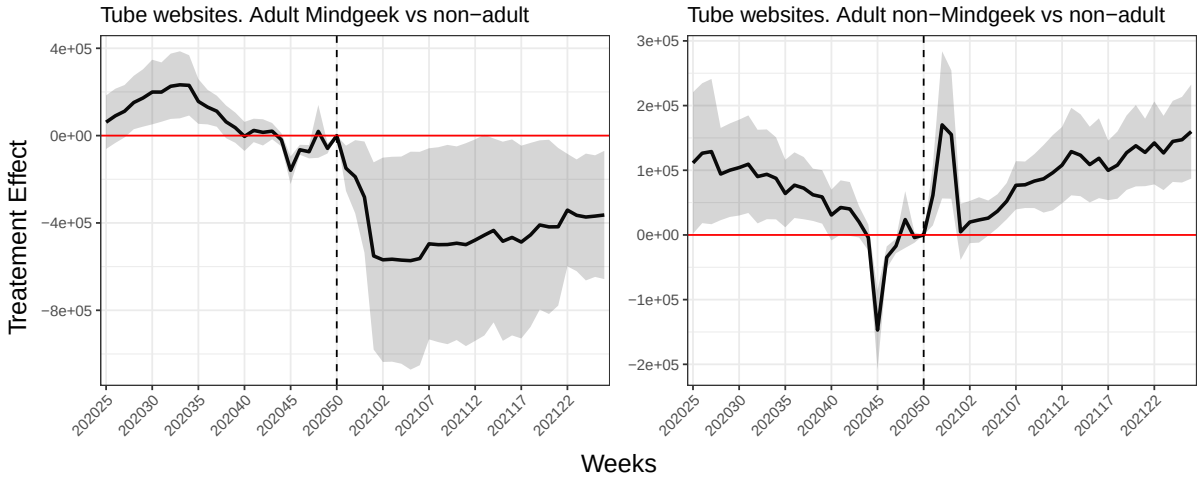


Figure C.2: Event study analyses for the daily average number of weekly visits. The x-axes show the week. The y-axes report the estimated event study coefficients. Each observation is a country-domain observation in a given week. Standard errors are clustered at the country-domain level. The gray area shows 95% confidence intervals. The left panel plots the average treatment effect of domains belonging to the MindGeek group compared non-adult domains. The right panel plots the average treatment effect of domains not belonging to the MindGeek group compared non-adult domains. Among the adult websites, we only focus on free tubes.

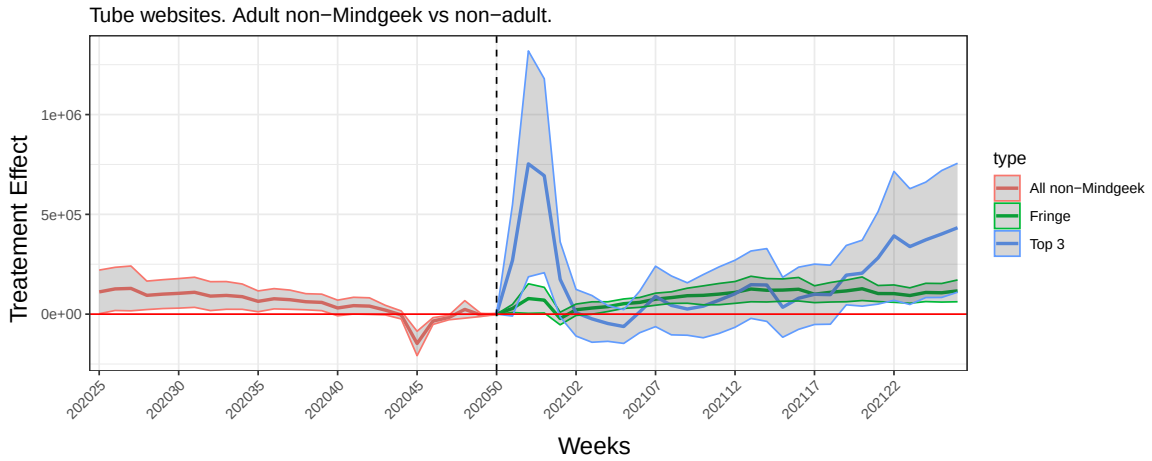


Figure C.3: Event study analysis for the daily average number of weekly visits of non-MindGeek domains distinguishing between mainstream (top 3) domains and fringe domains. The x-axes show the week. The y-axes report the estimated event study coefficients. Each observation is a country-domain observation in a given week. Standard errors are clustered at the country-domain level. The gray area shows 95% confidence intervals.

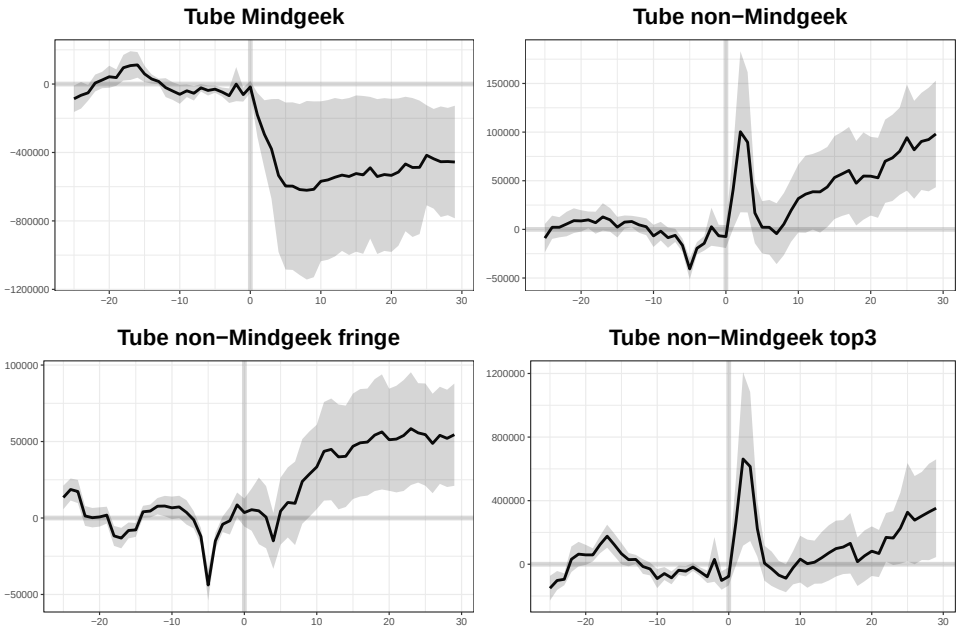


Figure C.4: Event study analyses for the daily average number of weekly visits using the MC-NN method. The x-axes show the week. The y-axes report the estimated event study coefficients. Each observation is a country-domain observation in a given week. Standard errors are clustered at the country-domain level. The gray area shows 95% confidence intervals. The top-left panel plots the average treatment effect of domains belonging to the MindGeek group compared non-adult domains. The top-right panel plots the average treatment effect of domains not belonging to the MindGeek group compared non-adult domains. The bottom-left panel plots the average treatment effect of non-MindGeek fringe domains compared non-adult domains. The bottom-right panel plots the average treatment effect of non-MindGeek mainstream (top 3) domains compared non-adult domains. Among the adult websites, we only focus on free tubes.

	Dependent variable:				
	Avg Daily Visits per Country-domain				
	(1)	(2)	(3)	(4)	(5)
After × Adult	−38,454	−3,814			
	(39,442)	(38,909)			
After × Adult MindGeek			−513,890**	−415,522**	−700,404**
			(207,123)	(205,008)	(282,373)
After × Adult non-MindGeek			48,263**	71,110***	−49,498**
			(19,873)	(20,465)	(19,476)
Year-week FE	✓	✓	✓	✓	✓
Domain FE	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓
Country-domain FE	✓	✓	✓	✓	✓
Domain Trend		✓		✓	✓
Holidays Included	✓	✓	✓	✓	
Observations	107,111	107,111	107,111	107,111	85,595

Table C.2: Estimates of the traffic diversion between non-MindGeek and MindGeek's domains under different model specifications using levels. Columns (1) to (4) consider the whole sample period, and Column (5) excludes weeks 42 to 53 of the year 2020. Among the adult websites, we only focus on free tubes. Standard errors clustered at the domain-country level. Note: *p<0.1; **p<0.05; ***p<0.01.

	Dependent variable:				
	Avg Daily Visits per Domain-Country				
	(1)	(2)	(3)	(4)	(5)
After × Adult MindGeek	−531,722.3**	−392,886.6*	−531,722.3**	−392,886.6*	−636,566.6**
	(215,213.200)	(202,249.700)	(215,215.200)	(202,251.600)	(257,998.700)
After × Adult non-MindGeek	1,749.303	90,631.86***	85,071.98	238,156.9***	−105,855.2*
	(17,827.270)	(23,938.050)	(86,127.570)	(87,376.960)	(58,316.630)
After × Adult MindGeek × t	1,262.02	11,585.52***	1,262.02	11,585.53***	15,758.45**
	(1,060.669)	(3,251.032)	(1,060.679)	(3,251.063)	(6,978.956)
After × Adult non-MindGeek × t	3,310.689***	9,949.475***	3,483.666*	14,862.67***	22,634.89***
	(611.362)	(2,395.654)	(1,971.688)	(3,356.921)	(6,928.979)
After × Adult non-MindGeek fringe			−96,528.86	−170,862.8**	114,992.5**
			(85,200.990)	(85,021.580)	(57,108.920)
After × Adult non-MindGeek fringe × t			−200.29	−5,692.719**	−14,039.5**
			(1,924.812)	(2,401.287)	(6,366.048)
Year-week FE	✓	✓	✓	✓	✓
Domain FE	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓
Country-domain FE	✓	✓	✓	✓	✓
Domain Trend		✓		✓	✓
Holidays Included	✓	✓	✓	✓	
Observations	107,111	107,111	107,111	107,111	85,595

Table C.3: Estimates of the traffic diversion to non-MindGeek mainstream (top 3) and fringe websites using levels. Columns (1) to (4) consider the whole sample period, and Column (5) excludes weeks 42 to 53 of the year 2020. Among the adult websites, we only focus on free tubes. Standard errors clustered at the country-domain level. Note: *p<0.1; **p<0.05; ***p<0.01.

D. Robustness Checks: PSM and MC-NN

D.1. PSM and MC-NN

Our baseline specification assumes that non-adult websites are a valid control group, and there may be systematic differences between adult and non-adult websites that could bias the results. To address this concern, we also employ Propensity Score Matching (PSM) to construct a potentially more suitable control group before estimating (1). PSM is a statistical technique that matches treated and control units based on observable characteristics, thereby reducing selection bias (Heckman et al., 1998; Caliendo and Kopeinig, 2008).

In our case, we match adult websites (both MindGeek and non-MindGeek) with non-adult websites that are similar in terms of pre-treatment characteristics such as website traffic, user demographics, and content type. The propensity score is estimated using a logistic regression model, where the dependent variable is a binary indicator for being an adult website. The matched non-adult websites serve as a control group to assess the robustness of our DiD analysis. Indeed, by using PSM, we mitigate concerns that unobserved heterogeneity between adult and non-adult websites may drive our results, ensuring that the estimated treatment effects more accurately reflect the causal impact of the regulatory intervention. The descriptive statistics of the resulting matching sample are provided in Table D.1.

	Adult Tubes Only		All Non-adult		Matched Non-adult Sample	
Metrics	Before	After	Before	After	Before	After
Visits	778,575	727,151	1,449,946	1,426,855	1,086,728	1,080,717
Pages views	6,974,414	6,607,644	11,494,923	11,633,863	10,240,787	9,939,111
Bounce rate	0.41	0.40	0.55	0.55	0.41	0.41
Pages per visit	6.96	7.13	4.09	4.11	6.22	6.25
Duration	386.91	392.43	275.02	275.58	366.54	363.02
Total time (×100)	4,960,210	4,594,278	8,311,461	8,518,290	6,612,417	6,314,381
N. domains	26	26	251	251	115	115
N. country-domains	208	208	1,834	1,834	208	208

Table D.1: Average daily statistics per country-domain aggregated at the week level. Non-adult websites are matched via PSM using only adult tubes at the domain-country level. Total time and duration are expressed in seconds.

To further strengthen our identification strategy, we also apply the Matrix Completion - Nuclear Norm (MC-NN) method, which addresses missing data and selection issues in high-dimensional settings. MC-NN hinges upon Machine Learning techniques to predict the potential outcomes of treated units, and it is particularly useful in cases where there is unobserved heterogeneity

or selection bias that is difficult to account for using traditional matching techniques. This method relies on the assumption that the observed data matrix (which contains website traffic data across different domains, countries, and weeks) is low-rank, meaning that the underlying relationships between the variables can be captured with fewer dimensions. Matrix completion is implemented using a nuclear norm minimization approach, which recovers the missing entries in the data matrix by solving an optimization problem (Athey et al., 2021).

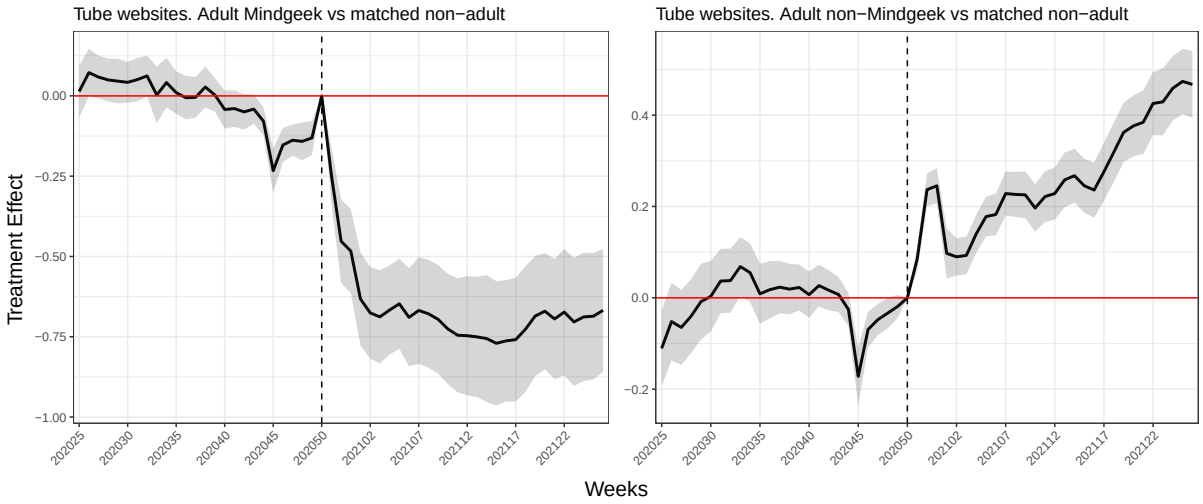


Figure D.1: Event study analyses (matching via PSM) for the logarithm of the daily average number of weekly visits. The x-axes show the week. The y-axes report the estimated event study coefficients. Each observation is a country-domain observation in a given week. Standard errors are clustered at the country-domain level. The gray area shows 95% confidence intervals. The left panel plots the average treatment effect of domains belonging to the MindGeek group compared to non-adult domains identified via PSM. The right panel plots the average treatment effect of domains not belonging to the MindGeek group compared to non-adult domains identified via PSM. Among the adult websites, we only focus on free tubes.

The results of both DiD-PSM, in Table D.2, Panel A, and Figure D.1, and MC-NN, in Table D.2, Panel B and Figure D.2, confirm the robustness of the findings presented in the main text. The same also applies to our results on the heterogeneity of the effects for the mainstream non-MindGeek sites *vis-à-vis* the fringe, see Table D.3.

	Dependent variable:				
	log (Avg Daily Visits per Website-Country)				
Panel A. DID-PSM Estimates	(1)	(2)	(3)	(4)	(5)
After × Adult	0.133***	0.003			
	(0.042)	(0.024)			
After × Adult MindGeek			−0.648***	−0.443***	−0.673***
			(0.067)	(0.050)	(0.055)
After × Adult non-MindGeek			0.275***	0.084***	−0.004
			(0.038)	(0.020)	(0.034)
Year-week FE	✓	✓	✓	✓	
Domain FE	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓
Country-domain FE	✓	✓	✓	✓	✓
Domain Trend		✓		✓	✓
Holidays Included	✓	✓	✓	✓	
Observations	22,617	22,617	22,617	22,617	18,087
Panel B. MC-NN Estimates	(1)	(2)	(3)	(4)	
After × Adult	0.136***		0.118***		
	(0.040)		(0.046)		
After × Adult MindGeek		−0.644***		−0.72***	
		(0.066)		(0.064)	
After × Adult non-MindGeek		0.278***		0.271***	
		(0.036)		(0.043)	
Year-week FE	✓	✓	✓	✓	
Country-domain FE	✓	✓	✓	✓	
Holidays Included	✓	✓			
Observations	107,111	107,111	85,595	85,595	

Table D.2: Estimates of the traffic diversion between MindGeek and non-MindGeek websites. Panel A shows results from equation (1) using Propensity Score Matching for the control group. Panel B uses Matrix Completion - Nuclear Norm. For panel A, Columns (1) to (4) consider the whole sample period, and Column (5) excludes weeks 42 to 53 of the year 2020. For panel B, Columns (1) and (2) consider the whole sample period, and Columns (3) and (4) exclude weeks 42 to 53 of the year 2020. Among the adult websites, we only focus on free tubes. Standard errors clustered at the country-domain level. Note: *p<0.1; **p<0.05; ***p<0.01.

	Dependent variable:				
	log (Avg Daily Visits per Website-Country)				
Panel A. DID-PSM Estimates	(1)	(2)	(3)	(4)	(5)
After × Adult MindGeek	−0.553***	−0.440***	−0.553***	−0.440***	−0.654***
	(0.047)	(0.054)	(0.047)	(0.054)	(0.060)
After × Adult non-MindGeek	0.102***	0.109***	−0.055***	0.085***	−0.088***
	(0.029)	(0.020)	(0.017)	(0.016)	(0.023)
After × Adult MindGeek × t	−0.007***	0.002	−0.007***	0.002	0.005*
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)
After × Adult non-MindGeek × t	0.012***	0.013***	0.004***	0.014***	0.013***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)
After × Adult non-MindGeek × fringe			0.181***	0.028	0.142***
			(0.032)	(0.021)	(0.036)
After × Adult non-MindGeek × fringe × t			0.010***	−0.002	−0.004*
			(0.001)	(0.002)	(0.002)
Year-week FE	✓	✓	✓	✓	
Domain FE	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓
Country-domain FE	✓	✓	✓	✓	✓
Domain Trend		✓		✓	✓
Holidays Included	✓	✓	✓	✓	
Observations	22,617	22,617	22,617	22,617	18,087
Panel B. MC-NN Estimates	(1)	(2)	(3)	(4)	(5)
After × Adult MindGeek	−0.66***	-0.64***	-0.64***	-0.72***	-0.72***
	(0.057)	(0.068)	(0.063)	(0.066)	(0.067)
After × Adult non-MindGeek	0.07***	0.003	-0.096***	-0.058***	-0.127***
	(0.03)	(0.018)	(0.015)	(0.022)	(0.019)
After × Adult MindGeek × t	0.0002				
	(0.0008)				
After × Adult non-MindGeek × t	0.017***		0.008***		0.008***
	(0.0007)		(0.0006)		(0.0006)
After × Adult non-MindGeek × fringe		0.10***	0.32***	0.079***	0.31***
		(0.032)	(0.039)	(0.036)	(0.047)
After × Adult non-MindGeek × fringe × t		0.018***		0.018***	
		(0.0007)		(0.0007)	
Year-week FE	✓	✓	✓	✓	✓
Country-domain FE	✓	✓	✓	✓	✓
Holidays Included	✓	✓	✓		
Observations	107,111	107,111	107,111	85,595	85,595

Table D.3: Estimates of the traffic diversion between MindGeek and non-MindGeek websites. Panel A shows results from equation (1) using Propensity Score Matching for the control group. Panel B uses Matrix Completion - Nuclear Norm. Columns (1) to (3) consider the whole sample period and report either the overall effect in the after period or the short-term effect (computed at exactly 5 weeks after the shock to avoid holidays) and associated trends (from the 5th weeks after the shock to the last week of the sample). Columns (4) and (5) exclude weeks 42 to 53 of 2020. Among the adult websites, we only focus on free tubes. The variable *Adult non-MindGeek* captures the reference category, which is the top 3 Adult non-MindGeek domains. Standard errors clustered at the country-domain level. Note: *p<0.1; **p<0.05; ***p<0.01.

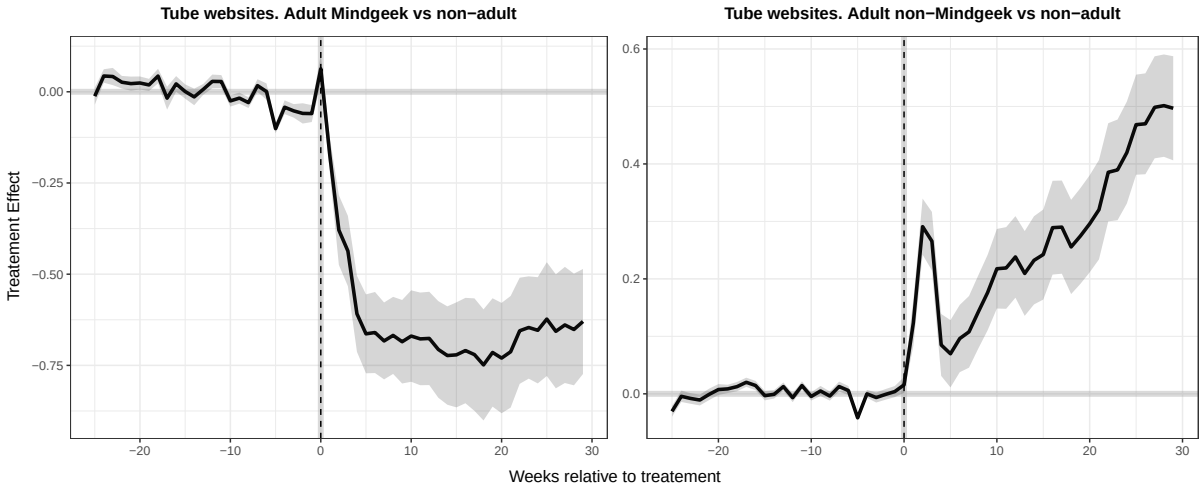


Figure D.2: Event study analyses (using the MC-NN method) for the logarithm of the daily average number of weekly visits. The x-axes show the week. The y-axes report the estimated event study coefficients. Each observation is a country-domain observation in a given week. Standard errors are clustered at the domain-country level. The gray area shows 95% confidence intervals. The left panel plots the average treatment effect of domains belonging to the MindGeek group compared to non-adult domains. The right panel plots the average treatment effect of domains not belonging to the MindGeek group compared to non-adult domains. Among the adult websites, we only focus on free tubes.

D.2. Visit-weighted MC-NN

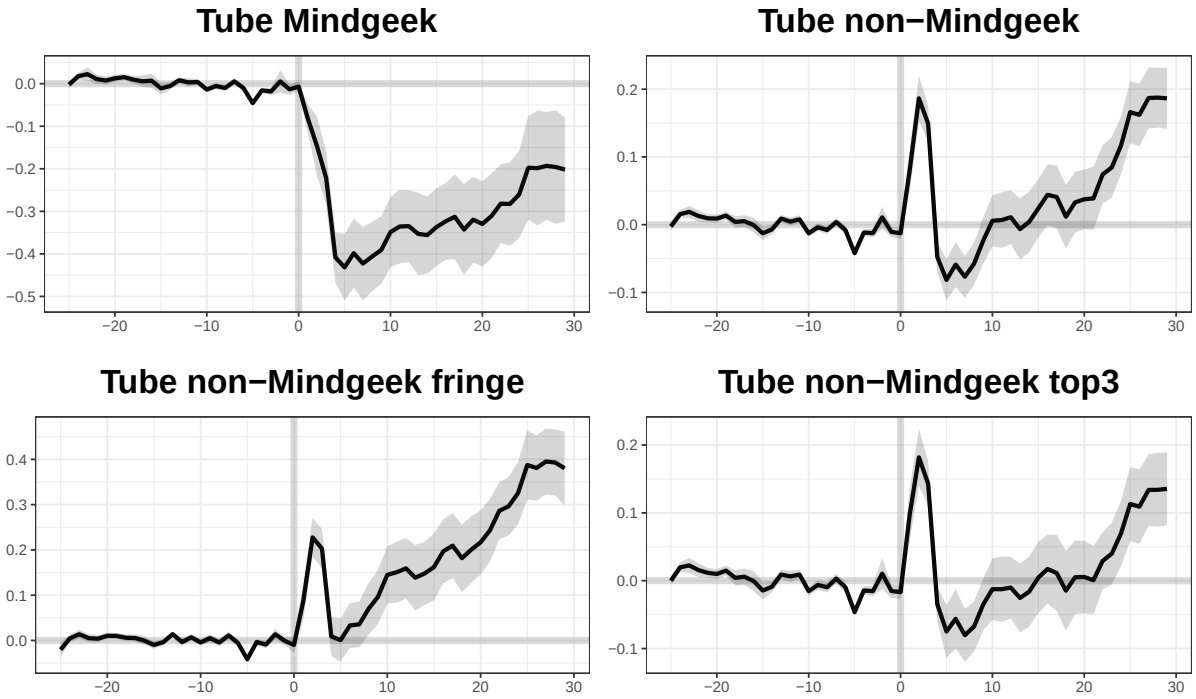


Figure D.3: Event study analyses (using the MC-NN method) for the logarithm of the daily average number of weekly visits, weighted by the average daily number of weekly visits before the shock for each country-domain. The x-axes show the week. The y-axes report the estimated event study coefficients. Each observation is a country-domain observation in a given week. Standard errors are clustered at the domain-country level. The gray area shows 95% confidence intervals. The top-left panel plots the average treatment effect of domains belonging to the MindGeek group compared to non-adult domains. The top-right panel plots the average treatment effect of domains not belonging to the MindGeek group compared to non-adult domains. The bottom-left panel plots the average treatment effect of non-MindGeek fringe domains compared to non-adult domains. The bottom-right panel plots the average treatment effect of non-MindGeek mainstream (top 3) domains compared to non-adult domains. Among the adult websites, we only focus on free tubes.

E. Back of the Envelope Calculations

In the following analysis, we estimate the losses incurred by MindGeek and the gains experienced by both fringe and mainstream competitors in terms of visits.

It is important to note that our estimates for the overall effect on fringe sites represent a lower bound, as our data includes only 19 fringe domains, while the actual number of fringe domains is significantly larger. Although smaller and smaller, these sites can jointly sum to a sizeable share of visits. Therefore, our estimates of the traffic gains for fringe sites are conservative and should not be directly compared with the traffic gains observed for mainstream sites.

Recall that in our sample, we have 152 country-domains for fringe sites, 24 country-domains for non-MindGeek mainstream sites, and 32 country-domains belonging to MindGeek.

From Table B.1, the average number of visits per country-domain is 88,000 for fringe sites and 3,322,000 for other non-MindGeek sites. The average number of visits per country-domain for MindGeek sites is 2,100,000.

Back-of-the-envelope without holidays. We begin with a back-of-the-envelope calculation using coefficients estimated in Column (3) of Table 4, which excludes holidays and, therefore, computes the effect at week 4, i.e., the first week of 2021. MindGeek experiences a reduction of 51% in traffic relative to our control group of non-adult sites. On the contrary, fringe sites experience no effect (i.e., $\exp(-0.119 + 0.142) - 1 \approx 2.3\%$ which is not significantly different from zero), and the top 3 mainstream sites experience a loss of approximately 11.2% relative to our control group of non-adult sites. When multiplied by the average number of visits before the shock, this implies a gain of 2,000 visits per country-domain for fringe sites, a loss of 395,000 visits for the top 3 sites, and a loss of 1.04 million visits for each MindGeek country-domain. Since there are 32 country-domains for MindGeek, the immediate losses amounted at approximately 34.7 million visits per day in the 8 countries in our dataset.

Next, we compute the weekly change in visits per country-domain. Here, MindGeek sites grow by 0.8% per week, corresponding to an increase of 17,000 visits per country-domain, while both fringe and mainstream sites experience an increase of 1.7%. This translates to an increase of 1,500 visits per week per fringe country-domain ($1.7\% \times 88{,}000$) and 56,600 visits per week per mainstream country-domain ($1.7\% \times 3{,}322{,}000$).

We can now calculate the effect at different time periods.

Week 5 after the shock ($t = 5$). Let us consider week 5 after the shock. This has the advantage to allows us to avoid conflating the effects with the holiday surge during the Christmas period. The effect at week 5 is calculated by summing the immediate effect, which in this case is at week 4, and the weekly growth, then multiplying it by the total number of domains within each category.

- *MindGeek.* The losses at week 0-4 are 34.7 million and MindGeek recovers 0.5 million visits per week. Thus, after 5 weeks of content moderation (that is, $t \times 1$), losses amount to 34.2 millions.
- *Fringe.* There is no immediate effect for the fringe. Since there are 214,000 more visits per week, after 5 weeks from the content moderation there are 0.2 visits overall.
- *Mainstream (top 3).* There are losses amounting at 7 million visits at week 0-4. Since there are gains by 1.3 million visits per week, after 5 week from the content moderation there is a 1.3 visits overall.

Week 15 after the shock ($t = 15$). The effect at week 15 is calculated by summing the immediate effect and weekly growth for fourteen weeks, then multiplying it by the total number of domains within each category.

- *MindGeek.* At week 15, losses amount at 28.7 million visits.
- *Fringe.* At week 15, gains are 0.22 per week amounting to a total of 2.5 million visits
- *Mainstream (top 3).* At week 15, gains amount at 7.1 million visits.

Back-of-the-envelope with holidays. Following the same procedure, we calculate a back-of-the-envelope calculation using the coefficients estimated in Column (2) of Table 4, which includes holidays. MindGeek experienced a reduction in traffic of 36.8%, amounting to approximately 25.1 million visits across all 32 country domains. Mainstream rivals and fringe sites immediately increased their traffic by 6.6%, gaining approximately 5.2 million and 0.88 million visits, respectively.

The weekly trend is not significantly different from zero for MindGeek, indicating that losses are persistent over time. Conversely, non-MindGeek sites see a 1.6% increase per week, amounting

to 214,000 visits per week for fringe sites and 1.3 million visits for the top 3 mainstream sites.

Week 5 after the shock (t=5). Since we include holidays, the effect at week 5 is calculated by summing the immediate effect and the weekly growth for five weeks only, then multiplying it by the total number of domains within each category.

- *MindGeek.* The losses at week 53 of 2020 (week 3 in the post-period) are 25.1 million. Since the weekly effect is not significantly different from zero, the drop is persistent.
- *Fringe.* The gains at week 0 are 0.8 millions Since there is a gain of 214 thousands visits more per week, at week 5 the overall gain is by 1.3 million visits.
- *Mainstream (top 3).* The gains at 0 are 5.3 million. Since there is a gain of approximately 2 million thousands visits more per week, at week 5 the overall gain is by 11.6 million visits.

Week 15 after the shock (t=15). Since we include holidays the effect at week 15 is calculated by summing the immediate effect and the weekly growth for fifhteen weeks , then multiplying it by the total number of domains within each category.

- *MindGeek.* Losses at week 0 are 25.1 million. Since the weekly effect is not significantly different from zero, the drop is persistent.
- *Fringe.* The gains at week 0 are 0.8 million. Since there is a gain of 214 thousands visits more per week (0.214$\times$ 15), the overall effect at week 15 is given by 4.1 million visits.
- *Mainstream (top 3).* The gains at week 0 are 5.3 million. Since there is a gain of 1.3 million thousands visits more per week, at week 15 the overall gain is by 24.4 million visits.

However, some caution is warranted. First, the fringe sites in our sample do not represent the entire population of fringe sites, so our estimates provide a lower bound. Second, we are measuring visits, and due to the role of aggregators, a single user who previously searched within MindGeek sites and consumed content there may now search via aggregators, being redirected to multiple sites (which contributes to an increase in the number of visits).

F. Market Concentration

In this section, we study how competition in the free tube adult market was affected by the shock. The left panel of Figure F.1 presents a measure of industry concentration focusing on the market for free tubes. We compute the Herfindal-Hirschman Index (HHI), in the adult-industry and show a decrease in correspondence and following MindGeek's radical content removal.

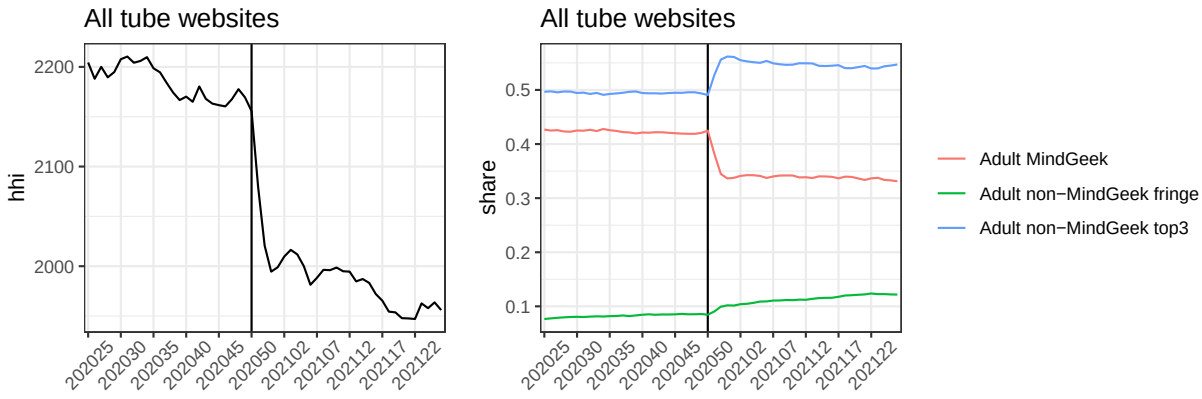


Figure F.1: The left panel plots the weekly HHI×1000 in the adult-industry market based on the free tube sites in our sample. The market shares of domains are computed summing average visits across countries. The right panel plots the weekly average of daily market shares across countries per groups: MindGeek, non-MindGeek mainstream domains (xnxx.com, xvideos.com, xHamster.com), and non-MindGeek fringe domains.

The right panel of figure F.1 illustrates the market shares of the main groups within the adult website industry, defined on the basis of the visits their free tube domains receive. The figure indicates significant shifts in market shares following MindGeek's intervention in content moderation. Before the removal of non-verified content, MindGeek, the largest platform ecosystem, dominated the market (red line), primarily driven by Pornhub. The combined market share of the other three main domains was slightly less than 50% (blue line), while the fringe had a share less than 10% (green line).

MindGeek's share dropped to approximately one-third of the market, while its three mainstream rivals increased their market share to over 50%. The fringe websites also benefited, growing to nearly 13% of the market, with an upward trend during the considered period.

G. Total Duration

Visits to websites are not the only metric for understanding user engagement, especially in a context where search activities change from within a major website to aggregators that redirect traffic to multiple sites. Moreover, one might argue that once traffic is diverted by aggregators, what used to be a single visit to MindGeek's platform for n movies may now result in visits to n different sites. Hence, we now focus on a composite measure: total duration, calculated as the average session duration multiplied by the number of visits. We show that our main insights continue to hold also when considering total viewership time.

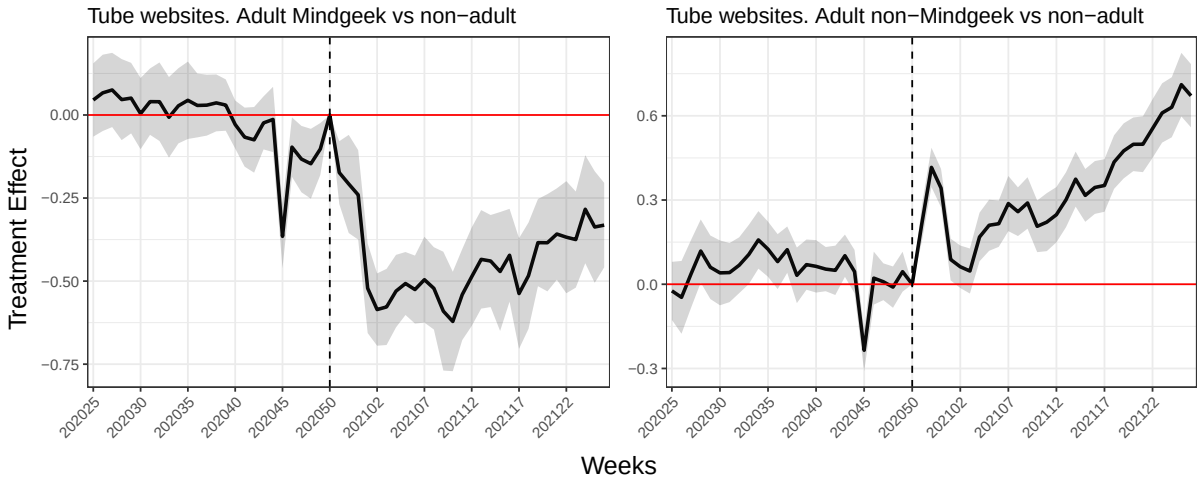


Figure G.1: Event study analyses on daily average total duration per week in log scale. The x-axes show the week. The y-axes report the estimated event study coefficients. Each observation is a country-domain observation in a given week. Standard errors are clustered at the domain-country level. The gray area shows 95% confidence intervals. The reference group consists of non-adult websites.

Note also that, in a context where free adult websites monetize traffic via ads—MindGeek through its in-house ad network, TrafficJunky, while competitors like XVideos and xHamster use TrafficFactory and TrafficStars—total duration serves as a useful proxy for the sites' ability to capture eyeballs, assuming ads are shown at a constant rate and have a uniform value.

Figure G.1 replicates our main event study analysis for both MindGeek and non-MindGeek tubes. Consistent with our main findings, the content purge led to a significant reduction in total duration for MindGeek (-45% on average), which stabilized within the first three months post-shock. However, some initial losses were recovered within six months, with the total viewership time ultimately decreasing by approximately 30%. In contrast, competitors experienced a substantial, stepwise increase in total duration over the same period, doubling

within six months.

We can now replicate the above back-of-the-envelope calculation by considering total time duration.

In Table G.1 we study how the logarithm of the average total time per country-domain changes after the shock. We observe the following. The coefficients closely resemble those in Table 4.

	Dependent variable:		
	log(Avg Total time per Domain-Country)		
	(1)	(2)	(3)
After × Adult MindGeek	-0.452^{***}	-0.340^{***}	-0.560^{***}
	(0.035)	(0.038)	(0.048)
After × Adult non-MindGeek	-0.100^{***}	0.037^{*}	-0.160^{***}
	(0.021)	(0.020)	(0.027)
After × Adult non-MindGeek fringe	0.184^{***}	0.060^{*}	0.140^{***}
	(0.040)	(0.037)	(0.052)
After × Adult MindGeek × t	0.002	0.011^{***}	0.016^{***}
	(0.001)	(0.002)	(0.003)
After × Adult non-MindGeek × t	0.007^{***}	0.017^{***}	0.020^{***}
	(0.001)	(0.002)	(0.002)
After × Adult non-MindGeek fringe × t	0.013^{***}	0.003	0.001
	(0.002)	(0.003)	(0.003)
Domain FE	✓	✓	✓
Country FE	✓	✓	✓
Domain-Country FE	✓	✓	✓
Week FE	✓	✓	✓
Domain Trends		✓	✓
Holidays Included	✓	✓	
Observations	107,111	107,111	85,595

Table G.1: Estimates of the time diversion to non-MindGeek mainstream (top 3) and fringe websites. Columns (1) and (2) consider the whole sample period, and column (3) excludes weeks 42 to 53 of the year 2020. Among the adult websites, we only focus on free tubes. Standard errors clustered at the domain-country level. Note: *p<0.1; **p<0.05; ***p<0.01.

We use estimates from Column (3) of 4 and pre-shock average of total duration time in Table B.1 to compute both the immediate drop and the diversion in total time. Recall that we exclude holidays, thus the immediate effect is calculated at week 4.

Since the pre-shock average of total time for MindGeek country-domains is 10.362 billion seconds and the reduction in total time at week 4 per-country domain for MindGeek is 42%, then there is an immediate loss in total viewership time equivalent to 4.352 billion seconds (approximately 1.21 million hours). The immediate effect for the fringe is not significantly different from zero

(i.e., $\exp(-0,16+0,14)-1 \approx -2$, but not statistically significant), whereas it is equivalent to a 3.8 billion seconds reduction for the top 3 sites, since their pre-shock average is 26.9 billion seconds.

Next, we compute the weekly change in visits per country-domain. Here, time spent on each MindGeek country-domain grows by 1.7% per week while both fringe and mainstream sites experience an increase of 2.2%.

We can now compute the effects at week 15. Since the analysis excludes holidays, the trend has to be consider over 11 weeks:

- *MindGeek.* Per domain, there is a loss of 1.8 billion seconds. With 32 domains, this totals approximately 21.4 million hours.
- *Fringe.* Per domain, there is a gain of 0.102 billion seconds per country-domain. With 152 country domains this totals approximately 15 billion seconds, approximately 3.3 million hours.
- *Mainstream (top 3).* Per domain, there is a gain of 4.3 billion seconds. With 24 domains, this totals approximately 16.9 million hours.

H. Diversion to OnlyFans

Figure H.1 presents a comparison of traffic trends for OnlyFans and Pornhub Premium. The descriptive analysis shows that while OnlyFans has experienced steady growth, Pornhub Premium saw a sudden drop in visits, particularly after the specific shock we study. The exclusive, studio-produced content on Pornhub Premium, which is curated and moderated, likely made it more vulnerable to this disruption. In contrast, OnlyFans operates on a subscription-based model where users directly support individual creators, who also in this case provide consensual content.

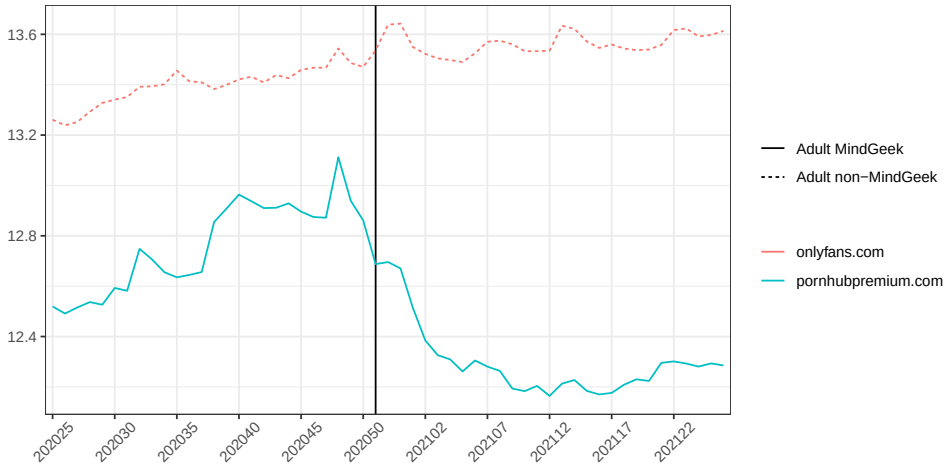


Figure H.1: Average visits per week in log scale between July 1st, 2020 and June 20th, 2021. The plot presents the logarithm of the weekly sum of visits across countries and domains per OnlyFans (red, dotted) and Pornhub Premium (light blue solid). The vertical dark line indicates the week in which MindGeek removed all non-verified content.

The growth of OnlyFans shows a consistent trend and does not appear to have been impacted by the shock, whereas the drop in visits to Pornhub Premium is sudden. Since the content on Pornhub Premium was moderated, featuring exclusives by studios or creators, as users struggle to renew their subscriptions, creators may have migrated to OnlyFans for direct monetization or to cam sites, which experienced a traffic surge (see Section 6.3). Notably, users return less frequently after transitioning to a premium provider, likely influenced by creators moving to platforms like OnlyFans, leading to lasting changes in user engagement.

I. DMCA URL Takedown Requests

In this section, we complement information provided in Section 6.4.

MindGeek's subsidiary, MG Premium, is the largest copyright-holder organization for DMCA takedown requests to Google. This is primarily because the platform hosts millions of copyrighted or licensed content items. These items are often downloaded by users and re-uploaded onto other websites that monetize traffic, typically small fringe sites. Once copyrighted content is copied to another website, the copyright holder can issue a DMCA takedown request, prompting the website to remove it. If the website does not comply, the copyright holder—or a specialized firm managing takedown requests for multiple clients—can request that search engines remove the URLs from search results.

Notably, even MindGeek's own tube sites contained copyright-infringing material prior to its content purge. Following a new policy on December 14, 2020, only verified users were allowed to post content, under the assumption that these users were legitimate owners or licensed holders of the uploaded material.

When a request is made, the request can contain multiple URLs. As an example, in a single week, ending on January 9, 2023, Google received DMCA notices for the removal of 4.7 million Daftsex URLs[46].

Table I.1 presents descriptive statistics for DMCA requests made by the top 500 organizations and copyright holders from June 2020 to June 2021. We focus on the top-500 in order to have a suitable control group since the dataset is made by a long tail of small copyright-holders.

	Other Organizations		MindGeek	
Metrics	Before	After	Before	After
Total requests	63,834	71,634	6,388,399	8,604,949
URLs removed	29,652	28,902	3,519,746	3,070,496
URLs not indexed	24,502	33,725	1,431,395	3,457,660
URLs with no action	9,678	9,007	1,437,257	2,076,791
N. Organizations	499	499	1	1

Table I.1: Descriptive statistics of the average monthly URLs requests for content removal under DMCA by the top 500 reporting organizations. Source: https://transparencyreport.google.com/copyright/overview.

[46] https://torrentfreak.com/90-million-dmca-takedowns-in-90-days-mindgeeks-32m-piracy-win-meets-reality-23030

Table I.1 shows that MindGeek's requests increased after the shock, rising from approximately 6.4 million to 8.6 million monthly requests, while other organizations submitted approximately on average 63,834 monthly requests before the shock and 71,634 monthly requests after the shock.

Although MindGeek's requests far exceeded those from other organizations, a notable portion of requests received no action from Google after the shock. For example, the number of MindGeek's URLs with no action rose from about 1.44 million to 2.07 million after the shock. On the other hand, there is no systematic difference in the URL requests followed by no action by Google for the other 499 organizations.

Furthermore, the number of MindGeek URLs not indexed by Google increased substantially to 3.5 million after the shock. In contrast, other organizations also experienced an increase, although to a lesser extent on average.

Finally, the average number of monthly URLs removed by Google remained almost stable for other organizations, it decreased from 3.5 million to 3.1 million for MindGeek.