

Front-Running the Red Carpet: A Measurement Study of Media Torrent Leaks and User Risks

Scott Seidenberger, Noah Pursell, Anindya Maiti

University of Oklahoma, USA

We study how streaming service fragmentation (SSF) can create media access friction (MAF), making illicit channels comparatively more attractive for some audiences. This friction is fueled by the current exclusivity model employed by the large streaming platforms. Decentralized, peer-to-peer (P2P) protocols like BitTorrent, through its trackerless DHT, present a robust illicit distribution channel. We introduce a cross-layer design that links the supply of torrents available via the BitTorrent DHT with the diffusion mechanics of specific torrent swarms around the release of new titles. Our macro panel (n=1,469 titles) shows that each additional legitimate provider of a given title reduces the instantaneous hazard of a first torrent by about 4.2%. In a targeted one-month micro panel of high-salience new releases, we tracked 244 torrent swarms involving 167,607 unique peer IPs across 220 jurisdictions and 7,277 ISPs. We find the torrents follow logistic S-curve diffusion with sharp and short-lived first mover advantage, while Bass imitation parameters are near zero, suggesting shock-driven growth rather than strong peer-to-peer contagion. Activity is highly concentrated by jurisdiction and network. Pre-release windows show elevated risk to users, with one-third of torrents carrying suspicious executables. Because BitTorrent operates as a federation of title-centric micro-networks with no stable chokepoint, broad interventions such as domain takedowns, CDN de-peering, or Layer-3 blocking do not map cleanly onto this distribution channel. The descriptive implication is that broader legitimate availability is a lower-collateral avenue for stakeholders to evaluate when seeking to reduce the demand conditions under which early leaks, early diffusion, and user cyber risk become most salient.

Keywords: streaming fragmentation; BitTorrent; media access friction; diffusion; measurement

Author contact: seidenberger@ou.edu

Introduction

Consumers of digital entertainment media feel squeezed. They either pay too much for what they get or they miss out on shows and movies they want at a price they view as fair. This streaming service fragmentation (SSF) creates media access friction (MAF). We use SSF to describe the splintering of rights and access to titles across many on-demand platforms, and MAF to capture the search, coordination, and subscription costs this creates for viewers. Subscription fatigue is real and households keep asking how many logins they really need. Access to popular media has long helped audiences cohere across lines of class, geography, and taste, yet fragmentation and access challenges may erode that shared culture over time Lotz (2021). U.S. Chamber of Commerce estimates put global annual loss from digital video piracy (for film and TV combined) in the range of 79 to 193 billion dollars Blackburn et al. (2019). These losses often mean loss of budgets for studios and fewer resources for creators. In this environment, torrenting often becomes the path of least resistance for filling the gaps in a subscription portfolio.

We analyze piracy as a substitution response to access friction in a mature platform market where exclusivity is the default strategy rather than a simple function of price sensitivity. Fragmented access forces users to juggle subscriptions, tolerate delays, or go without. When perceived value falls we should expect movement towards lower friction distribution channels. We also treat the household streaming platform mix as a portfolio choice problem, where a streaming service functions like a basket of content similar to how a fund is a basket of equities. This framing motivates individual level questions about how people choose their portfolio of services and how often they rebalance it.

Position relative to illicit Internet Protocol television (IPTV) and illegal streaming sites

This study complements, rather than substitutes for, the growing literature and actions against illegal centralized streaming sites. Illicit IPTV and cyberlocker ecosystems are fundamentally server mediated, that is, content traverses named domains, CDNs, and payment rails where researchers and law enforcement can often observe activity through targeted reconnaissance of DNS telemetry, advertising traces, and standard network traffic. However, BitTorrent is a decentralized peer-mediated system with no single observable vantage point. The meaningful unit of analysis moves from a single (or small set of) domain names to a loose collection of title-specific torrent swarms that spin up, peak, and collapse in relatively short time horizons. Theory and measurement must therefore be framed differently in this

context. Site-centric approaches that work for IPTV (site takedowns, CDN de-peering, merchant account interdiction, standard web analytics) do not transfer to P2P. Instead, P2P requires **longitudinal observation from multiple, stable vantage points** that jointly capture (i) torrent *supply emergence* on the DHT and (ii) *demand realization* inside swarms temporally aligned to release events. Our contribution implements this cross-layer design with a supply-side panel from *MagnetDB* to temporally situate and discover the onset of illicit availability of titles, and a demand-side panel (swarm monitoring) to empirically determine adoption curves, diffusion dynamics, and stakeholder concentration patterns. The approach surfaces *timing-sensitive outcomes* that cannot be determined from the methodologies employed against client-server based distribution platforms.

Methodologically, we develop a mechanism that links upcoming release calendars with continuous DHT discovery to measure early torrent swarm behavior. This approach produces reproducible diffusion estimates despite short lifespans and high churn, and reframes measurement from counting eventual file availability to identifying when leaks occur and what unfolds in the first two weeks. Using this framework, we document descriptive associations between streaming availability, leak timing, and early swarm dynamics. Titles with limited legitimate distribution tend to leak earlier, consistent with substitution toward lower friction channels when SSF and MAF raise the cost of legal access, though platform strategies and user demand are jointly determined.

Our study delivers four key findings:

- **Exclusivity and leak risk.** We treat SSF and MAF as consumer-side mechanisms linking exclusivity to elevated P2P leak risk. A Cox hazards model shows that each additional legitimate provider reduces the instantaneous leak hazard by roughly 4.2%, with exclusive titles reaching a 50% leak probability much earlier than multihomed titles.
- **Cross-layer measurement design.** We combine a supply-side panel from a large torrent discovery dataset with a demand-side panel from swarm monitoring. This framework produces reproducible early diffusion estimates and directly links leak timing to realized swarm demand.
- **Early swarm dynamics.** We find consistent timing-sensitive patterns in a targeted one-month panel of high-salience new releases. Adoption follows logistic S-curves for most torrents, and early well-seeded releases capture a brief lead in active peer share that lasts about a week. The

Bass imitation parameter is near zero for almost all hashes, so the diffusion evidence supports bounded S-shaped growth but not strong endogenous peer-to-peer contagion.

- **User risk and malware exposure.** Roughly one-third of pre-release torrents contain suspicious executables, confirmed through sandbox analysis. We also identify concentration of swarm activity in specific ISPs and jurisdictions, highlighting potential enforcement pathways.

Threat Model

Adversaries and Actors. We distinguish between two adversarial actors operating in the pre-release window. First, *Scene groups* produce authentic, high-quality early leaks of unreleased film and television titles. Their motivation is reputational: gaining status and credibility within an illicit gift economy. Scene groups intentionally provision healthy seeds and race to be first; in our threat model, they are distinct from actors who distribute malware or seek direct financial gain. Second, *malicious impostor actors* exploit user anticipation by publishing fake torrents that mimic Scene releases in filename, metadata, and file size but contain trojans or deceptive payloads.

Interaction Between Actors. Scene groups create a predictable period of heightened demand by normalizing high-quality early availability. Malicious actors exploit this window by impersonating trusted Scene-style releases. Because users rely on heuristic indicators such as early availability and seed ratios, the presence of authentic Scene uploads can inadvertently expand the attack surface for impostors.

Capabilities. Scene groups can obtain unreleased media, produce high-quality encodes, insert them into the DHT, and provision strong initial seeding. Malicious actors can inject forged torrents with realistic metadata, inflate file sizes to mimic legitimate encodes, and generate short-lived seeding shocks to appear swarm-healthy at release time.

Constraints. Scene groups, as modeled here, do not distribute malware and cannot prevent impostor uploads or swarm hijacking. Malicious actors cannot tamper with genuine Scene torrents and rely instead on impersonation and timing. Neither actor controls BitTorrent protocol integrity.

Victim Model. Victims are users who join early-stage public swarms and rely on superficial heuristics to judge authenticity. They lack specialized malware detection tools and may over-trust first-

available torrents due to the perceived reliability of Scene releases.

Security Implication. This ecosystem contains both non-malicious illicit actors (Scene groups) and malicious adversaries (impostors). Scene releases shape the timing and expectation structure of the underground distribution pipeline, while malicious actors front-run or parallel these releases to distribute harmful payloads. Our measurement characterizes how this dynamic is associated with elevated malware risk in the pre-release window and helps explain the diffusion patterns, seed concentration, and opportunistic abuse observed in early swarms.

Background and Related Work

The Scene

The illicit distribution pipeline for high-salience film and television releases is anchored by a long-standing subculture known as the Scene, a federated network of release groups that specialize in producing authentic, high-quality leaks ahead of official availability Goldman (2005). Unlike profit-driven cybercriminals, Scene groups operate within a gift-economy logic where status, reputation, and “first release” prestige are the primary incentives Huizing and van der Wal (2014). Their early uploads are typically clean, well-provisioned, and rapidly mirrored, which can train audiences to trust, and actively seek, first-available torrents. This reputational economy has an unintended security consequence: by normalizing early unauthorized availability, Scene activity creates a predictable window of heightened demand and minimal verification, which can be exploited by malicious impostors who publish fake torrents carrying malware or deceptive payloads David (2017). Understanding this ecosystem therefore requires treating the Scene not as an edge case but as a structural force that shapes leak timing, swarm composition, and user risk. Our cross-layer measurement approach captures both the authentic illicit supply generated by Scene groups and the opportunistic abuse introduced by malicious actors who front-run the same release window.

Economics of Exclusivity

Exclusivity in platform markets follows a lifecycle pattern. It is most common in nascent stages when platforms race to build sides, and then again in mature stages when incumbents protect share. Non-exclusivity tends to be most prominent in the intermediate stage when expanding reach matters more than locking in users Mantena et al. (2010). We are now observing a mature streaming market, yet

observation is not the same as optimality. The fact that exclusivity is prevalent today does not prove it maximizes welfare or long-run profits in every context.

Windowing and sequential release strategies let firms extract surplus from heterogeneous willingness to pay. By staging access across channels and time, providers practice temporal price discrimination that has been shown to be highly profitable for digital media markets Seifert et al. (2023). The economic appeal is straightforward. Early access and premium formats harvest high value users while later windows expand demand at often lower prices once scarcity and novelty decay.

The current streaming landscape layers a second set of incentives. To differentiate and build defensible moats, major platforms stood up in-house studios and prioritized platform exclusive originals. The strategy is to smooth subscription revenue while limiting dependence on third-party licensed works. These bets come with heavy fixed and organizational costs that must be justified through user retention and customer lifetime value, which in turn encourages continued emphasis on exclusive titles and tighter controls over their access Cardoso et al. (2022).

Streaming services are not interchangeable products. They differ on their geographic reach, library offerings, and ownership rights which produces distinct strategic postures towards exclusivity and licensing Lotz (2022). Global fragmentation follows from these differences. Non-US providers often compete most effectively by curating complementary local catalogs rather than confronting the largest US players head-on, which sustains a patchwork of access around the world Lotz and Eklund (2024).

For a comprehensive background on access-versus-exclusivity tradeoffs in digital creative and cultural markets, see Nasta (2025).

Piracy Measurement Approaches

Measuring digital piracy on a decentralized P2P network is fundamentally challenging because coordination, discovery, and state are diffused across thousands of ephemeral nodes rather than concentrated on observable servers. In BitTorrent specifically, centralized illicit IPTV and cyberlocker sites leave site-centric traces in DNS, HTTP, and payments, but BitTorrent externalizes coordination to the DHT and peer exchange, so “each peer becomes a tracker.” BEP 5 and BEP 9 formalize how discovery and metadata retrieval work without a portal or tracker, which makes early supply detection feasible but also

removes any single choke point to observe demand. Prior systems work shows the DHT is large, fast-churning, and idiosyncratic, which argues for longitudinal, multi-vantage designs rather than one-shot crawls Loewenstern and Norberg (2008); Hazel and Norberg (2008); Wolchok and Halderman (2010); Wang and Kangasharju (2013); Cuevas et al. (2010).

Existing measurement approaches only solve part of the problem. Web-panel clickstreams identify how users find substitutes and navigate finding new sources when authorities block piracy sites, but they are strongest for client-server ecosystems and link portals, not trackerless swarms; blocking many sites at once reduces piracy, while blocking a single dominant site mainly induces rerouting Danaher et al. (2020); Poort et al. (2014). ISP reports map protocol families and regional trends but are not title-resolved Sandvine (2023). DMCA transparency datasets capture enforcement volume and defects rather than actual consumption Urban et al. (2017); Seng (2015). Within P2P, tracker and portal crawls, active swarm enumeration, and DHT discovery each carry biases from churn, sampling frequency, and coverage Kryczka et al. (2011); Zhang et al. (2010).

Social-Media Gateways to Piracy

Social media platforms such as Reddit, Discord, and Telegram function as community-based gateways into digital piracy, lowering search and coordination costs while broadcasting norms that enable the behavior and, for many, keep it morally defensible Lee et al. (2018). Public forums, such as Reddit's r/Piracy, aggregate know-how, tool recommendations, troubleshooting, and "what works today" intelligence. Semi-public spaces like Discord servers and encrypted channels via Telegram then extend coordination into even higher-trust communities. Together, these spaces provide a social infrastructure through which novices encounter experienced peers, observe practices, and receive feedback, all classic ingredients for social transmission of behavior.

Social learning theory (SLT) and the theory of planned behavior (TPB), as applied to digital piracy, together explain how social media becomes a gateway to digital piracy. SLT holds that deviant behaviors are acquired through differential association, definitions, imitation, and differential reinforcement, and each maps onto what social media provides Morris and Higgins (2010); Lee et al. (2019). Threaded discussions and real-time chats increase association with peers who model illegal downloading and uploading. Highly upvoted posts and pinned "starter" guides supply definitions that normalize the act, often via a "victimless crime" framing. Step-by-step walkthrough guides and success stories

encourage imitation, and reputational feedback via upvotes, reactions, and mentions provide social reinforcement.

TPB adds the mechanism by which these exposures form intentions. Attitudes are made favorable because the behavior is framed as low harm, subjective norms are strengthened by visible endorsement and community rules that emphasize user safety over questions of legality. Empirically, prior work shows that virtual association with deviant peers is associated with more permissive views of digital piracy and that those who engage online with peers who break the law online are more likely themselves to participate in digital piracy Phau et al. (2014).

Contextual evidence from a public piracy gateway To anchor these theoretical claims in observable platform behavior, we collected the available historical corpus of posts and comments from *r/Piracy*¹. We focus on this subreddit because it is large, public, explicitly organized around digital piracy, historically archivable, and unusually transparent compared with Discord servers, Telegram channels, private trackers, or invite-only forums. It is therefore a useful reproducible case for observing public piracy know-how and P2P vocabulary, even though it is not representative of all piracy communities or all people who use BitTorrent. Figure 1 shows engagement with the subreddit has increased over time, consistent with the view that social platforms are durable hubs for coordination. Using a dictionary of P2P terms (e.g., *torrent*, *magnet*, *tracker*, *BitTorrent*, *seed/seeders*, *leech/leechers*, *DHT*), we compute the annual share of posts containing at least one P2P keyword/phrase. In the past several years, approximately 80% of posts reference P2P concepts, indicating that community discourse is heavily focused on P2P sharing practices rather than purely streaming. While descriptive, this pattern is consistent with our motivation that social media serves as a gateway to socialize and direct users to P2P distribution channels. Reddit’s audience is disproportionately English-language and U.S.-oriented relative to the global BitTorrent population. Additionally, subreddit moderation, deleted posts, API changes, and posthoc archival coverage can all shape what is observed. We therefore treat this series only as contextual evidence that P2P-centric piracy discourse is sustained over time, not as a direct measure of piracy volume, title-level demand, or the full social-media ecosystem surrounding piracy.

¹https://github.com/ArthurHeitmann/arctic_shift

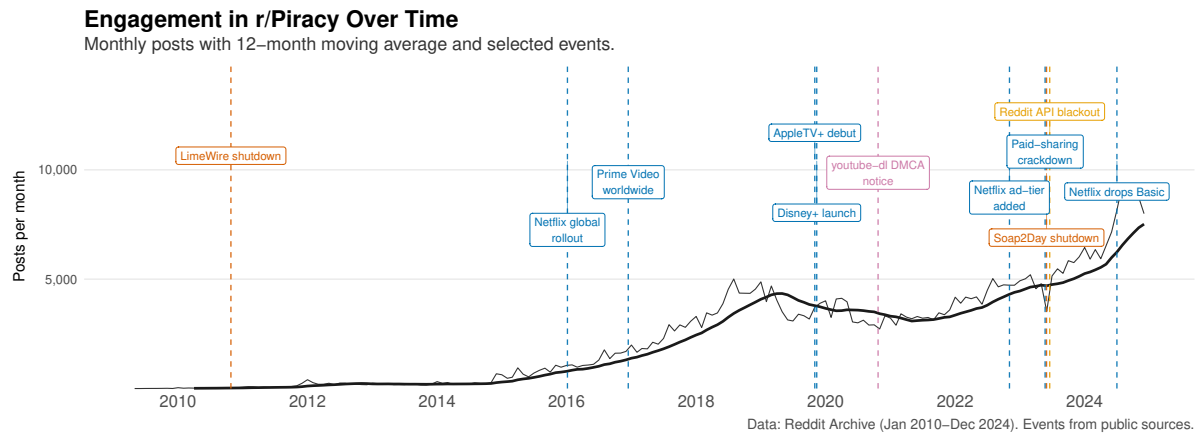


Figure 1. Engagement in the subreddit r/Piracy over time, indicating a rising trend.

Positioning and Novelty

Prior work establishes important pieces of the piracy ecosystem, but those pieces are often studied in isolation. DHT and tracker studies show that decentralized discovery can be measured, even though the resulting signals are noisy. Studies of site-blocking and clickstream behavior show how users substitute across centralized portals, while social-learning research explains how online communities can normalize piracy and malware research documents the direct risks posed by fake torrents. These literatures typically separate sites, users, files, and swarms rather than linking title-level supply emergence to title-specific demand realization around release events. Our contribution is a cross-layer measurement design that joins a supply-side DHT panel to a demand-side swarm-monitoring panel, aligns both to release timing, and uses that linkage to characterize leak timing, early diffusion, network concentration, and pre-release user risk.

Malware in Pirated Content

Prior work has established that pre-release windows are polluted by fake or booby-trapped torrents as malicious actors know that there is a population of users that are anticipating a torrent with specific metadata Cuevas et al. (2014); Schäfer and Malinka (2009). Furthermore, if a malicious torrent gains first-mover advantage in terms of swarm health, then that torrent will enjoy a rapid influx of new users in the swarm until a suitable, legitimate, new leader emerges.

Other works have examined the content of such malware propagated on BitTorrent, and the conclusions largely center on client-side preventative measures by using a heuristic-based approach such as the seeder/leecher counts and ratios, as well as the reputation of the torrent publisher Berns and Jung (2008); Kim et al. (2012). Lastly, in this work we confirm if suspicious files are actual malware by using a dynamic malware analysis Or-Meir et al. (2019).

Hypotheses

Based on current industry observations and public discourse, we group our hypotheses into four thematic blocks, each spanning one or both empirical data layers: the Macro Panel and the Micro Panel.

Platform Distribution and the Supply of Illicit Content

Streaming service fragmentation creates media access friction by increasing search, coordination, and subscription costs for subscribers. This friction can make illicit channels comparatively more attractive when viewers cannot access desired titles through their existing service portfolio. We therefore expect that **exclusivity accelerates leakage**. Titles available on fewer platforms should leak their first torrent earlier than those that are launched on many services (H1a). The flip side is that **wider distribution could inflate torrent volume**. When a title is available across more platforms, there will be more sources available to supply the torrent of the title, which means that a multi-homed title may accumulate a larger number of distinct torrent copies (H1b).

Social platforms remain important to the theory because they lower search costs, circulate technical know-how, and normalize P2P access. We therefore use public Reddit discourse, especially r/Piracy, as contextual evidence for the persistence of P2P-centered piracy discussion. We do not, however, treat aggregate Reddit activity as a title-level measure of piracy demand or include it as a modeled covariate in the title-level hazard and count regressions.

Lifecycle Dynamics of Illicit Swarms

While the prior section relies on the supply-side data of what torrents become available and when, there are several questions we can answer by using the fine-grain demand-side data from torrent swarms themselves. Once a torrent is seeded, its adoption should follow *diffusion theory*, albeit under legal risk. We hypothesize that **the vast majority of torrents conform to classic S-curves** Meyer et al.

(1999). We test this first as bounded logistic growth (H3a), and then test the stronger claim that Bass-style imitation contributes meaningful explanatory power beyond the initial seeding shock (H3b). We anticipate that saturation levels and growth rates will cluster within a relatively narrow range across titles, indicating rapid but bounded expansion.

Finally, when several releases of the same title compete, we foresee a first-mover advantage, where the torrent that gains an early lead in peer counts is likely to maintain higher share while later entrants decline more quickly (H4). These hypotheses will be evaluated in the Micro Panel using non-linear diffusion fits.

Spatial and Network Heterogeneity

Internet infrastructure shapes where and how intensively P2P piracy unfolds. We expect that higher fixed broadband capacity will shorten the time-to-first appearance of a title within a country (H5a). For an individual P2P session, upload bandwidth should increase the likelihood of an individual staying within the swarm to seed (H6a). Additionally, we anticipate that there will be a high degree of upload and seeder concentration at the ISP layer (H6b). To evaluate these hypotheses, we enrich the Micro Panel swarm data with detailed broadband data from Ookla.

Malware in Pre-release Fakes

We hypothesize that fraudulent torrent files carrying malware will concentrate in the short window preceding an official release, capitalizing on heightened user anticipation. Moreover, we expect a meaningful share of these pre-release fakes to contain harmful code, posing cybersecurity risks to downloaders (H7). The data to conduct this analysis is collected in the Micro Panel.

Together, these four thematic blocks link platform strategy to leak timing, chart the mechanics of swarm growth, expose the attendant cybersecurity risks, and reveal how infrastructure and network organization shape observed P2P piracy patterns.

Macro-Panel Analysis: Fragmentation & Early Leakage

Datasets

To answer questions around the supply of illicit content and the effects of factors such as genre, budget, and release strategy on streaming platforms, we leverage supply-side torrent data from *MagnetDB*²⁴ Seidenberger et al. (2025). This dataset collected torrents directly from the Bittorrent DHT from 2018 through 2024, and matched 1.56 million torrents to movies and TV shows from IMDb.

To further enrich this dataset, we used TMDB’s API to match titles to the streaming platforms they were available on The Movie Database (2025). Combined, this produced a complete dataset in which, for each title, we know when the torrents first appeared, how many appeared over time, on what streaming platforms the titles were available on legitimately, and other important metadata that serves as controls, such as the title’s budget, genre, and popularity.

Using the Arctic Shift Reddit archive, we acquired the available historical posts and comments of the *r/Piracy* subreddit, a large public community “Dedicated to the discussion of digital piracy” that reported over 2.3 million subscribers during data collection. This subreddit serves as a public coordination and troubleshooting community for people discussing digital piracy, while remaining subject to moderation, deletion, API/platform changes, and posthoc archival limits. Because this series is an aggregate discourse measure rather than a title-level measure of downloads, swarms, or piracy volume, we use it as contextual evidence in the Background section rather than as a covariate in the Cox and negative-binomial title-level models.

Methods

We adopt a two-stage empirical strategy that separates *when* the illicit supply of a title begins from *how much* ultimately circulates. First, to explain timing, we estimate a Cox proportional hazards model in which the event is a title’s first torrent appearance on the DHT. The covariate set contains the number of legitimate streaming providers, our operationalization of fragmentation, together with controls for cumulative demand and scale. These metrics include the TMDB reported popularity, log-transformed vote count, and log-transformed budget.

On TMDB, popularity and vote count are distinct metrics that reflect different aspects of a

movie’s engagement. While both are related to user interaction, popularity is a broader measure of a movie’s overall interest, while vote count specifically tracks the number of times a movie has been rated. A movie with a high popularity score might have many users interacting with it in different ways, even if it doesn’t have a huge number of ratings, such as users adding it as a favorite and adding it to their watchlists.

We set $t = 0$ to the earliest *public release* date for each title (theatrical premiere or first legitimate streaming availability, whichever comes first). Titles whose public release predates our observation window (January 1, 2018) enter the risk set with *delayed entry* at $\max\{t = 0, \text{Jan 1, 2018}\}$. Titles that never leak are right-censored at December 31, 2024. A small share of titles exhibit pre-release leaks and are excluded.

TMDB popularity and vote counts were retrieved in a single snapshot after the full observation window (2018–2024). As such, they reflect cumulative user engagement and should be interpreted as long-run demand proxies rather than ex ante expectations at release.

The model is:

$$\begin{aligned} h(t | X) = & h_0(t) \exp(\beta_1 n_{\text{providers}} + \beta_2 \text{popularity} \\ & + \beta_3 \log(\text{votes}) + \beta_4 \log(\text{budget})) \end{aligned} \quad (1)$$

where $h_0(t)$ is a nonparametric baseline hazard. The Cox framework is useful because the baseline hazard for leaks is unknown and may vary with time. This allows the underlying leak-risk curve to take any shape. Prior to estimation we assess the proportional-hazards assumption that each covariate’s relative effect is constant over time by inspecting Schoenfeld residuals. We also guard against additional uncertainty introduced by within-title correlation by calculating robust, *HC0* sandwich standard errors.

Second, to explain volume, we model the total number of torrents per title with a negative-binomial generalized linear model:

$$\begin{aligned} \log(\mathbb{E}[Y | X]) = & \alpha_1 n_{\text{providers}} + \alpha_2 \text{popularity} \\ & + \alpha_3 \log(\text{vote_count}) \\ & + \alpha_4 \log(\text{budget}) \end{aligned} \quad (2)$$

where $Y = n_{\text{torrents}}$. A likelihood ratio test strongly rejected the equidispersion assumption of a

Poisson distribution ($\theta \neq 0, p < 0.001$), justifying a negative-binomial.

To visualize how timing and volume interact, we scale the number of torrents, the time-to-first-torrent, and the other covariates and perform a principal component analysis. Then, titles are clustered together based on these dimensions.

$$\mathbf{z}_i = \begin{bmatrix} n_{\text{torrents},i} \\ t_{\text{event},i} \\ \log(\text{budget}_i) \\ \text{popularity}_i \end{bmatrix}, \quad i = 1, \dots, N. \quad (3)$$

Platform availability, marketing, and piracy risk are jointly chosen by studios. We condition on observable demand proxies (budget, popularity, vote count), but cannot rule out unobserved confounders. We therefore interpret our hazard and count models as characterizing associations, not estimating the structural effect of exclusivity on piracy.

Results

Figure 2 shows the resultant survival curves for titles released on varying number of streaming platforms. Table 1 reports the full Cox proportional hazards estimates. Three principal findings emerge.

Provider availability and first-leak timing The number of legitimate outlets to watch the title is strongly and inversely associated with the hazard of torrent appearance. Each additional provider corresponds to an approximately 4.2% lower instantaneous risk. Holding all covariates at their means, a title that is exclusive to a single service reaches 50% “survival” chance in about 120 days, whereas a title carried by eight services does not reach that point until roughly 275 days.

Audience engagement and earlier leakage User engagement, operationalized by the log of the TMDB vote count, is a significant positive predictor. A one-unit increase in $\log(\text{votes})$, equivalent to a 2.7 times rise in raw votes, is associated with a 19% higher hazard of first leak.

Once controlling for engagement, the title’s budget does not significantly contribute to the time-to-first torrent risk. This suggests that production budgets, and by proxy marketing scale, are

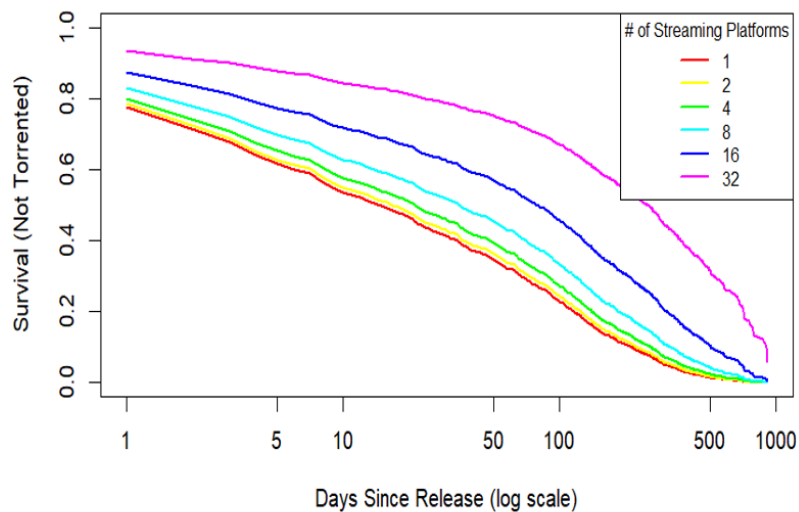


Figure 2. Predicted survival curves for time to first torrent as a function of the number of streaming-service providers. Covariates are fixed at sample means.

not independently associated with earlier torrent appearance once user engagement is modeled. The survival analysis supports the descriptive fragmentation hypothesis that exclusivity is associated with earlier torrent appearance, whereas broader platform availability is associated with a longer interval before the first observed torrent.

Torrent volume aligns with demand, not fragmentation The negative-binomial count model shows in Table 2 that once a title appears on the BitTorrent DHT, the *number* of streaming providers no longer predicts how many torrent files ultimately circulate. The coefficient is virtually zero. By contrast, user engagement and production scale (budget), are strong predictors of volume. A one-unit rise in the log of vote count (our cumulative demand proxy) is associated with a 29% higher expected torrent count, and every unit increase in the log of the budget adds about 2%. This result does not support H1b as stated: wider legitimate distribution is not associated with a larger number of torrent copies after leakage once demand proxies are modeled.

Table 1: Cox proportional-hazards model of time to first torrent ($n = 1,469$). Robust standard errors in parentheses.

Covariate	Coefficient	Hazard Ratio	p -value
$n_{\text{providers}}$	−0.0421 (0.0048)	0.959	< 0.0001
Popularity	−0.0552 (0.0460)	0.946	0.230
log(vote count)	0.1754 (0.0347)	1.192	< 0.0001
log(budget)	−0.0040 (0.0051)	0.996	0.435

Table 2: Negative-binomial model of total torrent files per title ($n = 1,469$). Standard errors in parentheses; incidence-rate ratios (IRR) are $\exp(\beta)$.

Covariate	Coefficient	IRR	p -value
$n_{\text{providers}}$	0.00012 (0.00498)	1.000	0.980
Popularity	−0.0704 (0.0527)	0.932	0.181
log(vote count)	0.2565 (0.0401)	1.292	< 0.0001
log(budget)	0.0219 (0.0057)	1.022	0.0001

Dimensional structure of piracy risk We z-score four variables: total torrent count, time-to-first torrent, log(budget), and TMDB popularity, then project titles into the first two principal components. The loadings in Table 3 justify the axis labels. PC1 loads positively on torrent count and negatively on time-to-first torrent, so it indexes overall piracy intensity. PC2 loads most strongly and positively with time-to-first torrent, with smaller positive weights on budget and popularity, so it represents leak latency. Clustering this PC plane with $k = 4$ yields a compact typology. Instant Blockbusters (upper right; leaks within days with very high volumes), High-Profile Early Leaks (early leaks within a month and dozens of torrents), Mid-Tier Mainstream (moderate delays and modest volumes), and Slow-Leak Minimal Piracy (late leaks and few torrents). Genre tallies indicate that sci-fi, fantasy, and action are over-represented in the two high-risk groups, while drama and documentaries are concentrated in the minimal-risk cluster.

Table 3: PCA loadings for the first two principal components.

Variable	PC ₁	PC ₂
n_{torrents}	0.460	−0.394
Time-to-first torrent	−0.279	0.786
log(budget)	0.606	0.320
Popularity	0.586	0.352

Annotated cluster map Figure 3 plots titles in the piracy-intensity space (x: number of torrents, y: time-to-first-torrent), with point size proportional to budget and colors matching the four clusters. A diagonal “risk line” is overlaid as a visual guide to increasing exposure that moves toward the lower-right quadrant where torrents are both numerous and early. Annotated exemplars illustrate each archetype: *Oppenheimer* and *Doctor Strange in the Multiverse of Madness* typify the high-intensity side of the line; *Insidious: The Red Door* sits in the mid-tier region; *Our Christmas Mural* and *Do Not Expect Too Much from the End of the World* fall in the minimal-piracy zone. The figure complements the model results by showing how title-level attributes translate into concrete risk profiles.

Taken together, the survival model characterizes the association between provider exclusivity and earlier piracy onset, while the count model and clustering reveal that torrent volume is primarily demand-driven rather than fragmentation-driven once a title has leaked.

Effect identification and limitations Our macro analysis estimates associations, not causal effects. Platform availability is chosen by studios and may correlate with unobserved determinants of piracy risk. In the reported models, we condition on observed demand and scale proxies and use a longitudinal panel spanning genres and media types. We therefore interpret the estimates as substantively meaningful descriptive associations: greater exclusivity is associated with earlier first leakage, while broader reach is not associated with larger post-leak torrent volume once demand proxies are modeled.

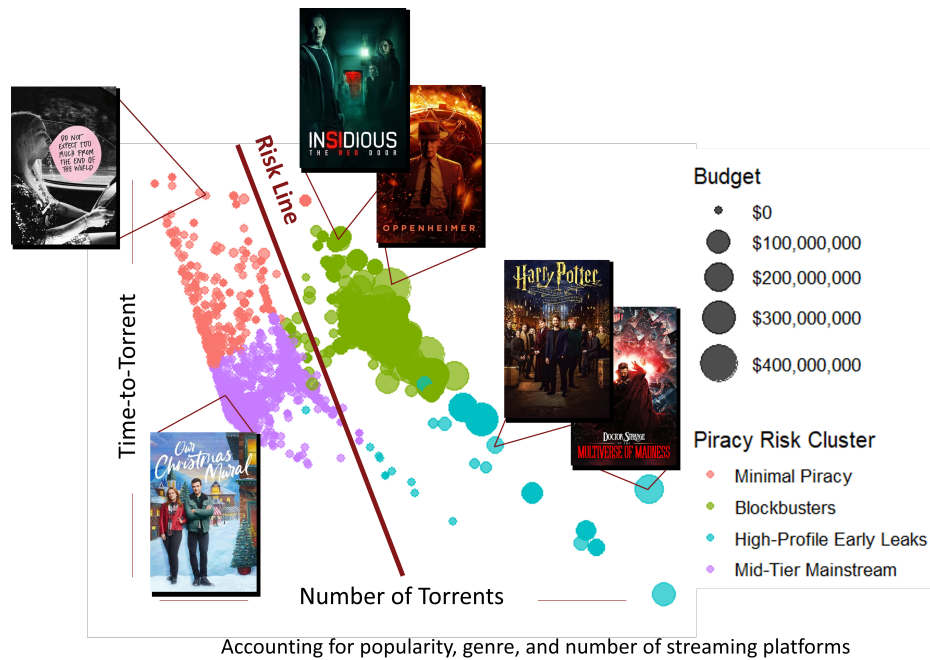


Figure 3. Annotated cluster map in piracy-intensity space (x: number of torrents, y: time to first torrent). Point size encodes budget; colors denote the four clusters. The diagonal risk line highlights higher-exposure titles, and exemplar posters mark representative cases. All copyrights of posters are retained by copyright holders.

Micro-Panel Lifecycle Analysis: Diffusion, Geography & Malware

Dataset

The micro panel monitored 28 new titles, aggregating per-peer swarm observations for a resulting 244 unique torrents, covering 167,607 distinct IPs across 220 ISO-3166 countries/territories and 7,277 ISPs. For each unique torrent *infohash*, we reconstruct time-stamped adoption curves (cumulative unique peers since the first observed peer t_0), identify parallel releases of the same title, and attach network performance data from Ookla at the session start quarter Ookla (2025).

To collect this data, we deployed an open source torrent swarm monitoring tool, *Torrent-Monitor* Martinez (2025). To know what torrent infohashes to monitor, we referenced upcoming re-

leases on IMDb² and JustWatch³, and wrote a custom script that queried those titles against the live *MagnetDB* discovery stream. When there was a hit for the query, just as a normal user would use their torrent search tools, we added that torrent to the Monitor. This monitor ran consistently for a one-month period from 14 July to 15 August 2025. The monitoring tool joined swarms but did not engage in actual file transfer past exchanging metadata; additionally, in order to avoid misuse we only publish hashed IPs and do not publicly share magnet links. We make our data available on OSF⁴.

We selected 28 upcoming film and high-profile TV titles from IMDb and JustWatch calendars that (i) had clear release dates between 14 July and 15 August 2025, (ii) were likely to generate visible public interest, and (iii) could be queried unambiguously against the live *MagnetDB* discovery stream. Candidate titles were excluded when title strings were ambiguous, release timing could not be verified, no public DHT activity appeared during the monitoring window, or observed activity appeared confined to ecosystems that our public-DHT monitor could not observe. We constrained the window to one month because continuous swarm monitoring requires persistent participation in each swarm, repeated peer enumeration, geolocation enrichment, and responsible handling of sensitive network identifiers. The resulting design prioritizes measurement depth over breadth. As a result, our diffusion results speak primarily to mainstream, high-salience releases in a short release window rather than to the long tail of all media content. We treat these 28 titles as a purposive case sample, not a representative draw of all new media content on BitTorrent. Because selection favors salient, unambiguously named titles with observable public-DHT activity, the panel likely overstates swarm sizes and the strength of early first-mover advantage while understating long-tail, non-English, and closed-ecosystem (e.g., private-tracker) behavior; moreover, by conditioning on titles that produced an observable public leak, it characterizes diffusion given that a leak occurred rather than the incidence of leakage across all releases. The appendix summarizes the sampling protocol and boundary conditions.

Diffusion & First-Mover Dynamics (H3-H4)

Do torrents grow like classic S-curves, and when several releases of the same title compete, does the earliest viable torrent hold a material advantage? To test this, we fit logistic and Bass diffusion curves to each hash's cumulative peer-trajectory, and we run a title-level event model that compares the leader's

²<https://www.imdb.com/calendar/>

³<https://www.justwatch.com/us/new>

⁴https://osf.io/xgmsq/overview?view_only=d6dd664d0eaf4665a819242145a5c9aa

daily share of active peers to later entrants, aligning time at the first release.

Why S-Curves for torrents? In diffusion models, the cumulative number of adopters typically follows a sigmoid path where adoption starts slowly, accelerates as access and awareness expands, then decelerates as the addressable audience saturates. The logistic curve captures this parsimoniously with a carrying capacity K (the effective ceiling for a title’s active peers), a midpoint x_0 , and a growth rate k . In this context, K can be interpreted as the maximum contemporaneous demand that a specific title can realistically attract on public trackers during its lifecycle. As the interested population is exhausted, growth naturally slows even if the swarm remains healthy Meyer et al. (1999).

Bass in the torrenting context The Bass model decomposes adoption into the two forces of innovation p and imitation q . Intuitively, p picks up exogenous shocks, such as a well-seeded release by a reputable group, front-page placement on trackers, or cross-site mirroring. q captures endogenous contagion where other peers join the swarm because “others are already downloading it.” This visibility can provide a type of social proof that the torrent is the “best” one when users lack strong prior information Park and McCallister (2023). In consumer technology, p and q jointly generate the familiar S-curve; in torrents, we expect p to be tightly linked to initial seeding and distribution, whereas q should matter only if discovery is social and capacity constrained.

This mapping is consistent with the mechanics of BitTorrent swarms. Early availability and *seed bandwidth* strongly shape observed performance, more seeds (and higher aggregate seed capacity) typically shorten download times and make the swarm more attractive to new peers Bharambe et al. (2005). These features mean a large fraction of early growth can be explained by supply-side “seeding shocks” rather than by strong P2P imitation.

Results Torrent adoption follows S-curves and the earliest well-seeded release enjoys a sharp first-mover advantage. Logistic curves fit 87% of hashes (median $R^2 \approx 0.71$), with median saturation $K \approx 318$ peers. Adding the Bass imitation term does not improve fit; the estimated imitation parameter is near zero for almost all hashes, which is consistent with growth driven by initial seeding and tracker visibility rather than peer-to-peer contagion. At the title level, the event study shows a discrete day-

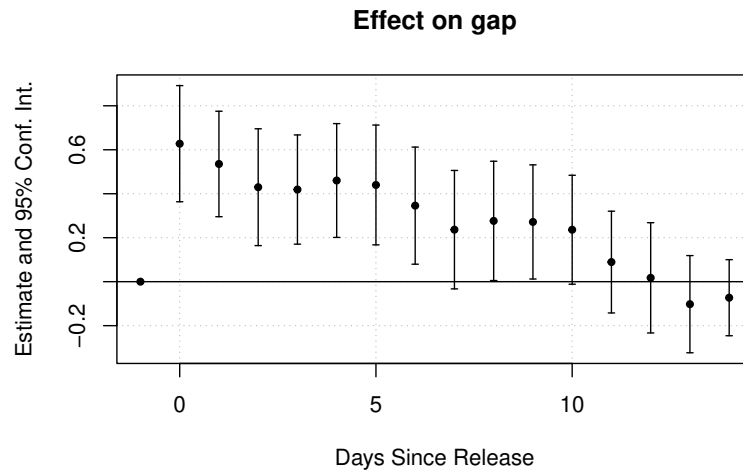


Figure 4. First-mover advantage in competing torrents. Points plot the estimated *difference in daily active-peer share* between the earliest well-seeded hash (“leader”) and later entrants for the same title from an event-study model that aligns time at the leader’s release (day 0). Vertical line marks day 0; bars show 95% confidence intervals. The leader’s share jumps on day 0, stays significantly higher for roughly the first week, and then attenuates toward zero within about two weeks, indicating a strong but transient early advantage.

0 jump in the leader’s active-peer share that remains statistically positive for roughly one week and tapers to indistinguishable from zero within about two weeks (Figure 4). These patterns support H3a: most hashes display bounded S-shaped growth well captured by a logistic curve. They do not support the stronger H3b expectation of meaningful endogenous imitation in the Bass sense: the near-zero imitation parameter indicates that early growth is better explained by supply-side seeding shocks and tracker visibility than by peer-to-peer contagion. The event-study evidence supports H4, showing a material but time-limited first-mover advantage.

Geography & Network Heterogeneity (H5–H6)

We ask whether internet capacity shapes *when* swarms first appear and *who* seeds once they do, and whether participation is concentrated in a small number of places and networks. Concretely, we relate country-level time-to-first appearance (TFA) to fixed broadband performance, and we model seeding at the session level while allowing for unobserved differences across titles and countries. We then quantify

concentration across country codes and ISPs.

Timing vs. Capacity Regressing the log median TFA by country on median fixed speeds shows a modest, intuitive relationship whereby countries with faster download capacity see titles earlier. A 50 Mbps increase in country-median download speed corresponds to about a 14% shorter TFA; upload speed is small and not significant, and overall fit is modest (as expected given unobserved policy and market factors). The takeaway is directional rather than deterministic: capacity helps titles surface sooner in a given jurisdiction, but it is only one piece of the timing puzzle.

Who seeds vs. who leeches At the peer level, we define a session as “seeding” if at least 50% of observed minutes are spent uploading. A logistic mixed model with random intercepts for a title and country finds no reliable association between local capacity and the odds of seeding once we account for those heterogeneities. The odds-ratios for download, upload, and latency are near 1 with wide, overlapping CIs. By contrast, the title-level effects dominate with random-effect variance for titles an order of magnitude larger than country-level variance ($sd \approx 2.41$ for titles vs. ≈ 0.77 for countries). In practical terms, *what is being shared* matters much more than a user’s internet connection in predicting if they stay in the swarm to seed. This pattern is consistent with behavioral mechanisms rather than a purely topological account of seeding. In this panel, many users appear to use BitTorrent mainly as a distribution channel rather than as a network to which they meaningfully contribute, unlike some P2P systems with stronger bidirectional incentive mechanisms.

Concentration and network concentration points Participation in swarms is highly concentrated across both jurisdictions and networks. The country-level Gini is high (0.84), with the top 10 countries accounting for roughly two-thirds of unique peers; the ISP-level Gini is even higher (0.89), with the top 10 ISPs near one-quarter of peers (because there are thousands of ISPs in total, inequality is sharper). This concentration, combined with the seeding results above, suggests that early public swarms are carried by a relatively small set of network concentration points. A small number of large access providers and hosting/transit players carry an outsize share of swarm activity for early torrent leaks of titles.

Pre-release Fakes & User Risk (H7)

Are “fake” torrents for a given title (e.g. filenames carrying .exe, .scr, keygen, README.url) more common before the initial release date? Are there specific targeted efforts around popular releases? To explore this, we join each hash to the title-level release date and compute the fraction of suspicious filenames by day relative to release.

We find that the share of suspicious files in the torrent swarms spikes pre-release. In our sample, roughly one-third of pre-release torrents flag as suspicious versus about two percent post release, a 14x increase. The day-by-day curve peaks in the short window prior to the official release date.

To further enrich this analysis of specific risks and attack vectors, we created a sandbox environment and analyzed six suspicious files with the open source platform CAPEv2 CAPE Sandbox Developers (2019). This environment allowed us to open the suspicious files and, when malware was present, detonate it and collect detailed forensic information. The virtual machine emulating the victim’s machine was Windows 10 22H2 Home edition.

These six files were selected as an illustrative case sample based on having both suspicious filenames and anomalous static characteristics (e.g., inflated file size, low entropy). This sandbox analysis demonstrates the types of behaviors that at least some pre-release impostor torrents exhibit and provides qualitative insight into attacker tactics. We therefore interpret this analysis as a forensic case study that complements our quantitative measurement of suspicious-file prevalence, rather than as an attempt to characterize malware families or population-wide distributions.

The static analysis of the files suggests that the files are intentionally inflated in size to imitate the expected file sizes of the titles. Very low entropy indicators show that the files are bloated with zero-padding so that they seem legitimate in size, one of the core heuristics users may use when deciding which torrent to download. Compression ratios in our sample range from about 800 to 1000x. The static analysis did not produce any hashes of known malware samples as indexed by Malware Bazaar⁵.

Dynamic analysis, where we detonated the malicious executables, yielded patterns consistent with Trojan behavior. All of the samples first conducted reconnaissance of the environment by running

⁵<https://bazaar.abuse.ch/>

locale and keyboard checks via the Windows registry. One of the samples clearly had a date gate, as once it checked the local time and date, it exited its own process. The other samples showed signs of unpacking and further injection, initiating a remote process on another thread that then lay dormant. Finally, in these samples with the follow-on process launch, the mouse was taken over, moved around the screen, and then settled in the lower left corner above the search bar. This is likely the ending of the initial takeover phase, after which the executable waits for further triggers. We were not able to meet those triggers and leave further dynamic analysis of this malware to future work.

Cross-Layer Synthesis & Policy Implications

How the two layers fit together

The macro panel shows that platform exclusivity is associated with earlier leaks: each additional legitimate provider platform reduces the instantaneous risk of a first torrent by $\approx 4.2\%$ (hazard ratio ≈ 0.959), and exclusives hit the 50% leak probability far sooner than widely distributed titles (≈ 120 vs. ≈ 275 days at covariate means). Volume, in contrast, scales with demand proxies (popularity, budget) rather than with the number of providers. Taken alone, these are substantively meaningful descriptive associations. Narrower availability is associated with earlier leak timing; demand proxies are more closely aligned with eventual torrent scale.

The micro panel explains what happens once a leak exists. Adoption inside swarms is shock-driven, not sustained by classical patterns of P2P imitation diffusion. Logistic curves fit the majority of hashes (median $R^2 = 0.71$; median saturation $K = 318$ peers), while the Bass imitation term is near zero across fits. Thus, the evidence supports H3a's bounded S-curve prediction but not H3b's stronger imitation claim. A first viable encode captures an immediate share lead on day zero that remains statistically positive for roughly a week and then attenuates by about week two. Geography matters for *timing*, as faster fixed broadband areas see earlier first appearances, but not for *who* is responsible for seeding. After accounting for title and country heterogeneity, the local internet capacity (upload, download, latency) does not reliably predict seeding propensity. Participation is highly concentrated (high Ginis in both country and ISP layers), revealing a small set of networks that shoulder an outsize share of early activity. Finally, pre-release fakes that contain malicious programs spike in the short window before official availability, with 14x higher suspicious-file rates than post-release.

The combined pattern does not show fragmentation mechanically raising the file-count of tor-

rents per title. Instead, the title-level and swarm-level evidence is consistent with a timing pathway where earlier leaks arrive in a high-attention release window where first-mover advantage, shock seeding, and network concentration shape early swarms and user risk.

The macro panel shows that some titles leak substantially earlier than others; the micro panel shows that once a leak exists, the first two weeks concentrate both user activity and malware risk. Cross-layer, this means that factors associated with earlier leaks also identify the period in which users are most exposed to malicious impostors.

User risk is concentrated in the pre-release window Joining swarms to release dates shows a sharp increase in suspicious executables and related lures immediately before official availability. We observe roughly one-third of pre-release torrents versus 2% post-release, and our malware analysis indicates behavior consistent with staged compromise. This aligns with the macro–micro story: fragmentation and staggered access can increase anticipation during the no-access period; early leaks meet that demand; the same window attracts impostor payloads that exploit users’ urgency.

What this means for how we think about BitTorrent

Taken together, these patterns imply that BitTorrent is less a monolithic P2P network and more a *loose federation of ad-hoc, title-centric distribution micro-networks*. Swarms spin up around specific releases, a small number of well-provisioned seeders account for much of early diffusion, and many leechers connect briefly, download, and churn out. That architecture aligns with our diffusion evidence of shock-driven growth and a clear first-mover advantage. Fragmented legitimate access is best understood here as a contextual source of media access friction rather than a directly identified cause of any individual leak; once a leak exists, the diffusion pattern resembles a concentrated distribution system with high churn more than a balanced, many-to-many mesh.

Alternative explanations and boundary conditions The macro associations are observational. Platform availability is chosen by studios and may correlate with unobservables such as marketing intensity, scene attention, or enforcement salience. We condition on demand/scale proxies and document robustness, but we do not claim structural causal effects. The micro panel covers a one-month window of recent releases and focuses on public trackers and the DHT; as such it may overweight

high-salience titles and under-represent closed ecosystems or private trackers. The Reddit evidence is similarly scoped: r/Piracy is a large public gateway community, but it is not representative of all piracy communities and can be affected by Reddit demographics, moderation, posthoc archive coverage, and displacement to Discord, Telegram, private trackers, or non-English forums. Country codes reflect IP geolocation and include territories as defined by ISO-3166; shares are computed over unique peers, not bytes. These caveats do not overturn the central pattern that earlier opportunity, shock-driven diffusion, and concentrated provisioning jointly shape early torrent risk, but they do help to properly scope this work.

Implications for Policy

P2P technologies are **dual-use** and rely on public internet routing. Heavy-handed Layer 3 censorship or blanket targeting of P2P traffic risks collateral damage and civil-liberty concerns. Our results do not, by themselves, identify or justify a specific enforcement intervention. They do show that the relevant measurement window is narrow and title-specific as early activity is concentrated, first-mover advantages fade within roughly two weeks, and a small set of networks carries an outsize share of peers. For policymakers, the descriptive implication is that any intervention should be narrow, evidence-triggered, time-limited, and attentive to collateral effects rather than framed as broad protocol- or Layer-3 blocking.

Where enforcement cooperation is considered, the evidence points toward verified infringing swarms, aggregate network-level indicators, procedural safeguards, and automatic sunsets rather than standing restrictions on a protocol or access network. In parallel, reducing media access friction through broader legitimate availability remains a lower-collateral avenue for stakeholders to evaluate when seeking to reduce the demand conditions under which early leaks, large early swarms, and pre-release impostor files become most salient.

Conclusion

Combating the illegal consumption of copyrighted media is an ongoing challenge. In this paper we present a two-panel descriptive study of how legitimate availability, release timing, and P2P swarm behavior relate to early media torrent leaks. The evidence should be read descriptively. The macro panel shows that narrower legitimate availability is associated with earlier first torrent appearance, while post-leak torrent volume is better explained by demand and scale proxies than by provider count. The

micro panel then shows why timing matters, as early swarms grow quickly, first viable releases enjoy a short-lived advantage, participation is concentrated across networks, and pre-release impostor files create direct user risk. At the same time, our Reddit evidence is contextual rather than representative, and the one-month micro panel is a targeted case sample of high-salience releases. These limitations define the scope of the claims rather than undermining the central measurement contribution that cross-layer observation can reveal when illicit supply emerges, how demand materializes inside swarms, and where user risk concentrates during the most time-sensitive release window.

Acknowledgments

The authors thank the editors and anonymous reviewers for their constructive feedback.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

- Berns, A. D. and Jung, E. (2008). Searching for malware in bittorrent. *University of Iowa Computer Science Technical Report*, pages 1–10.
- Bharambe, A. R., Herley, C., and Padmanabhan, V. N. (2005). Analyzing and improving bittorrent performance. *Microsoft Research, Microsoft Corporation One Microsoft Way Redmond, WA, 98052:2005–03*.
- Blackburn, D., Eisenach, J. A., and Harrison, Jr., D. (2019). Impacts of digital video piracy on the U.S. economy. Technical report, Global Innovation Policy Center (U.S. Chamber of Commerce) and NERA Economic Consulting, Washington, DC.
- CAPE Sandbox Developers (2019). Capev2: Malware configuration and payload extraction. <https://github.com/kevoreilly/CAPEv2>.
- Cardoso, B., Abreu, J., and Prata, A. (2022). The tv ecosystem is fragmented and in need of new business and technological strategies. In *Proceedings of the 2022 ACM International Conference on Interactive Media Experiences*, pages 133–140.

- Cuevas, R., Kryczka, M., Cuevas, A., Kaune, S., Guerrero, C., and Rejaie, R. (2010). Is content publishing in bittorrent altruistic or profit-driven? In *Proceedings of the 6th International Conference*, pages 1–12.
- Cuevas, R., Kryczka, M., González, R., Cuevas, A., and Azcorra, A. (2014). Torrentguard: Stopping scam and malware distribution in the bittorrent ecosystem. *Computer Networks*, 59:77–90.
- Danaher, B., Hersh, J., Smith, M. D., and Telang, R. (2020). The effect of piracy website blocking on consumer behavior. *MIS Quarterly*, 44(2):631–660.
- David, M. (2017). *Sharing: crime against capitalism*. John Wiley & Sons.
- Goldman, E. (2005). The challenges of regulating warez trading. *Social Science Computer Review*, 23(1):24–28.
- Hazel, G. and Norberg, A. (2008). Bep 9: Extension for peers to send metadata files. https://www.bittorrent.org/beps/bep_0009.html. Last modified 2017-03-26; accessed 2025-08-31.
- Huizing, A. and van der Wal, J. A. (2014). Explaining the rise and fall of the warez mp3 scene: An empirical account from the inside. *First Monday*.
- Kim, S., Han, J., Chung, T., Kim, H.-c., Kwon, T. T., and Choi, Y. (2012). Content publishing and downloading practice in bittorrent. In *International Conference on Research in Networking*, pages 97–110. Springer.
- Kryczka, M., Cuevas, R., Guerrero, C., Azcorra, A., and Cuevas, A. (2011). Measuring the bittorrent ecosystem: Techniques, tips, and tricks. *IEEE Communications Magazine*, 49(9):144–152.
- Lee, B., Jeong, S., and Paek, S. Y. (2019). Determinants of digital piracy using deterrence, social learning and neutralization perspectives. *International Journal of Comparative and Applied Criminal Justice*, 43(4):295–308.
- Lee, B., Paek, S. Y., and Fenoff, R. (2018). Factors associated with digital piracy among early adolescents. *Children and Youth Services Review*, 86:287–295.
- Loewenstern, A. and Norberg, A. (2008). Bep 5: Dht protocol. https://www.bittorrent.org/beps/bep_0005.html. Last modified 2020-01-21; accessed 2025-08-31.

- Lotz, A. D. (2021). Unpopularity and cultural power in the age of netflix: New questions for cultural studies' approaches to television texts. *European Journal of Cultural Studies*, 24(4):887–900.
- Lotz, A. D. (2022). *Netflix and streaming video: The business of subscriber-funded video on demand*. John Wiley & Sons.
- Lotz, A. D. and Eklund, O. (2024). Beyond netflix: Ownership and content strategies among non-us-based video streaming services. *International Journal of Cultural Studies*, 27(1):119–140.
- Mantena, R., Sankaranarayanan, R., and Viswanathan, S. (2010). Platform-based information goods: The economics of exclusivity. *Decision Support Systems*, 50(1):79–92.
- Martinez, M. F. (2025). Torrentmonitor. <https://github.com/N4rr34n6/Torrent-Monitor>. GitHub repository; AGPL-3.0 license; commit 43a86ba (2025-04-14).
- Meyer, P. S., Yung, J. W., and Ausubel, J. H. (1999). A primer on logistic growth and substitution: the mathematics of the loglet lab software. *Technological forecasting and social change*, 61(3):247–271.
- Morris, R. G. and Higgins, G. E. (2010). Criminological theory in the digital age: The case of social learning theory and digital piracy. *Journal of Criminal Justice*, 38(4):470–480.
- Nasta, L. (2025). Access versus exclusivity in the digital creative and cultural landscape. In *Navigating the Paradoxes of Digital Transformation in the Creative and Cultural Industries: Embracing Innovation*, pages 23–52. Springer.
- Ookla (2025). Speedtest by ookla global fixed and mobile network performance maps. <https://registry.opendata.aws/speedtest-global-performance/>. Accessed 2025-08-15.
- Or-Meir, O., Nissim, N., Elovici, Y., and Rokach, L. (2019). Dynamic malware analysis in the modern era—a state of the art survey. *ACM Computing Surveys (CSUR)*, 52(5):1–48.
- Park, S. and McCallister, J. (2023). The effects of social proof marketing tactics on nudging consumer purchase. *Journal of Student Research*, 12(3).
- Phau, I., Lim, A., Liang, J., and Lwin, M. (2014). Engaging in digital piracy of movies: a theory of planned behaviour approach. *Internet Research*, 24(2):246–266.

- Poort, J., Leenheer, J., Van der Ham, J., and Dumitru, C. (2014). Baywatch: Two approaches to measure the effects of blocking access to the pirate bay. *Telecommunications Policy*, 38(4):383–392.
- Sandvine (2023). Global internet phenomena report 2023. Technical report, Sandvine. Accessed 2025-08-31.
- Schäfer, J. and Malinka, K. (2009). Security in peer-to-peer networks: Empiric model of file diffusion in bittorrent. In *2009 Fourth International Conference on Internet Monitoring and Protection*, pages 39–44. IEEE.
- Seidenberger, S., Pursell, N., and Maiti, A. (2025). Magnetdb: A longitudinal torrent discovery dataset with imdb-matched movies and tv shows. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 19, pages 2575–2586.
- Seifert, R., Otten, C., Clement, M., Albers, S., and Kleinen, O. (2023). Exclusivity strategies for digital products across digital and physical markets. *Journal of the Academy of Marketing Science*, 51(2):245–265.
- Seng, D. K. B. (2015). 'who watches the watchmen': An empirical analysis of the reasons for rejecting copyright take-down notices. *Available at SSRN 3687861*.
- The Movie Database (2025). TMDb API – Version 3. Data accessed 17 March 2025.
- Urban, J. M., Karaganis, J., and Schofield, B. (2017). Notice and takedown in everyday practice. Public Law Research Paper 2755628, UC Berkeley School of Law.
- Wang, L. and Kangasharju, J. (2013). Measuring large-scale distributed systems: case of bittorrent mainline dht. In *IEEE P2P 2013 Proceedings*, pages 1–10. IEEE.
- Wolchok, S. and Halderman, J. A. (2010). Crawling {BitTorrent}{DHTs} for fun and profit. In *4th USENIX Workshop on Offensive Technologies (WOOT 10)*.
- Zhang, B., Iosup, A., Pouwelse, J., Epema, D., and Sips, H. (2010). Sampling bias in bittorrent measurements. In *European Conference on Parallel Processing*, pages 484–496. Springer.

Appendix

Responsible Measurement

Our swarm monitoring recorded IP addresses only transiently in memory to join and enumerate peers; all persistent storage used one-way hashes and country/ASN-level aggregation. We did not attempt to deanonymize peers or link them to individuals, and we release only aggregate statistics at the country and ISP level. Following institutional guidelines, this work does not constitute human-subjects research because it observes publicly routable endpoints engaged in P2P file-sharing. We explicitly discuss potential risks of enabling targeted enforcement against specific ISPs or jurisdictions and limit the granularity of any public data release to mitigate such harms.

Micro-Panel Sampling Protocol

The micro panel is a targeted release-window sample designed for depth of swarm observation rather than population representativeness. Candidate titles were identified from IMDb and JustWatch release calendars for 14 July–15 August 2025. We retained titles with clear release timing, high expected public salience, unambiguous title strings, and observable public-DHT activity through the live MagnetDB query process. We excluded titles when names could not be matched reliably, release dates were ambiguous, no public DHT hit appeared during the monitoring window, or the available activity appeared outside the public swarms our measurement infrastructure could responsibly observe. Continuous monitoring required persistent swarm participation, repeated peer enumeration, geolocation enrichment, and de-identification of sensitive network identifiers, which limited feasible breadth. The resulting 28-title panel should therefore be interpreted as a purposive sample of high-salience new releases.