UNIVERSIDADE ESTADUAL DE MARINGÁ
DEPARTAMENTO DE FÍSICA

Mônica Valério Prates

Structure and behavioral dynamics in a dark web child sexual abuse network

Maringá, 14 de abril de 2025.

UNIVERSIDADE ESTADUAL DE MARINGÁ
DEPARTAMENTO DE FÍSICA

Mônica Valério Prates

Structure and behavioral dynamics in a dark web child sexual abuse network

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Física da Universidade Estadual de Maringá.

Orientador: Prof. Dr. Haroldo Valentin Ribeiro

Maringá, 14 de abril de 2025.

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá - PR, Brasil)

P912s

Prates, Mônica Valério

Structure and behavioral dynamics in a dark web child sexual abuse network / Mônica Valério Prates. -- Maringá, PR, 2025.

61 f. : il. color.

Orientador: Prof. Dr. Haroldo Valentin Ribeiro.

Dissertação (mestrado) - Universidade Estadual de Maringá, Centro de Ciências Exatas, Departamento de Física, Programa de Pós-Graduação em Física, 2025.

1. Redes criminais. 2. Ciência de redes. 3. Sistemas complexos. 4. Cibercrime. 5. Redes bipartidas. I. Ribeiro, Haroldo Valentin, orient. II. Universidade Estadual de Maringá. Centro de Ciências Exatas. Departamento de Física. Programa de Pós-Graduação em Física. III. Título.

CDD 23.ed. 530

Síntique Raquel de C. Eleutério - CRB 9/1641

MÔNICA VALÉRIO PRATES

STRUCTURE AND BEHAVIORAL DYNAMICS IN A DARK WEB CHILD SEXUAL ABUSE NETWORK

Dissertação apresentada à Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de mestre.

Aprovado em: Maringá, 14 de abril de 2025.

BANCA EXAMINADORA

_____

Prof. Dr. Haroldo Valentin Ribeiro
Universidade Estadual de Maringá – UEM

_____

Prof. Dr. Marcelo Kaminski Lenzi
Universidade Federal do Paraná – UFPR

_____

Prof. Dr. Angel Akio Tateishi
Universidade Tecnológica Federal do Paraná – UTFPR/Pato Branco

_____

Prof. Dr. Ervin Kamiski Lenzi
Universidade Estadual de Maringá – UEM

Acknowledgments

I gratefully acknowledge the support of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Programa de Cooperação Acadêmica (PROCAD) - Segurança Pública e Ciências Forenses for awarding the Master's scholarship.

Título em português:

Estrutura e dinâmica comportamental em uma rede de abuso sexual infantil na dark web

Resumo

Esta dissertação de mestrado investiga a estrutura, o comportamento e a evolução de um fórum *online* de abuso sexual infantil hospedado na *dark web*, com base em dados privilegiados obtidos por meio da Operação Darknet, uma investigação de longo prazo conduzida pela Polícia Federal do Brasil. Utilizando uma abordagem que integra teoria de sistemas complexos, ciência de redes e contribuições da criminologia, analisamos como os usuários interagiam entre si e com os conteúdos ilícitos ao longo do tempo. Ao modelar o fórum como uma rede bipartida composta por usuários e postagens, identificamos diversas características estruturais, como assimetria extrema de atividade, organização hierárquica e formação de comunidades ou estruturas modulares na rede. Verificamos que aproximadamente 31% dos usuários foi responsável pela maior parte do conteúdo e das interações, enquanto a maioria dos participantes teve um envolvimento passivo, atuando principalmente como consumidores. Observamos também que, à medida que se expandiu, o fórum desenvolveu subgrupos cada vez mais complexos e especializados, refletindo a segmentação temática e a evolução das preferências dos usuários. Nossos resultados revelaram ainda que usuários altamente ativos tendiam a evitar as postagens mais populares, indicando padrões de interação disassortativa e o surgimento de núcleos isolados de conteúdo. Além disso, os indivíduos identificados e capturados pelas autoridades policiais apresentaram assinaturas comportamentais distintas, incluindo frequências de acesso mais elevadas, maior duração de atividade e inserção mais profunda em subcomunidades hierárquicas específicas. Esses padrões sugerem que propriedades estruturais e comportamentais podem ser utilizadas para detectar usuários de alto risco e priorizar ações investigativas. Ao combinar dados empíricos com métodos de redes complexas, nosso trabalho contribui para a compreensão do comportamento criminal em ambientes *online* e oferece subsídios para uma melhor compreensão da estrutura e funcionamento de redes de abuso sexual infantil.

Palavras-chave: Redes criminais. Ciência de redes. Comunidades ilícitas *online*. Sistemas complexos. Redes bipartidas. *Dark web*. Cibercrime.

Abstract

This dissertation investigates the structure, behavior, and evolution of an online child sexual abuse forum hosted on the dark web, drawing on privileged data obtained through Operation Darknet, a long-term investigation conducted by the Brazilian Federal Police. Using a framework that integrates complex systems theory, network science, and criminology insights, we analyzed how users interacted with one another and with illicit content over time. By modeling the forum as a bipartite network of users and posts, we identify key structural features, such as extreme activity asymmetry, hierarchical organization, and modular community formation. Approximately 31% of users generated most content and interactions, while most participants engaged passively, primarily as consumers. As the forum expanded, it developed increasingly complex and specialized subgroups, reflecting thematic segmentation and evolving user preferences. Our results reveal that highly active users tended to avoid the most popular posts, indicating disassortative interaction patterns and the emergence of isolated content hubs. Moreover, individuals identified and apprehended by law enforcement exhibited distinct behavioral signatures, including higher access frequencies, longer durations of activity, and deeper embedding within specific hierarchical subcommunities. These patterns suggest that structural and behavioral markers can be used to detect high-risk users and prioritize investigative efforts. By combining empirical data with advanced network modeling, this work contributes to the understanding of criminal behavior in online environments and offers insights for better understanding the structure and functioning of child sexual abuse networks.

Keywords: Criminal networks. Network science. Illicit online communities. Complex systems. Bipartite networks. Dark web. Cybercrime.

Contents

Introduction

Complex systems encompass a variety of phenomena, both in natural and social contexts, characterized by a number of fundamental interconnected relationships within their own structures, as well as interaction with the external environment [1–4]. Complex networks – in which nodes denote entities and edges are the relationships between them – serve as powerful tools to represent complex systems, providing models for interactions and techniques for analyzing patterns and dynamics [5, 6], thereby enhancing our understanding of system behavior and enabling predictions. Although a general definition of complexity remains elusive [7], the intricate network connectivity inherent to these systems significantly influences their behavior, underscoring that a solid grasp of network theory is essential for any comprehensive approach to studying complex systems [8].

The mathematical study of networks, known as graph theory, originated with Leonhard Euler's solution to the Königsberg bridge problem in the 18th century [9]. While initial research in the field was predominantly theoretical, empirical analysis of networks emerged in the 1930s, primarily through studies of social networks. Since the late 1990s, the field has experienced remarkable growth, largely driven by physicists who have significantly contributed to foundational research on the statistical properties of networks [10]. Over the past decade, discoveries highlighting the universality of various network topological characteristics have facilitated its expansion into diverse disciplines [11]. Reflecting its inherently multidisciplinary nature, network theory has been employed to analyze a myriad of systems, including the neuronal network of the nematode worm *C. elegans* [12], ecological food webs [13], scientific collaboration networks [14], and individuals involved in political scandals [15, 16].

In particular, applying network science to social systems in the context of crime – with nodes representing individuals or organizations involved and edges representing their social interactions – is an effective way of illustrating their structure and dynamics. For example, analyses of relationships within mafia groups, such as the Sicilian Mafia, have shown that leaders exhibit the lowest degree of centrality, indicating a strategic distancing from central network positions [17–19]. Similarly, a comprehensive study examining a criminal intelligence network compiled by the Brazilian Federal Police, encompassing records of 23,666 individuals involved in crimes such as drug and arms trafficking, bank robbery, and money laundering, demonstrated that its largest connected component displays critical social network characteristics, including small-world properties and a heavy-tailed degree distribution [20]. Additionally, this network component exhibited high modularity, rendering it susceptible to targeted, module-based attacks; indeed, removing approximately 2% of its nodes or edges was sufficient to disrupt the network significantly.

These findings align with numerous recent studies exploring the structure and dynamics of organized crime [20, 21], political corruption networks [15, 22–24], and various offenses committed through individual associations [25–29]. The emerging field of criminal network analysis, significantly bolstered by collaboration between law enforcement and academia, continues to reveal structural insights with substantial implications for theoretical and practical domains. For instance, empirical assessments of law enforcement interventions, such as the disruption of a dark web pedophile network, have demonstrated that interventions resembling random node removals are notably less effective than optimized strategies targeting critical nodes [26]. Moreover, recent research has demonstrated the efficacy of combining criminal network structures with machine learning methodologies, achieving high accuracy in reconstructing missing criminal connections, classifying criminal associations, predicting financial flows among criminals, and forecasting re-offending behaviors [30, 31]. Studies analyzing security intelligence data have also introduced innovative protocols for disrupting networks by identifying key agents, substantially enhancing law enforcement's ability to detect and respond to criminal activity [32, 33]. Additionally, recent analysis of user behaviors within dark web criminal networks has provided critical insights into temporal activity patterns, content preferences, and clustering tendencies among offenders [34].

Collectively, these studies underscore that leveraging network analysis can reveal hidden structures, interactions, and vulnerabilities within criminal networks. They significantly

enhance our understanding of the roles, strategies, and underlying mechanisms of organized crime, simultaneously contributing to theoretical criminological knowledge and informing practical law enforcement strategies for more effective interventions against illicit enterprises. Extending beyond traditional criminal organizations, network analysis is also applicable to cybercrimes occurring on the dark web.

The Internet can be divided into three distinct layers: the surface web, the deep web, and the dark web. The surface web refers to the portion of the internet that is accessible to the general public and indexed by standard search engines (*e.g.*, Google). Despite its visibility, it represents only about 5% of the total online content. Beneath it lies the deep web, which comprises approximately 90% of all internet traffic. The dark web is a small part of the deep web, and it requires specific software to access. This environment is mostly used for illicit activities, such as online drug and weapons markets and the distribution of child sexual abuse material[1] (CSAM) [35].

The dark web is primarily accessed through The Onion Router (TOR), which conceals users' IP addresses by routing traffic through multiple random servers, making the original address practically impossible to trace [36]. TOR sites are identifiable by their *.onion* URLs and are commonly referred to as TOR hidden services, which are not indexed by public search engines. When using TOR, data transmission is encrypted and inaccessible to third parties, allowing users to remain fully anonymous and untraceable. Consequently, the dark web has become a particularly attractive platform for individuals with illegal intentions.

In this context, the 2023 annual report published by the Internet Watch Foundation (IWF) highlighted the alarming prevalence of CSAM online [37], defined as content depicting or describing child abuse. This material encompasses photographs, videos, audio files, texts, drawings, and any other media form involving or addressing such offenses. Among the hidden services reviewed by the IWF responsible for distributing CSAM in 2023, 74% were identified as commercial platforms dedicated to the sale of such material. On the surface web, of the 392,665 web pages assessed, approximately 70% contained, linked to, or advertised CSAM – an 8% increase compared to the 2022 report [37]. The report also revealed that 97% of the depicted children were girls, alongside a disturbing downward trend in victims'

[1]The term *child pornography* can be misleading, as it may imply that the child consented to participate in the production of sexual content. In contrast, the term *child sexual abuse material* is more appropriate, since it conveys the abusive and non-consensual nature of the content, removing any suggestion of voluntariness on the part of the child.

ages. In 2021, 68% of URLs featuring CSAM involved children aged 11-13 years, while 23% involved children aged 7-10 years. By 2023, these proportions shifted to 54% for ages 11-13 and 41% for ages 7-10. This issue is especially concerning on the dark web, where CSAM sites are reported to be nearly 2,000 times more prevalent than on the surface web [36]. A six-month data collection from TOR hidden services revealed an average daily activity of 45,000 sites, with approximately 16% related to drug marketplaces and around 2% associated explicitly with child abuse. In contrast, more than 80% of search queries were seeking CSAM [38]. Additionally, a comprehensive five-year investigation into 176,683 TOR *.onion* domains found that 20% (35,337) were associated with CSAM [39]. Furthermore, 11.1% of search sessions were related to CSAM, with queries such as "child porn" being among the most frequent. The study also analyzed the most visited TOR search engines, discovering that nearly 81% returned results related to CSAM, with some actively advertising this material. Additionally, 40.5% of users searched specifically for content involving children aged 11 years or younger, underscoring an unsettling trend of online offenders targeting increasingly younger victims.

Contrary to common assumptions, accessing the dark web requires minimal technical expertise. For instance, during the FBI investigation known as Operation Pacifier into "Playpen," one of the largest dark web child sexual abuse (CSA) forums, investigators discovered that the individual arrested and charged with creating and administering the forum was a 58-year-old owner of an excavation trucking company who possessed no formal technical training [40]. As the anonymity provided by the dark web reduces users' fear of exposure, numerous forums have emerged as hubs for sharing and exchanging CSAM. Although relatively uncommon, previous research has identified that these forums typically exhibit hierarchical structures based on members' importance and contributions, with designated roles such as administrators, moderators, and VIP members, among others [36]. Similar to Playpen, these forums constitute primary platforms for individuals seeking CSAM online. At present, research on the networked structures of dark web CSA forums remains limited, and the actual impact of law enforcement interventions is not yet thoroughly documented. This research gap primarily arises from the illegal nature of the content hosted on these platforms, making it inaccessible to academic researchers unless conducted in collaboration with law enforcement agencies.

Upon joining a CSA forum, members encounter various categories subdivided into spe-

cific topics, typically organized according to the age and gender of child victims, as well as the types of acts perpetrated against them. Users can comment on posts and express approval through acknowledgments, such as a "thanks" (similar to a "like" on social media). Analyzing the popularity of topics and the number of thread views offers valuable insights into the behavior of online sexual offenders, including their preferences and patterns of community engagement. For example, a study found significant variations in attention based on average view counts, with threads containing photos or videos attracting greater user interest compared to text-based posts [41]. Additionally, posts specifying victim ages between 6 and 12 years old, along with terms such as "incest," received higher average views. Posts with explicitly sexualized and abusive titles garnered the highest levels of attention, underscoring a pronounced interest in hardcore acts involving children. Furthermore, Operation Darknet, conducted by the Brazilian Federal Police, investigated strategies for network disruption by analyzing data collected from a dark web CSA forum monitored between 2013 and 2014 [42]. The study concluded that achieving complete network dismantling would require removing approximately 60% of core users. The operation also identified a significant positive rich-club effect – a measure of connectivity among nodes exceeding a certain degree threshold. This positive effect typically signals cooperation among individuals; in this context, it implies that influential, high-degree members tend to interact preferentially with each other.

A study comparing the use of public forums (*e.g.*, Reddit) with dark web forums (*e.g.*, Pedo Support Community) found that threads are significantly shorter in dark web forums. This suggests reluctance among most users to leave digital footprints on the dark web [43]. By constructing an interaction network – where nodes represent forum members, and directed edges connect individuals replying to others' posts, with edge weights corresponding to the number of replies – they also observed that the degree distribution in public forums is more homogeneous compared to dark web forums, as users tend to have similar activity levels. In contrast, approximately 80% of nodes in the analyzed dark web forums had a total degree equal to one, indicating that high activity is concentrated among only a few users. This appears to be a common characteristic of dark web forums, as documented in existing literature. Indeed, previous research identified a small subset of highly active members in another dark web CSA forum who disproportionately contributed to the overall activity. Of the 14,838 users studied, only 0.7% (109 individuals) contributed over 500 posts, collectively accounting for 40% of total forum activity [44]. Similarly, an investigation into the "lurkers" of

a CSA forum – users who observe, browse, and consume content silently without engaging in visible interactions – revealed that, among the 417,438 registered members, only 3.4% (14,088 individuals) actively engaged verbally within the forum [45]. Nevertheless, the same study found that 93.6% of users had attempted to download CSAM at least once, demonstrating that although the majority refrains from openly participating, they actively consume CSAM. An analysis of the top three dark web forums with the highest volume of CSAM-containing posts identified 10,490 unique CSAM creators [46]. Remarkably, within each forum, the most active user accounted for an alarming 91%, 99%, and 44% of all posts, respectively [46]. These findings highlight the prevalence of a small number of influential members who play a critical role in content dissemination and sustaining these forums' existence and growth. Such individuals are particularly significant to law enforcement agencies, as they not only facilitate the online distribution of CSAM but also actively encourage participation among other forum members [34].

Motivations for joining a CSA forum may extend beyond satisfying one's sexual interest in children; they can also offer validation and a sense of social belonging within a community where crime and sexual deviance are endorsed and normalized. New members are frequently greeted with welcoming messages from other users, reinforcing feelings of acceptance and shared purpose [47]. Additionally, receiving positive feedback on posts, such as comments or likes, can further encourage engagement and contribution to the forum. Research has shown that prolonged exposure is positively correlated with increased user activity, thus fostering ongoing participation [48]. Given the dark web's appeal to individuals seeking anonymity, including those involved in criminal activities, law enforcement agencies increasingly face growing online communities of pedophiles, where participants can communicate, share, and sell CSAM on an unprecedented global scale. As many of these forums host hundreds of thousands of members, it is impractical for authorities to identify each user individually. Consequently, there is a pressing need to prioritize individuals who pose the greatest potential threat, such as those involved in CSAM production, forum administrators, and members who occupy central positions within the interaction networks. For instance, an international operation led by German law enforcement in 2021, in coordination with several other police agencies, shut down the dark web CSA forum known as "Boystown," which had over 400,000 registered members [49]. Among the four men apprehended, three were forum administrators, and the fourth was identified as one of the platform's most prolific contributors, having

authored over 3,500 posts.

Given the urgent need for more effective investigative strategies in such environments, this work aims to elucidate the behavioral patterns of users within these forums using network and statistical analysis and building on a partnership with the Brazilian Federal Police, which granted secure access to information collected during Operation Darknet – a long-term investigation that monitored users in a CSA forum. In doing so, it enhances our understanding of online CSAM consumption and facilitates the identification of key participants. The remainder of this dissertation is organized into three chapters. The first presents the data and explores the demographic characteristics of the forum. The second investigates user and post networks, examining their basic topological properties, modular structure, and how they evolved over time. Finally, the third chapter concludes this dissertation with a summary of findings and perspectives on future research directions.

CHAPTER 1

Demographic characteristics of a child sexual abuse forum

This chapter explores the demographic and behavioral characteristics of an online forum dedicated to child sexual abuse (CSA), which was covertly monitored by the Brazilian Federal Police as part of a long-term investigation named Operation Darknet (2013–2016). The analysis reveals a highly asymmetric structure, where a small group of users generated most of the content and interactions while the majority remained passive consumers. User activity exhibited diverse patterns, ranging from short-term, intensive engagement to sustained, habitual use over time. Forum content was classified into multiple categories, with certain types – particularly those linked to real-world offending – receiving heightened attention. Despite the anonymity offered by the platform, interaction metrics suggest varying levels of user caution and participation, especially in response to more extreme or incriminating material. The findings shed light on the forum's internal organization, the distinct roles adopted by its members, and the social dynamics that sustained the community, offering critical insights for understanding and disrupting such online environments.

1.1 Operation Darknet

Between 2013 and 2016, the Brazilian Federal Police conducted Operation Darknet [50, 51], one of the first investigations in Brazil to delve into the deep web, with the primary

objective of identifying users on the Tor network who trafficked CSA material. Authorized through a judicial order, the operation employed an innovative tool developed by the Brazilian Federal Police, which was capable of detecting users sharing videos and photos containing CSA content. Largely coordinated under the 11th Federal Court in Porto Alegre, the investigation represented a pioneering approach in terms of both technical methodology and legal strategy.

Initially, the police established a CSA forum on the dark web in October 2013, beginning with an opening post that questioned the sexual age of individuals. After this introductory message, law enforcement refrained from further direct intervention, opting instead for passive monitoring of the forum's growth and user behavior. By the end of the surveillance period in October 2014, the forum had expanded to include 25,943 posts, attracted over 6 million views, and was frequented by 9,614 registered members. These users engaged in activities that included sharing CSA content as well as posting opinions and personal experiences. Although the forum was predominantly in Portuguese, it also contained a few posts in other languages and attracted users from countries other than Brazil. In addition, all forum members were male, which is consistent with prior findings indicating that males are the predominant gender of child sexual abuse material (CSAM) consumers [52, 53].

The operation unfolded in two distinct phases. The first phase was officially launched on October 15, 2015 with the execution of more than 100 search and seizure warrants nationwide. This phase led to 51 arrests and the seizure of digital media, with Federal Police experts conducting on-site forensic examinations of the recovered evidence. In many instances, additional media were collected for later analysis, a necessary measure to uncover videos or images that may have been partially erased or concealed. In the second phase of Operation Darknet, approximately 70 additional suspects were targeted. This phase extended across 37 Judicial Subsections spanning 17 states, including further operations in Porto Alegre. Search and seizure warrants were executed in these diverse locations, and the seized digital media remains under forensic examination. Overall, Operation Darknet is recognized as a landmark investigation by the Brazilian Federal Police that not only broke new ground in the realm of digital forensics and legal process within the deep web but also set a precedent for subsequent operations targeting digital crimes involving child sexual abuse material.

Through our partnership with the Brazilian Federal Police, established via the "PROCAD – Segurança Pública e Ciências Forenses" initiative, we obtained detailed information on

the CSA forum monitored during Operation Darknet. These data comprise comprehensive internal statistics on forum usage. It is important to note that all original data and media were handled exclusively by Brazilian Federal Police agents with the requisite legal clearance. Data processing complied with the Brazilian Law for Data Protection (Act No. 13709 of 2018), the Brazilian Individual Rights Act (Brazilian Constitution of 1988), the Brazilian Criminal Code (Act No. 2848 of 1940), the Brazilian Criminal Procedure Code (Act No. 3689 of 1941), and Brazilian Federal Police internal procedures. Moreover, all data were anonymized prior to being provided to the authors.

This privileged access enabled us to elucidate the behavioral patterns of users within these environments, offering rare and valuable insights into the dynamics of CSA-related online communities on the dark web, as detailed in the following sections and chapters.

1.2 Growth dynamics of the forum

The data analyzed comprised detailed records of user activity, including counts of post views, likes, and comments, as well as the corresponding date and time of each event. All forum content was originally classified by users into 81 distinct categories, which were later analyzed by law enforcement and organized into 11 groups based on the content's nature, as follows:

- Uncategorized CSAM;
- General topics, rules, and suggestions;
- Texts and drawings;
- Non-nude;
- Jailbait;
- Toddler;
- PedoMOM;
- Bondage/hurtcore;
- Boylover;
- ZooPedo;
- and Potential offenders.

Each group refers to users who were likely engaged in the active perpetration of sexual abuse against a child.

The evolution of forum membership is depicted in Figures 1.1A and 1.1B. On the day the forum became publicly available, 95 users registered; within a month, the platform had attracted over 1,000 members. This rapid surge in user registrations reached a peak around August 2014, with 150 new members joining in a single day. Figure 1.1B further highlights the forum's sustained and growing popularity, indicating continued interest in its content over time. However, in October 2014, members discovered that the forum was under police surveillance, after which no additional registrations were recorded. Additionally, a 40-day gap in the data between late December 2013 and January 2014 suggests that the forum may have been inactive, potentially due to technical issues.

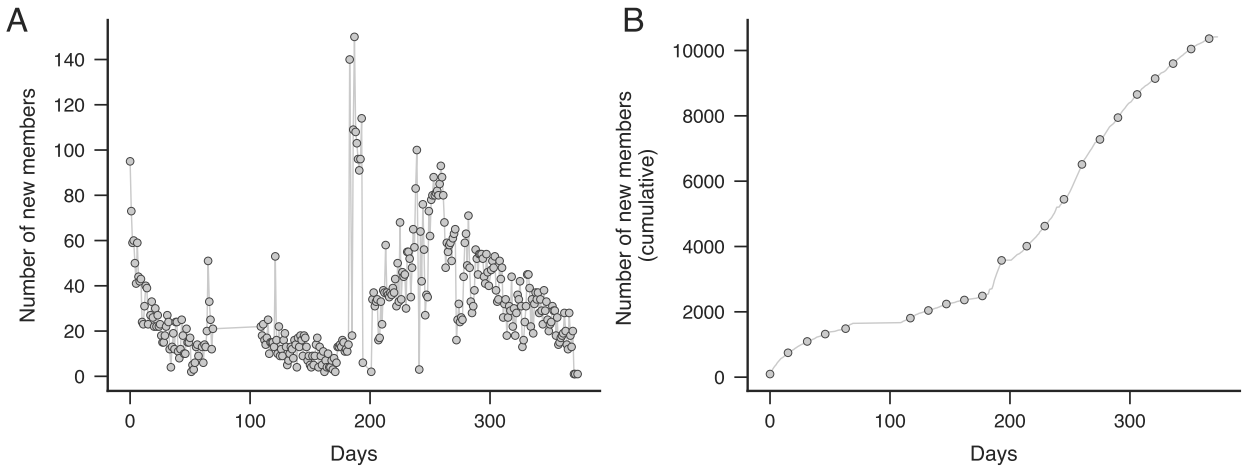

Figure 1.1: Number of new registered forum members from October 2013 to October 2014. Panel (A) displays the daily number of new members, while Panel (B) shows the cumulative total of new members over the same period.

To better understand the dynamics of user interaction, Figures 1.2A-1.2D present the temporal counts of forum activities. The forum's early stages were characterized by a rapid increase in the number of daily accesses, followed by a period of significant growth which was marked by notable spikes in user activity, with daily accesses surpassing 50,000 at its peak. These spikes likely corresponded to specific events, such as the release of new CSA content. Furthermore, post comments serve as a key indicator of user engagement, reflecting whether the forum fostered active discussions and interactions. While the number of daily comments remained relatively low overall, there was a notable spike around August 2014, with nearly 400 comments posted in a single day. This was possibly driven by a simultaneous peak in the

number of posts created that day. A dramatic spike occurred in October 2014, the final stage of the forum. On the day members became aware that the forum was being monitored by law enforcement, over 3,000 comments were posted. This behavior can be attributed to mass communication about potential risks, content removal, or attempts to cover their tracks. The surge in activity during this crisis underscores a highly interconnected community capable of rapidly disseminating information about perceived threats.

Additionally, it becomes evident from Figure 1.2C that certain posts gained considerable popularity, as reflected by distinct days with significantly higher numbers of post likes than others. A positive correlation between the number of likes and downloads of a post in a CSA forum has been demonstrated in previous research [41], and such observations may suggest that the days with the highest number of likes on posts indicate the sharing of particularly engaging media content, such as photos or videos featuring CSA.

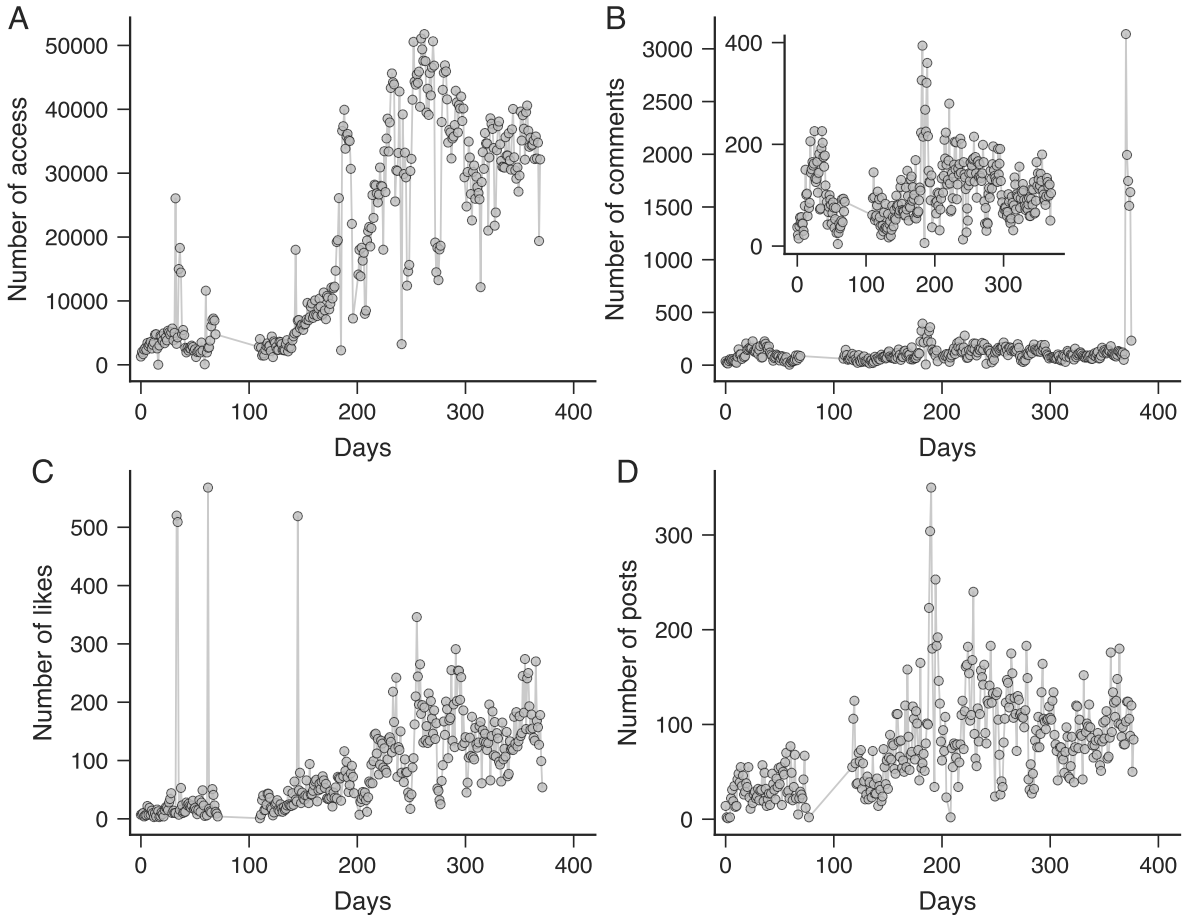

Figure 1.2: Temporal counts of forum activities between October 2013 and October 2014. Panel (A) illustrates the number of accesses; Panel (B) details the number of comments; Panel (C) shows the number of likes; and Panel (D) depicts the number of posts over time on the forum.

The manner in which members accessed and interacted with forum content exhibits a high degree of skewness, as illustrated by Figures 1.3A-1.3C. Panel (A) reveals a gradual decline in the number of views per post, with the most viewed post receiving nearly 10,000 views. The inset of Panel (A) displays the decay rate for the top 3,000 rankings, compared to a power-law decay with an exponent of 0.3. This suggests that while the majority of posts received a substantial number of accesses, only a small subset attracted attention from a limited number of members. A similar pattern is observed in Panel (C), which shows the distribution of likes given to posts by members. In contrast, Figure 1.3B demonstrates that only a select number of posts garnered an exceptionally high number of comments, while most received very few. For example, the most commented post accumulated 10,273 comments, far surpassing the second-most commented post, which received 1,406 comments. This disparity is particularly striking given that the most commented post revealed that the forum was being monitored by the police. A long tail is evident, with 64.2% $(8,141)$ of posts receiving only a single comment. Potential contributing factors to this phenomenon may include the nature of the content itself. For example, previous studies have shown that only a small number of forum users express interest in CSAM involving children aged 0-3 years [39], as well as in niche fetishes [45]. It is also well established in the literature that the majority of CSA forum members are not verbally active; although they access and download content, only a small percentage of users account for most of the forum communication.

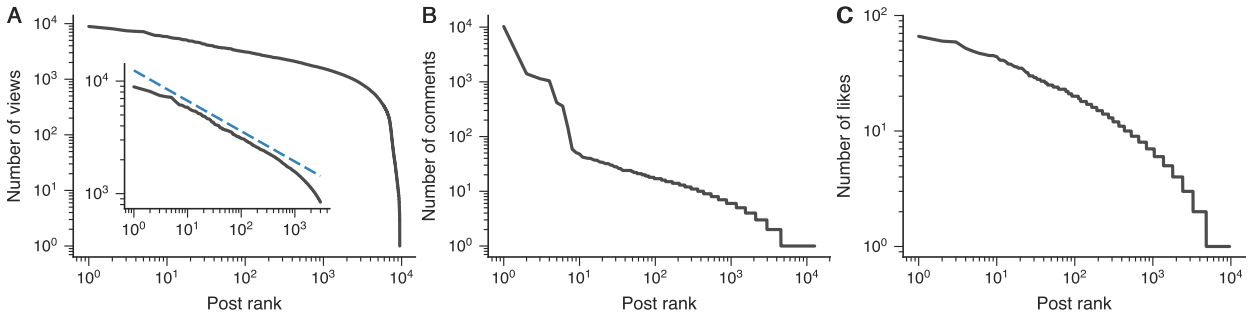

Figure 1.3: Ranking of forum posts based on their (A) number of views, (B) number of comments, and (C) number of likes, presented on a log-log scale. The inset in Panel (A) shows the decay rate for the top 3,000 rankings, compared to a power-law decay with an exponent of 0.3 (blue dashed line), indicating an approximate relationship of $\text{rank}^{-0.3}$.

1.3 Hierarchical structure of the forum

As previously mentioned, all forum content was analyzed by Brazilian Federal Police agents who classified posts into 11 distinct categories. The distribution of posts across these categories is shown in Figure 1.4. A significant portion of posts containing CSAM were labeled as "Uncategorized," leaving the nature of their content unknown. The category "General topics, rules, suggestions" has the second-highest post count, a pattern also observed in other CSA forums [44]. Prior research has shown that dark web CSA forums are often highly organized and efficient environments characterized by clear member hierarchies [54]. For instance, some forums explicitly prohibit "hurtcore" – CSAM involving pain and distress – and posting this type of content would result in a member being banned[36]. Thus, the high volume of posts in this category reflects the forum's emphasis on structure and shared guidelines. The "Potential offenders" category ranks third, with nearly 1,000 posts. Members contributing to this category were considered by the police as individuals who might be producers of CSAM. Moreover, the categories "Non-nude" and "Texts and drawings" may serve as entry points for users exploring such forums or represent less severe levels of engagement compared to explicit content. On the other hand, the last three categories appeal to a very small subset of users, reflecting their extreme preferences and niche fetishes.

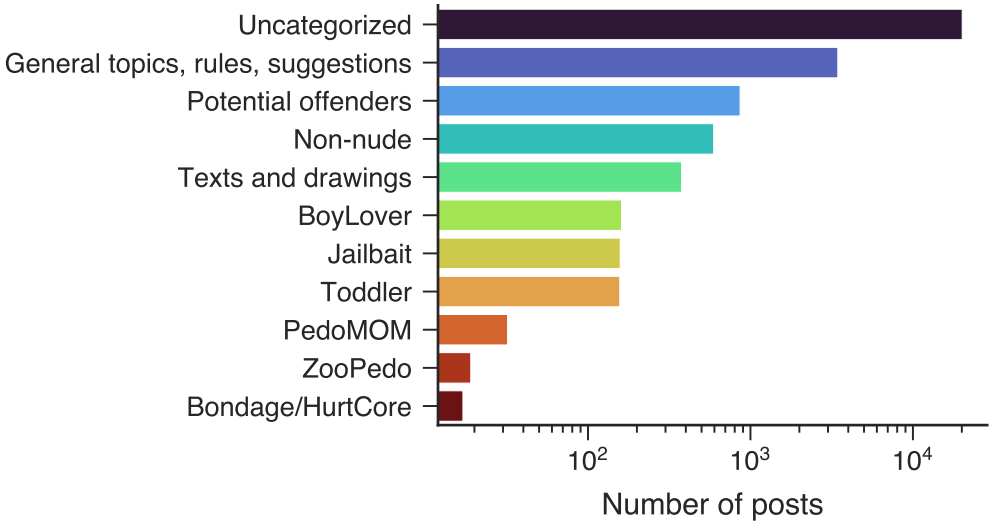

Figure 1.4: Distribution of posts across content categories.

In addition to examining the distribution of posts across different categories, we also assessed the ranking of posts within each forum category based on their number of views. Figure 1.5 shows that the majority of posts across most categories received a significantly

high number of views. Also, some interesting patterns emerge. For example, although the "Bondage/HurtCore" category contains fewer than 20 posts overall, the top-ranked post in this category had nearly 4,000 views. This indicates that while the category may be considered a niche within this particular forum, extreme content of this nature remains highly sought after, even if it is less frequent or less widely discussed. Furthermore, the top-ranked post in the "General topics, rules, suggestions" category also received a notably high number of accesses, highlighting that a large portion of members engage with content related to the forum's guidelines and community standards. As previously mentioned, this behavior is typical of individuals involved in dark web CSA forums. A large number of posts in the "Potential offenders" category also received considerable attention, raising concerns that users may be particularly interested in content linked to individuals who could actively be engaging in child abuse, whether by sharing their activities in textual form or through media, such as photos and videos.

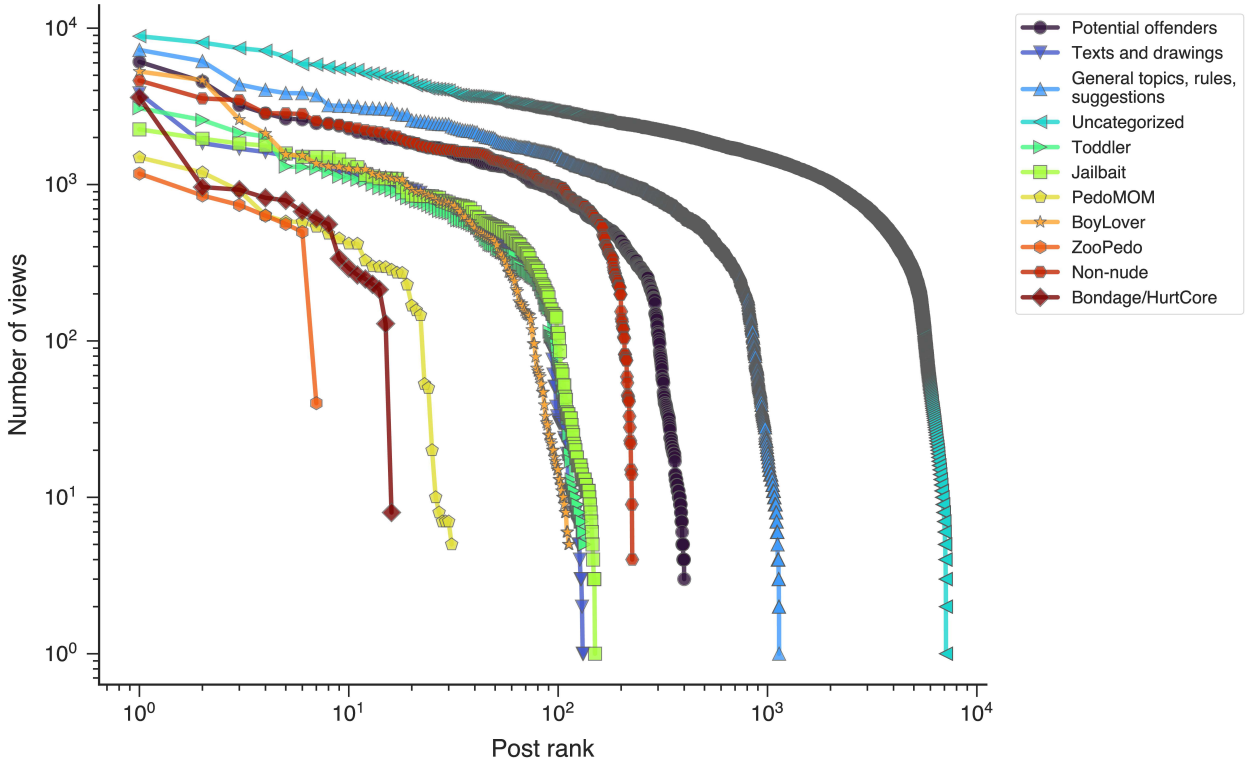

Figure 1.5: Ranking of posts in each of the forum's categories based on their number of views.

Further analysis of the forum's post rankings for each category, this time considering the number of likes received, revealed that the most-liked post belonged to the "Potential offend-

ers" category. This finding reinforces our earlier argument that members showed notable interest in content shared by individuals who may be real-life child abusers. Additionally, such posts with a high number of likes might include links to download CSAM, making them of particular interest to law enforcement. A notable discrepancy emerges between the results in Figures 1.5 and 1.6 across most categories, particularly for the more extreme content, such as "Bondage/HurtCore" and "ZooPedo." This could be attributed to users' reluctance to interact with or leave any traceable mark on these types of content, even within the anonymity of the dark web.

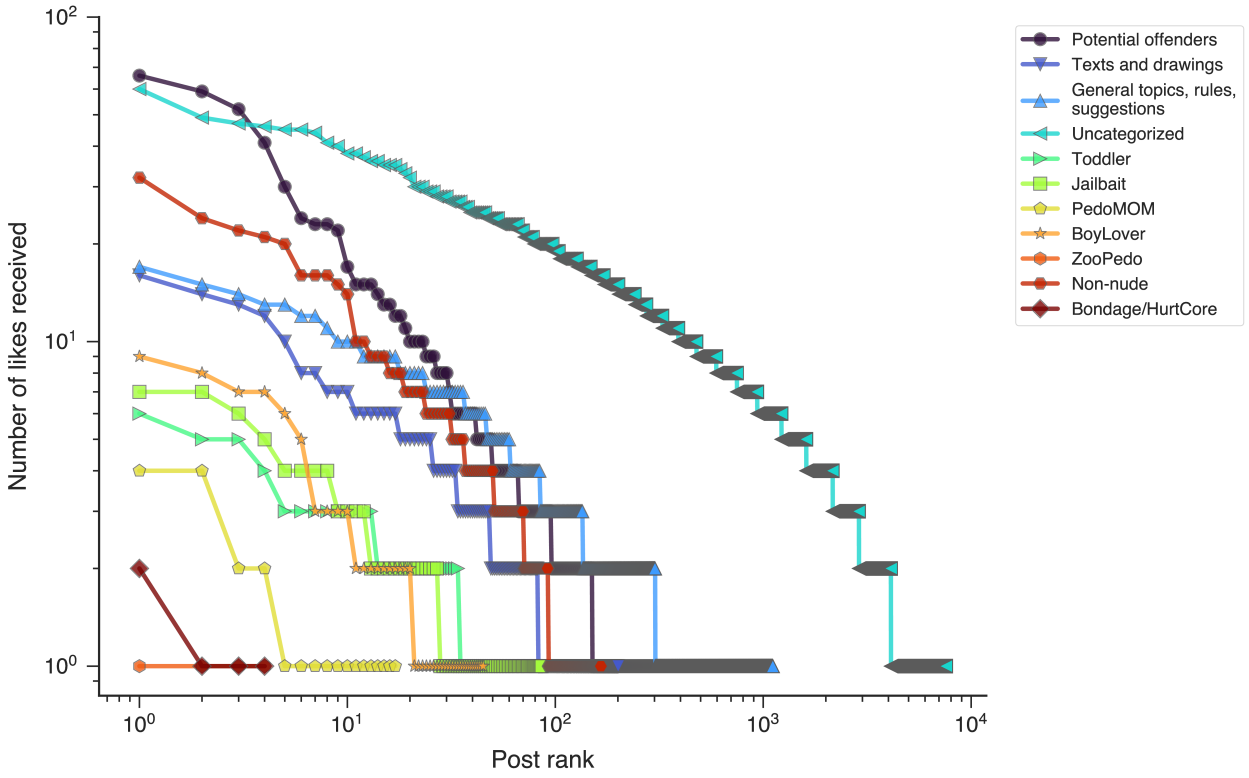

Figure 1.6: Ranking of posts in each of the forum's categories based on their number of likes received.

1.4 User profiles, activities, and dynamics

The ranking of forum members based on the number of views, posts, and comments is depicted in Figures 1.7A-1.7C. Panel (B) suggests that the forum heavily relied on a core group of highly active users, with the most active member responsible for creating over 1,000 posts. This observation is further supported by the fact that, although the forum had 9,614 members, only 31% (2,985 individuals) created a post during the period

when the forum was active. This core group may include moderators, administrators, or content creators who drive much of the interaction and activity. A similar trend is evident in Panel (C), which shows that a small number of top-ranked users posted an exceptionally high volume of comments relative to the rest. This dynamic highlights a distinction in the typology of online CSA offenders: specialized content creators, who contribute regularly, and consumers, who primarily observe or download content without engaging in discussions. This distinction is further illustrated in Panel (A), which demonstrates that the majority of forum users explored a substantial number of posts. Notably, the top-ranked member alone accessed posts over 200,000 times. These patterns illustrate the presence of social structures within the forum, in which a small number of users held central roles in setting the tone and establishing norms. Furthermore, the disparity in user post counts suggests that most users did not feel compelled or safe enough to contribute actively.

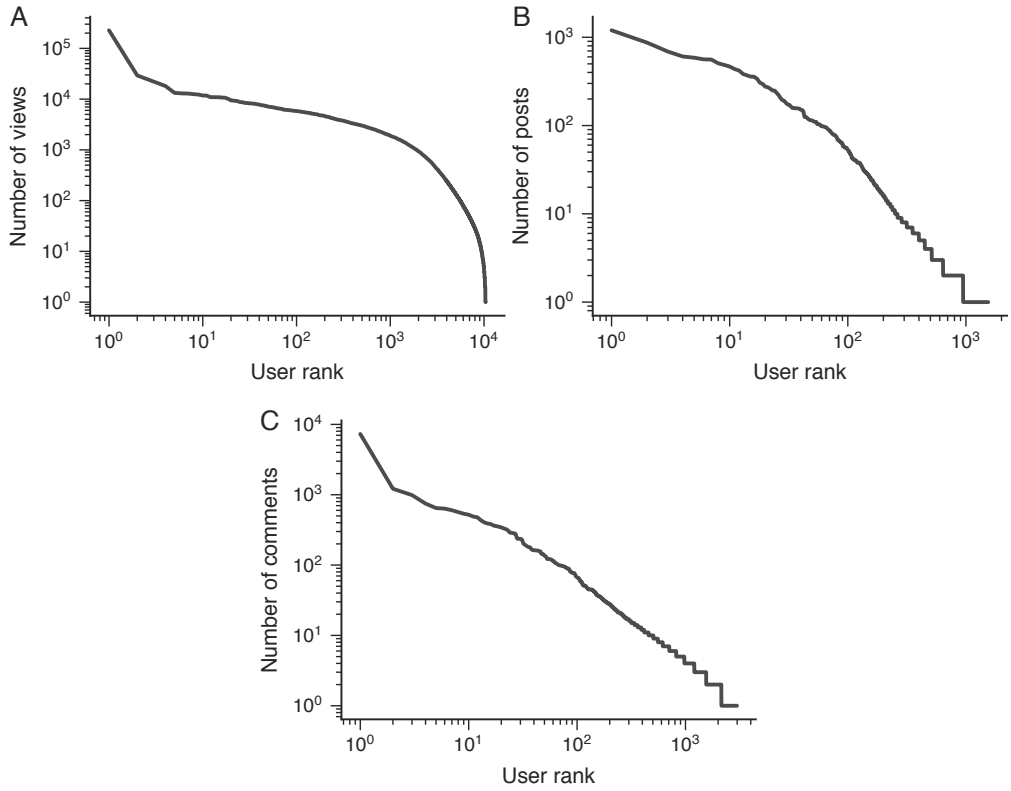

Figure 1.7: Ranking of forum members based on their (A) number of views, (B) number of posts, and (C) number of comments.

To further elucidate the dynamics of user activity and engagement within the forum, Figure 1.8 shows members' ranks based on the number of likes they sent and received on their posts. Users who frequently liked posts exhibited higher levels of community interaction,

actively appreciating and validating others' content. The top-ranked members in Figure 1.8A can be characterized as active consumers of content, reinforcing social bonds by endorsing posts they perceive as valuable. While they may be less visible than content creators, they play a critical role in the supply and demand dynamics of online CSAM. In contrast, the ranking of users based on the number of likes received on their posts, as shown in Figure 1.8B, reflects their popularity and influence within the forum. This suggests that the CSA content they produce either resonates deeply with the audience or is highly valued.

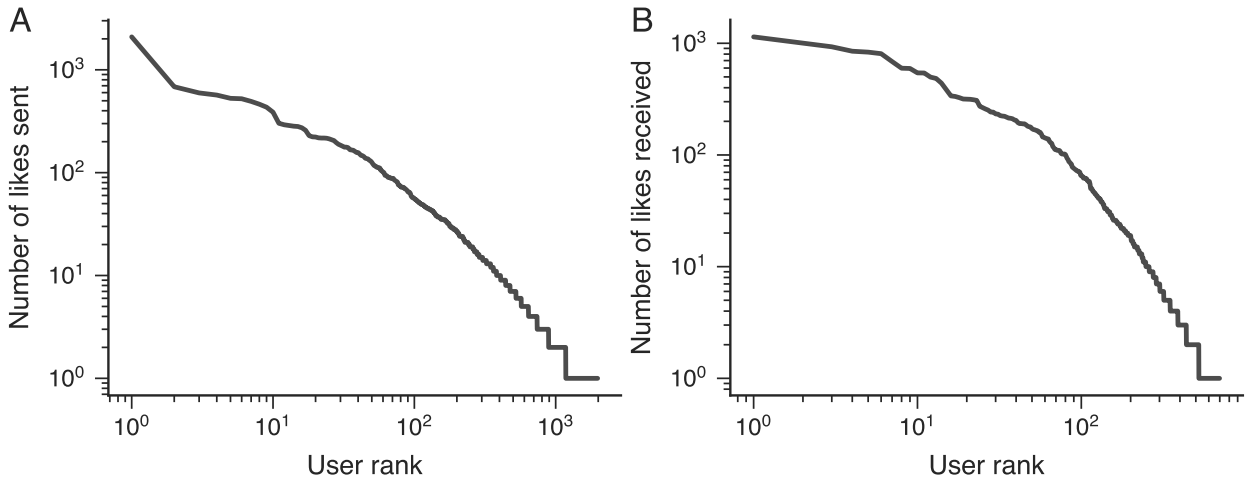

Figure 1.8: Ranking of forum members based on their (A) number of likes sent, and (B) number of likes received.

The data also revealed that members who frequently gave likes did not necessarily receive many in return, suggesting that producing content and receiving recognition were two distinct dynamics. For example, the top-ranked users in Panels (A) and (B) are different individuals. Moreover, members who both gave and received the most likes were possibly the most influential, as the overlap of users between the top ranks in Figure 1.8 indicates they were the forum's central figures, often taking on dual roles as content creators and consumers. It is also evident that the majority of members were passive participants, which aligns with our previous observations and is a "lurking" behavior commonly observed in most online communities [55].

When analyzing the relationship between the number of unique posts accessed and the total number of views per user, Figure 1.9A reveals a general trend of repeated engagement with the same posts. The dashed blue line in this figure represents a one-to-one linear relationship, indicating that each post was accessed once per member. Most users are relatively

distant from this line (note that values are on log-log scale), which suggests focused and compulsive CSAM consumption and implies that the majority of members are habitual consumers rather than one-time forum visitors. In contrast, the clustering of points along the 1:1 line might represent members who accessed the forum primarily to download content. Such behavior could be linked to the requirement of supplying some form of CSAM to gain access or maintain membership in other closed forums [56]. Previous research on dark web CSA forums has demonstrated that users often participate in multiple platforms simultaneously [47]. As a result, individuals who do not produce CSAM themselves may collect content from other forums to use as a form of currency. Panel (A) also reveals one clear outlier – a highly active user who accessed 8,030 unique posts over 200,000 times. Users exhibiting such extreme activity should be prioritized by law enforcement agencies for further investigation to understand their roles within the forum and potentially identify offenders. Additionally, examining which posts were repeatedly viewed by multiple users could highlight high-impact or central content within the network.

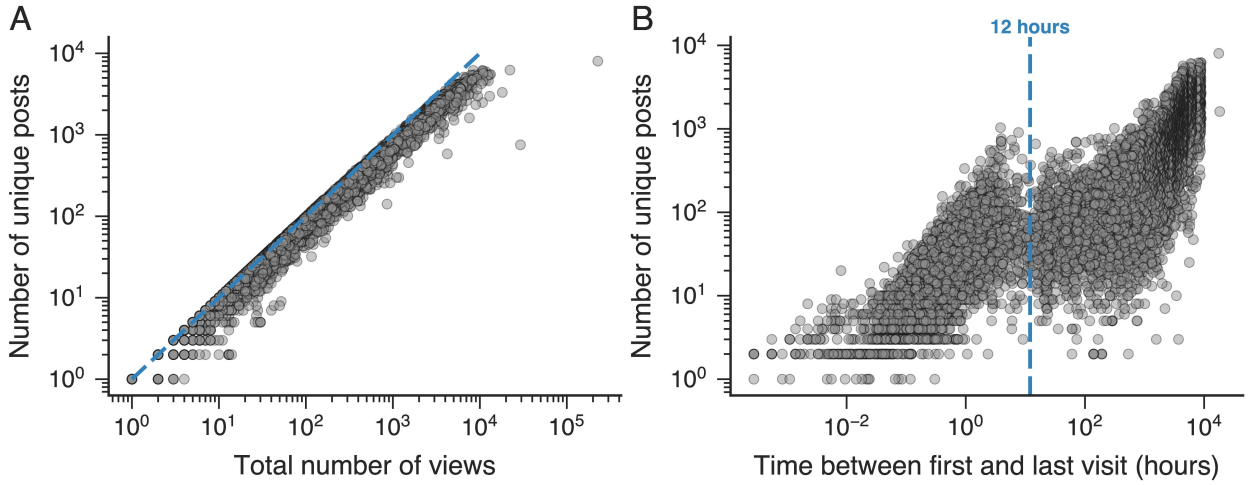

Figure 1.9: Number of unique posts as a function of (A) the total number of views per user, and (B) the time (in hours) between their first and last visit. The dashed blue line in Panel (A) represents a one-to-one linear relationship, while in Panel (B) it indicates the 12-hour mark.

To gain deeper insights into user behavior, Figure 1.9B illustrates the number of unique posts as a function of the time elapsed between a user's first and last visit to the forum. The dashed blue line at the 12-hour mark appears to divide members into two distinct groups: those with short-term and those with long-term engagement. The first group exhibited high activity within a very limited time frame. For instance, one user accessed nearly 1,000 posts

in under 12 hours. This behavior might reflect specific objectives, such as downloading large volumes of content or consuming material in a concentrated session. These users might not revisit the forum, potentially representing new members exploring the platform or one-time visitors with focused intent. They may also prioritize security and anonymity, minimizing their exposure by limiting their time on the forum. Conversely, long-term users demonstrated sustained activity over extended periods, with some spending thousands of hours on the forum and viewing thousands of posts. This group is likely to display more habitual or community-oriented behaviors, such as participating in discussions, revisiting threads, or maintaining ongoing collections of CSAM. Overall, Figure 1.9B implies that the forum catered to both transient and long-term user types, underscoring its adaptability in serving diverse needs. Recognizing this behavioral segmentation can inform law enforcement strategies, such as identifying patterns of short-term, high-volume activity or targeting long-term users to disrupt network structures.

Figures 1.10A-1.10C provide a comprehensive summary of member behavior in terms of content interaction. The Venn diagrams in Panels (A) and (B) illustrate that the majority of forum users were exclusively passive CSAM consumers. An interesting observation in Panel (B) is that every member who ever posted content also wrote at least one comment. Additionally, Panel (C) highlights active engagement actions, such as posting, liking, and commenting. In this case, a distinct pattern emerges: commenting was the predominant form of interaction. Notably, the 914 members present in the intersection of all three sets in Panel (C) represent the forum's most active and involved users. Given their heightened level of participation, these individuals should receive particular attention from law enforcement, as they likely exhibit a stronger commitment to criminal activity.

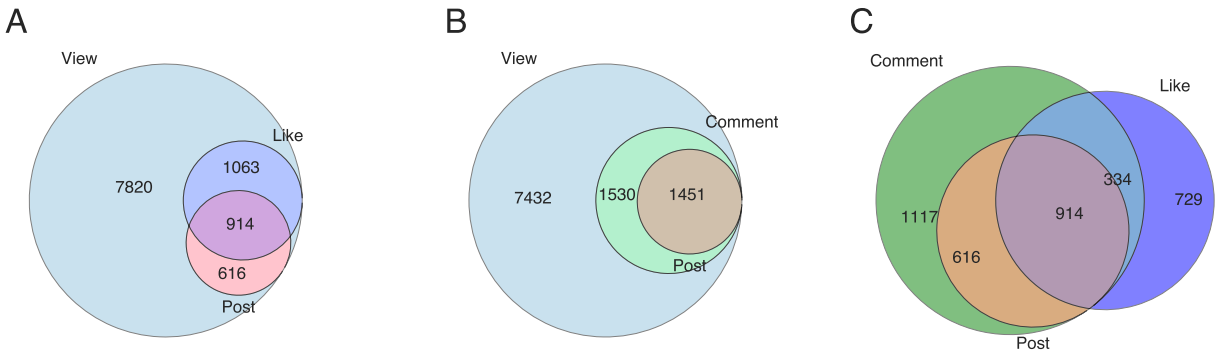

Figure 1.10: Venn diagrams of user engagement dynamics in the forum.

CHAPTER 2

Network structure and dynamic of a child sexual abuse forum

This chapter presents a descriptive analysis of the networked structure of individuals participating in the child sexual abuse forum. We begin by outlining the process used to construct the network, followed by an examination of its temporal evolution. To gain deeper insights into user dynamics, we then described their key network and statistical properties. Finally, we analyzed the subset of users identified by the police, assessing their positions within the network and evaluating whether they exhibit distinguishing characteristics compared to other forum members.

2.1 User-post bipartite network

Formally, a graph or network G is defined as a triple consisting of a vertex or node set $V(G)$, an edge set $E(G)$, and a relation that associates with each edge two vertices. A bipartite graph is a network structure in which the nodes can be divided into two disjoint sets, with edges occurring only between nodes of different sets [5, 57]. Such a graph is also referred to as a two-mode network. In the context of the CSA forum, a bipartite network is especially useful because it naturally captures the inherent structure of the online platform by representing users and posts as separate node sets.

In this framework, one set of nodes represents forum users while the other corresponds to

forum posts. An edge is formed between a user and a post if the user has viewed it at least once, with the edge weight reflecting the frequency of access. Since the network is bipartite, its structural evolution is largely driven by user browsing patterns and by the growth of both users and content, ultimately leading to the emergence of well-defined communities. To identify these communities or network modules, we adapted a nonparametric topic modeling approach based on the hierarchical clustering of two distinct node sets [58]. This method builds upon existing works on hierarchical stochastic block models (hSBMs) [59–61], first extending their application to infer topical structures within networks of texts and words and here adapted to our bipartite networks of forum members and posts. The model identifies the best modular partition of the network, generating a bipartite structure divided into groups across multiple levels and organized as a hierarchical tree. This approach enables the identification of structural patterns at various scales of resolution, providing insights into the depth of the hierarchy and the number of groups in both node sets.

We investigated the evolution of this network. Recognizing that daily analysis would have been computationally costly and time-consuming – for instance, considering the entire forum activity led to a network of 4,472,809 edges connecting 9,614 registered members and 25,943 posts – we constructed the networks using snapshots based on selected percentiles of total forum accesses. For example, the 21st day of forum activity corresponded to the first percentile, the 32nd day to the second percentile, and so on until the final day of forum operation. We aggregated all forum activity up to each percentile day to construct 100 snapshots of the network configuration. This approach was further motivated by the highly uneven daily distribution of forum accesses (see Figures 1.1 and 1.2), with some days exhibiting an exceptionally large number of accesses compared to others. By employing percentile-based networks, we accounted for this variability while ensuring a more efficient and representative analysis of the forum's activity over time.

Figure 2.1 illustrates the user-post network on the 21st day of monitoring, which corresponded to 1% of the total number of forum accesses. For clarity, all networks presented in this section display only 10% of the total number of edges for each day, with node sizes proportional to their respective degrees. There are four levels ($l = 0, 1, 2, 3$) in the hierarchical modular structure of the network for both user and post nodes, with user nodes on the left and post nodes on the right. For posts, the second-highest hierarchical level ($l = 2$) is represented by solid lines, dividing the nodes into only two distinct groups (red circles).

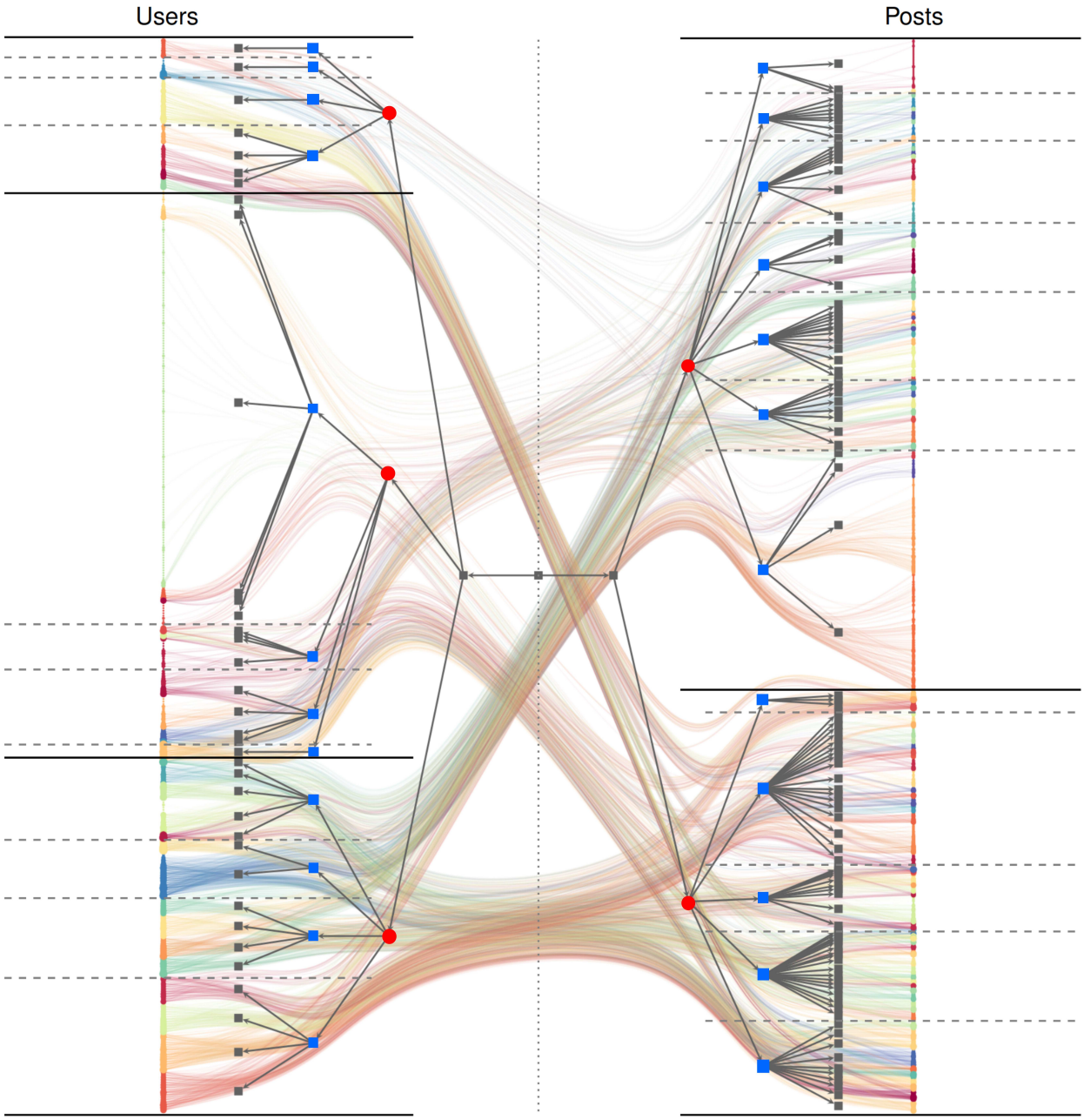

Figure 2.1: Modular structure of the user-post forum network on the 21st day of monitoring, corresponding to 1% of the total number of forum accesses. For posts, the hierarchical clustering reveals two groups at the second-highest level (red circles) and twelve groups at the third-highest level (blue squares). Similarly, for users, the second-highest level consists of three groups (red circles), which are further divided into twelve topics at the third-highest level (blue squares). For clarity, only 10% of all edges are shown in this visualization.

These groups are further subdivided at the third-highest hierarchical level ($l = 1$), as indicated by dashed lines. At this level, posts are organized into twelve distinct topics (blue squares). For users, the second-highest hierarchical level forms three groups (red circles),

which are subsequently split into twelve subgroups at the third-highest level (blue squares). At the lowest hierarchical level ($l = 0$), both sets of nodes are divided into a large number of groups, reflecting the network's complex structure.

2.2 Network growth and evolution

The evolution and growth of the user-post network are illustrated in Figure 2.2. Panels (A)-(C) depict the network on the 226th, 270th, and 323rd days of monitoring, corresponding to percentiles 25%, 50%, and 75% of the total forum accesses, respectively. Notably, half of the total forum accesses occurred within just 97 days. The modular structure of the user-post network reveals that as users' content preferences became more distinct, groups of posts started merging and forming more clearly defined clusters. Panels (A) and (B) exhibit eight hierarchical levels, while Panel (C) shows nine levels, indicating an increase in structural complexity over this period.

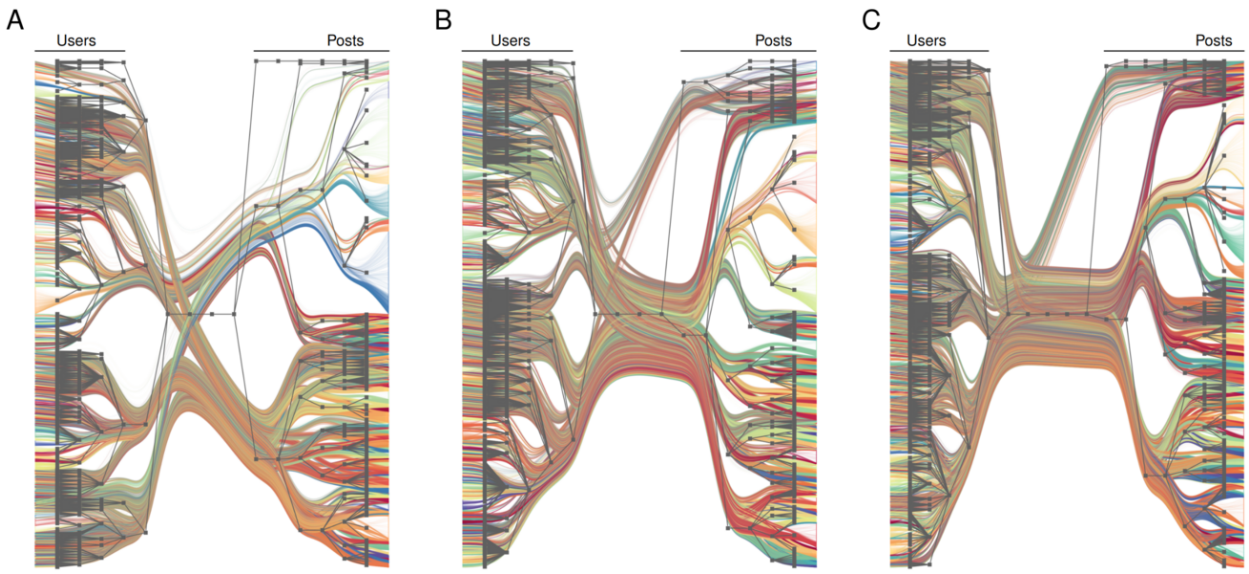

Figure 2.2: Structural representation of the user-post network after (A) 226 days, (B) 270 days, and (C) 323 days of monitoring, corresponding to 25%, 50%, and 75% of the total forum accesses, respectively. For clarity, only 10% of all edges are shown in each network visualization.

The user-post network after 375 days of police monitoring is presented in Figure 2.3, encompassing 100% of the total forum accesses. The hierarchical structure of post nodes consists of seven levels, and the pie charts illustrate the distribution of posts across content categories for the five groups in the fourth level, providing an overview of how posts were

clustered using our nonparametric modular modeling. To enhance category distinctions within each group, all posts labeled "Uncategorized" were removed from the charts.

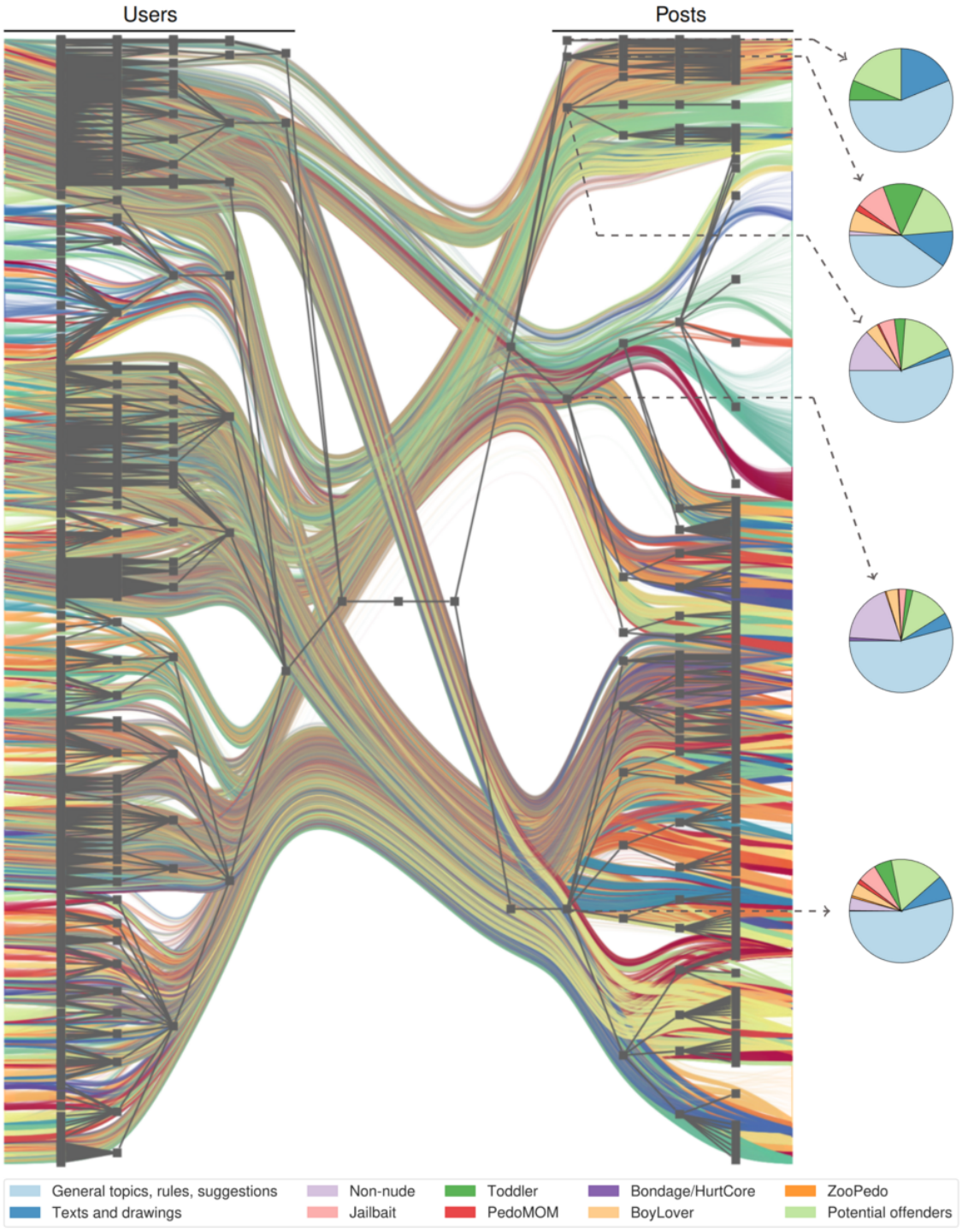

Figure 2.3: Structural representation of the user-post network after 375 days of monitoring, corresponding to 100% of the total forum accesses. The pie charts illustrate the distribution of posts across content categories for the five groups in the fourth hierarchical level, with colors corresponding to the legend.

Across all five groups, the majority of posts fall under the "General topics, rules, suggestions" category, emphasizing users' interest in following the forum's guidelines. The first group contains only 31 posts, primarily categorized as "Texts and drawings," "Potential offenders," and "Toddler." While similar categories appear in the second and third groups, distinct differences emerge in the distribution of posts from "Texts and drawings," "Toddler," "Jailbait," "BoyLover," and "Non-nude." The fourth and fifth groups are the largest, with 6,980 and 7,009 posts, respectively. Notably, posts from the "Bondage/HurtCore" category appear exclusively in these two groups. Their main distinction lies in the proportion of posts categorized as "Potential offenders" and "Non-nude."

To gain deeper insights into how content organization and user interactions develop over time, we analyzed the evolution of hierarchical levels within the user-post networks. Figure 2.4 shows that the number of post and user groups at the lowest hierarchical level ($l = 0$) increases significantly over time, reflecting the forum's expansion. Panel (A) displays a surge in user participation and a rapid creation of new content, which gradually fragmented into diverse topics and categories. Consequently, members' content preferences became more specialized. Initially, the forum exhibited only 4–5 hierarchical levels, but in later stages, this structure had as many as nine levels. This shift indicates that new content was not merely accumulating but also being systematically organized into thematic structures. Overall, the forum presented a dynamic evolution rather than remaining static, characterized by increasingly refined content consumption patterns.

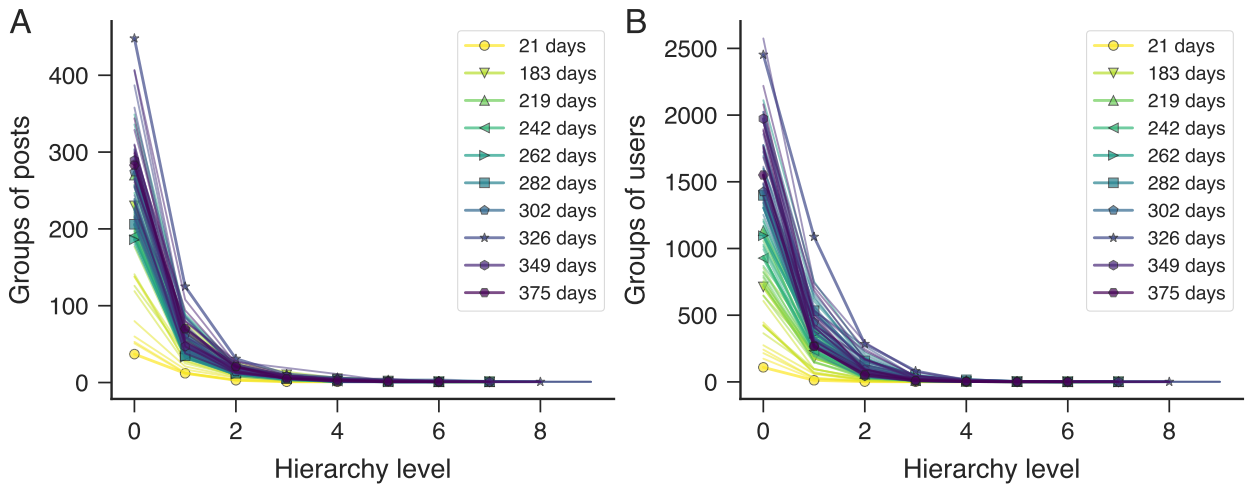

Figure 2.4: Evolution of the hierarchical structure within the user-post network during the monitoring period. Panels (A) and (B) indicate the number of post and user groups, respectively, at each hierarchical level over time.

We observed a similar pattern in Figure 2.4B, where the number of user groups escalates with the forum, indicating a segmentation of user behavior likely due to diverse content preferences. The increase in the number of groups at lower hierarchical levels also suggests the emergence of sub-communities or dedicated user clusters. This trend reflects different forms of engagement within the forum – such as lurkers, active contributors, and moderators – each occupying distinct hierarchical positions. The presence of more hierarchical levels points to a social stratification, with some users consolidating into broader groups while others remained more fragmented. Additionally, there was a co-evolution between users and content, with greater content diversity attracting a more varied user base.

Further assessment of the network's finer hierarchical levels revealed an almost exponential growth pattern in the number of post and user groups during the forum's active period, particularly at level 0, as shown in Figure 2.5. As previously discussed, this surge is driven by rapid content diversification, fragmenting the network into more distinct clusters. This aligns with findings in other online discussion forums or social media dynamics, where a few broad topics initially dominate, but as more content is produced, discussions become more specific [62, 63]. Moreover, saturation appears to occur at higher levels ($l = 1$ and $l = 2$), suggesting that broader categories might have formed and stabilized, although level 0 continues to grow exponentially.

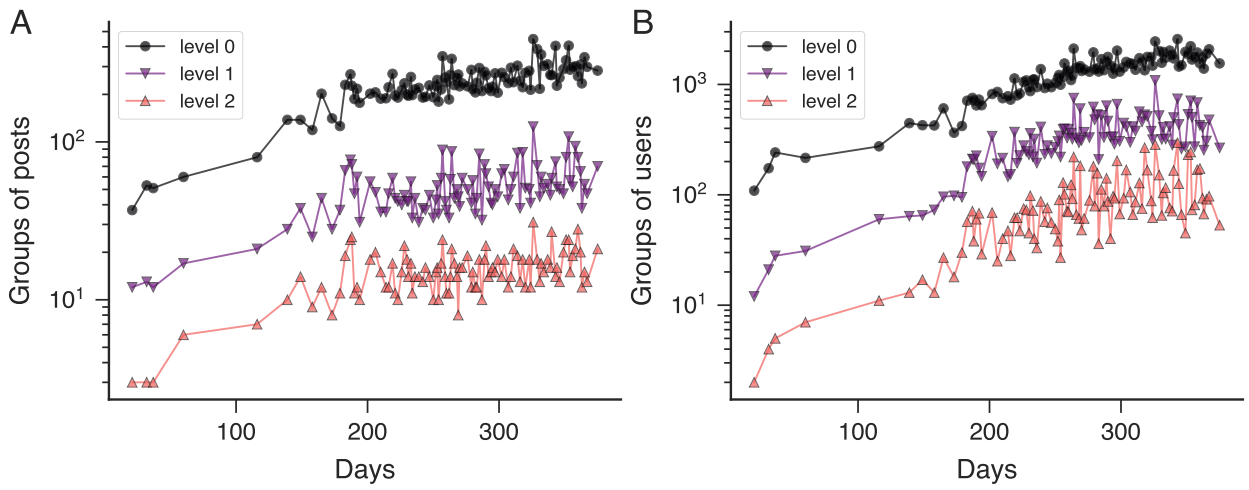

Figure 2.5: Hierarchical structure dynamics of the forum's user-post network during the monitoring period. Panels (A) and (B) indicate the number of post and user groups, respectively, at the three lowest hierarchical levels over time. Note that the vertical axes are on log scale, such that the approximate linear behavior indicates an exponential growth in the number of post and user groups.

Subsequent analysis of the user-post networks' evolving hierarchical depth is shown in Figure 2.6, revealing that the network remained relatively shallow during the early stages of the forum, with a maximum of six hierarchical levels in the first 100 days. This period likely reflects a phase of generalized content and broad topics, with only a few initial subgroups beginning to emerge. It may also suggest that members were still in the process of defining the major categories of interest within the forum – for example, some dark web CSA forums are typically centered on content involving either girls or boys. As new topics emerged, the hierarchy deepened, and the growing number of levels indicates that discussions became increasingly specific, with users actively organizing content into smaller, more defined subgroups.

Around 200 days into the forum's active phase, the hierarchy appeared to stabilize at eight levels, implying that the network structure had mostly solidified. Occasional days with up to ten levels may reflect significant events within the forum, such as the introduction of extremely niche topics or the addition of new content that leads to the creation of multiple subgroups. From a law enforcement perspective, the increasing fragmentation of the forum into smaller subgroups could raise concerns. As the network becomes more segmented, illegal activities might disperse across various subgroups, making it harder to monitor the full scope of content within the forum.

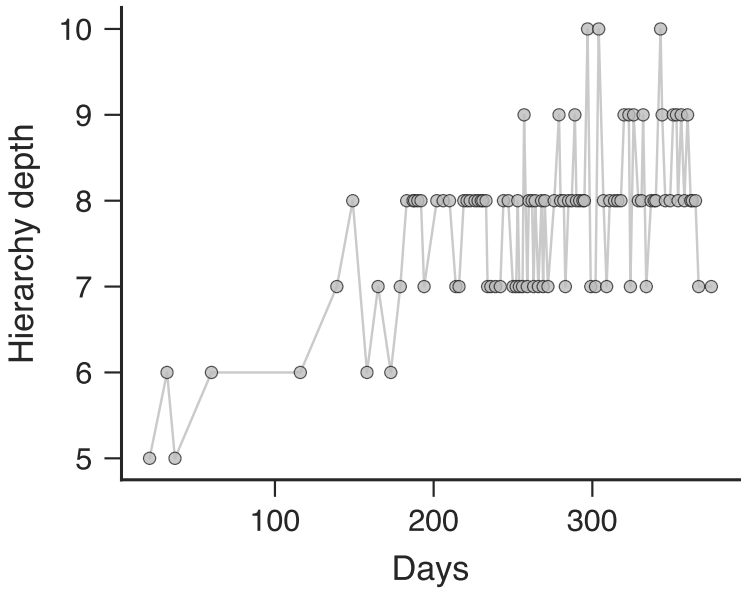

Figure 2.6: Hierarchy depth of the user-post networks over time.

2.3 Topological properties of the user-post network

The density of a network is a fundamental measure for assessing the overall connectivity of the system [5,64]. By letting $m = |E|$ denote the total number of edges, n_p the number of post nodes, and n_u the number of user nodes, the density of a bipartite network is defined as [64]

$$\delta(G) = \frac{m}{n_p n_u},$$

with its value lying strictly between 0 and 1 [64].

Given the highly specialized nature of the forum under analysis, we found that the network density is very low and continues to decline over time, as depicted in Figure 2.7. This decrease possibly occurred because members tended to repeatedly access the same content, resulting in many missing edges between users and posts. The sharp decline in density also reflects the network's expansion, as an influx of new members and the creation of additional posts dilute user engagement across a broader range of content. As the forum branches into increasingly specific subtopics, the probability of any single user engaging with a large portion of the available posts diminishes.

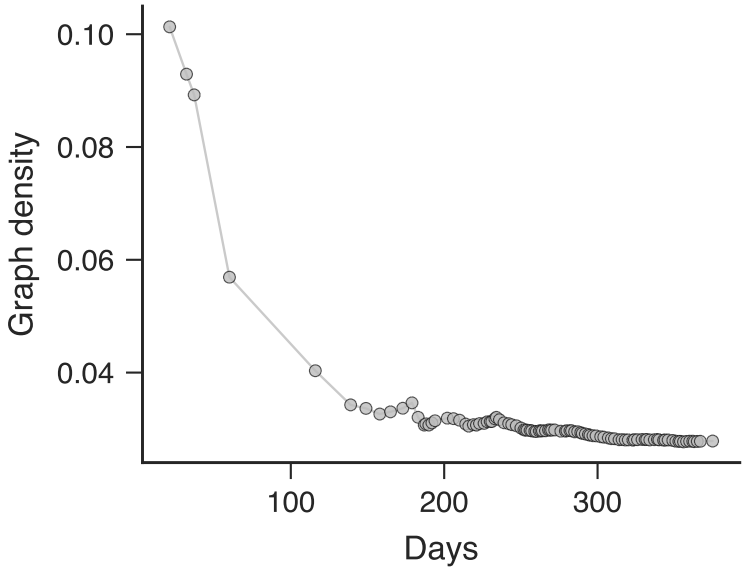

Figure 2.7: Density of the user-post networks over time.

Another key aspect of network characterization is its degree distribution, which, in this case, reveals the dynamics of content dissemination and user behavior within the forum. If p_k denotes the fraction of nodes in the network with degree k, then p_k is the probability

that a randomly selected node will have degree k [65]. The cumulative distribution function (CDF) of a random variable X is defined as:

$$F_X(x) = P(X \leq x), \quad \forall\, x \in \mathbb{R},$$

representing the probability that X takes a value less than or equal to x [66]. Given that the forum's user-post networks are bipartite, we estimated the CDFs of degree for post and user nodes separately, as shown in Figures 2.8A-2.8B, respectively.

We observed in Panel (A) that most posts had low degrees, with less than 0.01% experiencing high levels of interaction in the later stages of the forum. Conversely, some posts even reached degrees exceeding 3,000, marking extreme outliers in terms of engagement. The small values of the CDF at this degree suggest that highly popular posts were exceptionally rare. The degree distributions indicate that the majority of content remained overlooked while a few posts captured significant user attention. A similar pattern is evident in Panel (B), which shows the user degree distribution, revealing a very uneven engagement pattern, with a mix of casual and highly active members. This finding aligns with the existing literature on dark web CSA forums and other social networks, where the presence of a few high-degree users is a commonly observed phenomenon.

Additionally, Figures 2.8C-2.8D present normalized versions of Panels (A) and (B), where post and user degrees were rescaled by their means to fit a common reference frame. This normalization allows for a comparison of degree distributions across different snapshots of the network's evolution while controlling for fluctuations in total activity, enabling us to test whether the distributions remain stable over time. If the rescaled distributions converge onto a single curve, it would suggest that the underlying distribution is consistent over time. Comparing Panels (C) and (D), we observe that the curves for different time periods mostly overlap in Panel (D), suggesting that the distribution of user degrees remained relatively stable despite the forum's growth. This implies that user behavior did not fundamentally change over time regarding the disparity in content visualization. Since the overall structure of user participation remained consistent, it points to new members adopting similar engagement patterns as older members. Furthermore, the shapes of these distributions do not conform to standard forms such as exponential or lognormal distributions, reinforcing the presence of heterogeneous behavior in the user–post dynamic.

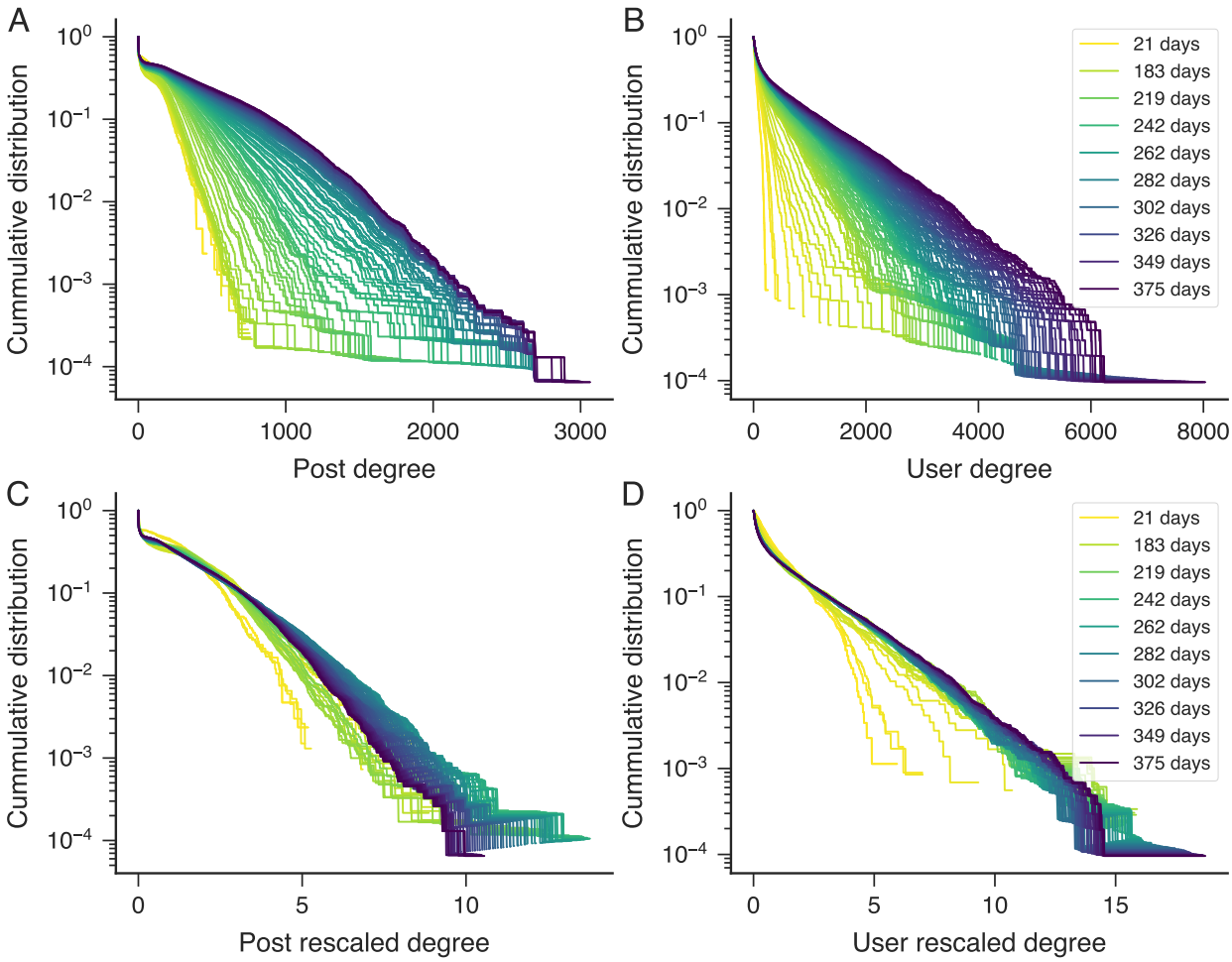

Figure 2.8: Cumulative distribution functions for the (A) post degree, (B) user degree, (C) rescaled (normalized) post degree, and (D) rescaled (normalized) user degree. The vertical axes in all panels are in log scale.

Extending the same analysis while weighting the degree by access frequency, we obtained the distributions shown in Figure 2.9. Comparing these distributions with their non-weighted versions of Figure 2.8 reveals a shift in the distribution shape. In the weighted version, the distributions resemble more a heavy-tailed distribution, indicating that when interaction frequency is considered, high-degree posts and users become even more dominant. This might indicate an increasing interaction pattern throughout the forum's active period – for example, the longer a post existed, the more accesses it received. Figure 2.9A also shows that forum members revisited posts multiple times rather than engaging with a diverse range of content. Moreover, it becomes evident that high-degree users accessed content very frequently, highlighting a core group of regularly active members. This observation can aid law enforcement agencies in identifying high-risk users and posts, as frequency-weighted

engagement patterns may distinguish habitual consumers or content producers from one-time visitors.

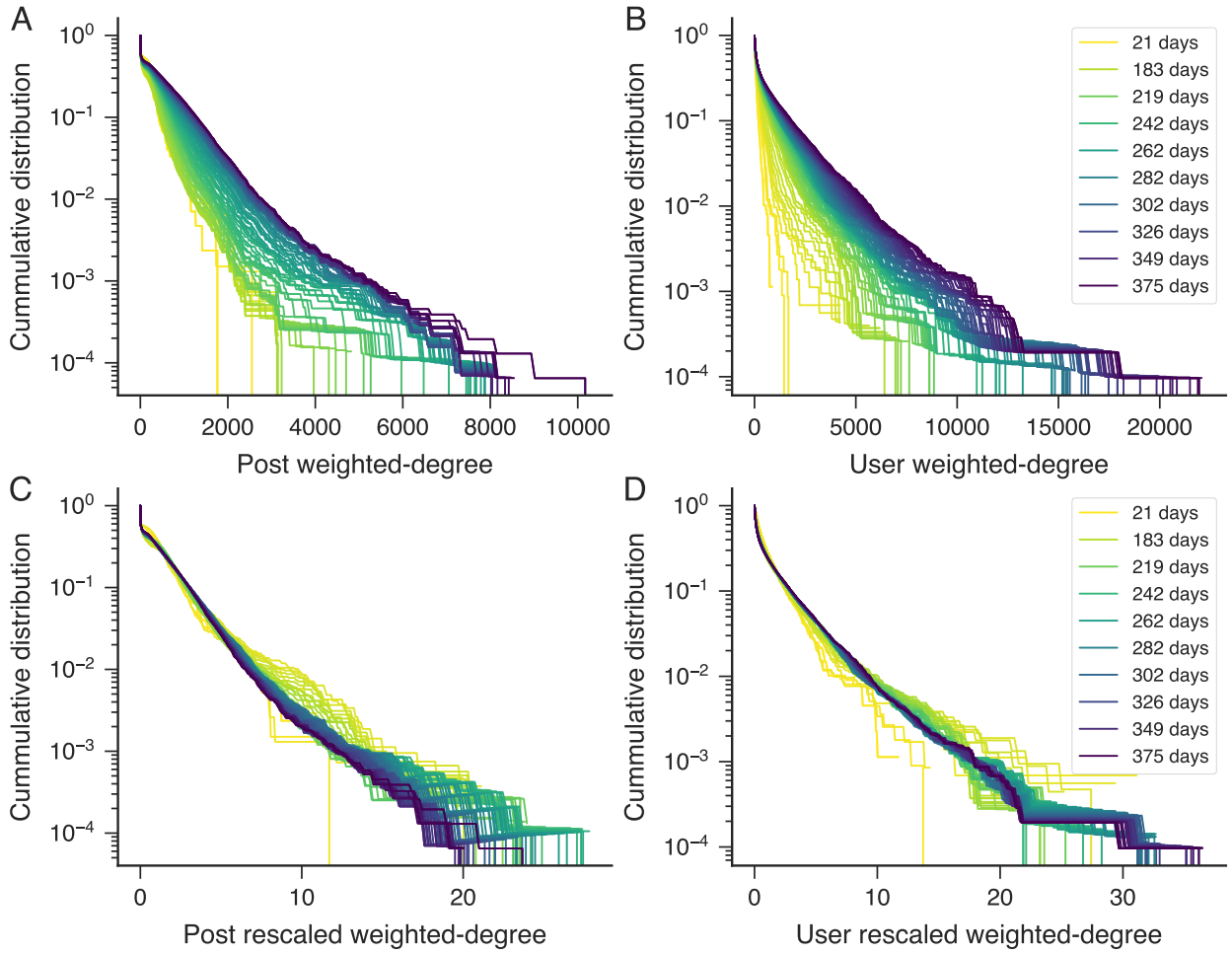

Figure 2.9: Cumulative distribution functions for the (A) weighted post degree, (B) weighted user degree, (C) rescaled (normalized) weighted post degree, and (D) rescaled (normalized) weighted user degree.

We now turn to the temporal evolution of node degrees in the networks, examining the average post and user degrees to provide a complementary perspective on how user engagement and content dissemination changed throughout the forum's lifespan. Figure 2.10A shows that the average post degree initially remained around 100, experiencing some fluctuations in the early days. However, a sharp increase occurred around day 200, suggesting an acceleration in content consumption. This aligns with the demographic results in Figures 1.1 and 1.2, which previously demonstrated a surge in new user registrations and overall forum activity during the same period. The general upward trend indicates that posts received more interactions over time, driven by an expanding user base that engaged with more con-

tent and exhibited a persistent behavior of revisiting posts. A similar trend is observed in Figure 2.10B, which illustrates the standard deviation of post degrees. The simultaneous increase in both the average post degree and its standard deviation suggests a polarization effect, pointing to a growing disparity between popular and unpopular posts in terms of views. This pattern implies that a select number of posts became disproportionately dominant, potentially achieving viral status and attracting repeated user engagement.

Regarding the average user degree and its standard deviation, the near-linear growth observed in Figures 2.10C-2.10D is concerning, as it suggests an intensification of user interaction over time. New members may have begun with limited activity but gradually turned highly active, developing habitual patterns of CSAM consumption. The forum's growth was not uniform; rather, it became increasingly fragmented and hierarchical, with key members emerging while others remained as lurkers.

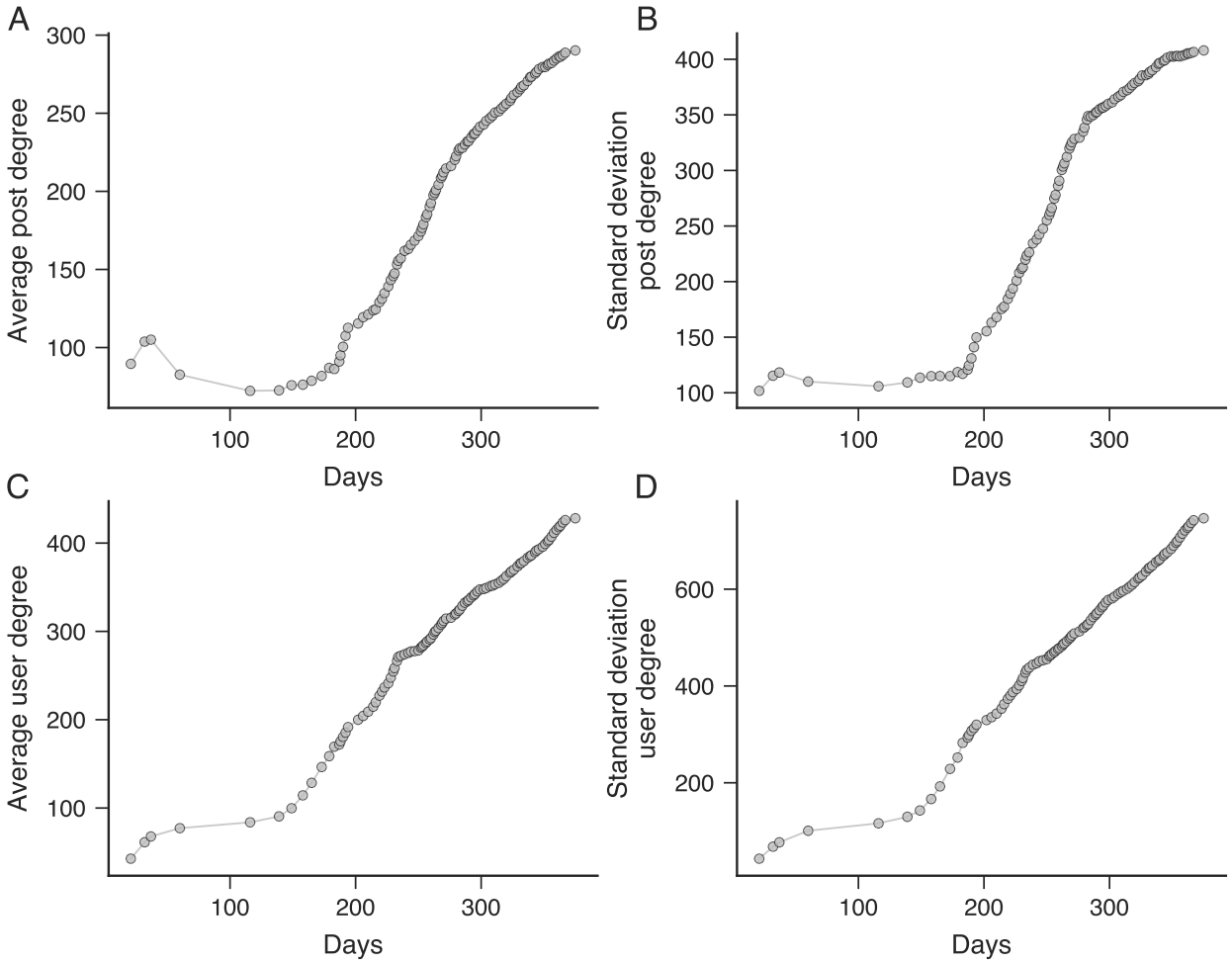

Figure 2.10: Temporal evolution of the (A) average post degree, (B) standard deviation of post degree, (C) average user degree, and (D) standard deviation of user degree.

We further expanded these findings by considering interaction frequency in the average degree of users and posts, as shown in Figure 2.11. Comparing Figures 2.10A and 2.11A, we noticed that the average weighted post degree is considerably higher, indicating that certain posts served as recurring engagement points, potentially containing highly sought-after CSA content that attracted repeated accesses from the same users. Likewise, the average weighted user degree in Figure 2.11C is substantially higher than in Figure 2.10C, with an exceptionally large standard deviation. This reinforces our findings that a small number of central figures played a dominant role in the network. Given their likely influence on the forum's supply and demand of content, these individuals should be prioritized as high-risk targets for law enforcement.

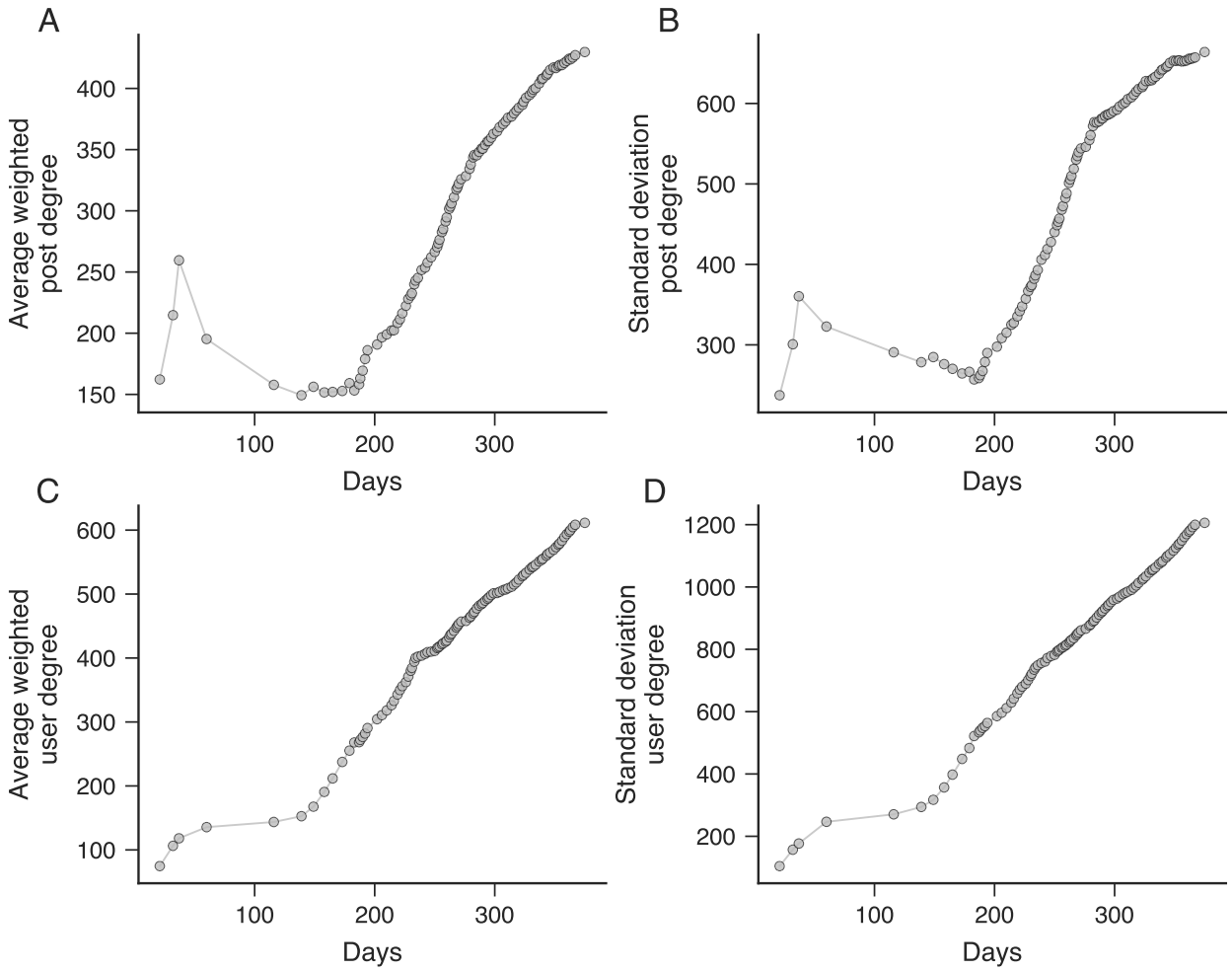

Figure 2.11: Temporal evolution of the (A) average weighted post degree, (B) standard deviation of weighted post degree, (C) average weighted user degree, and (D) standard deviation of weighted user degree.

2.4 Degree correlations in the user-post network

An additional characteristic of the user-post network investigated was how nodes with different degrees interacted, focusing on the correlation between their connectivity. This degree correlation pattern determines whether high-degree nodes tend to link with similarly connected nodes or with less connected ones. Since the user-post networks are bipartite, there are no direct connections between users – in this framework, every user's first neighbor is a post. Therefore, examining the degree of a user's neighbors provides insight on whether they engaged with content that was widely viewed by other members. In this context, Figure 2.12 presents the degree correlation between user weighted degree and the average post weighted degree across eight different days.

To quantify this linear relationship, we calculated the Pearson correlation coefficient ρ, a widely applied measure of similarity. Given two random variables X and Y, with means μ_X and μ_Y, respectively, the correlation coefficient is defined as:

$$\rho = \frac{\sum(X - \mu_X)(Y - \mu_Y)}{\sqrt{\sum(X - \mu_X)^2(Y - \mu_Y)^2}},$$

where $-1 \leq \rho \leq 1$. Positive values of ρ indicate that Y increases as X increases, while $\rho < 0$ implies that Y decreases as X increases. If $\rho = \pm 1$, the variables exhibit a perfect linear relationship, whereas $\rho = 0$ signifies no correlation [67, 68]. Thus, each panel in Figure 2.12 also includes the corresponding value of ρ, illustrating the strength and direction of the relationship between the two variables.

We observed an overall negative correlation, which implies that the most active users tended to interact with posts that, on average, received fewer views from other forum members. This suggests a pattern of specialized engagement, where high-degree members focused on niche content rather than widely popular posts. Over time, the correlation coefficient ρ decreased in magnitude, reflecting a weakening of this negative correlation. This trend suggests that user interactions became more evenly distributed across different posts.

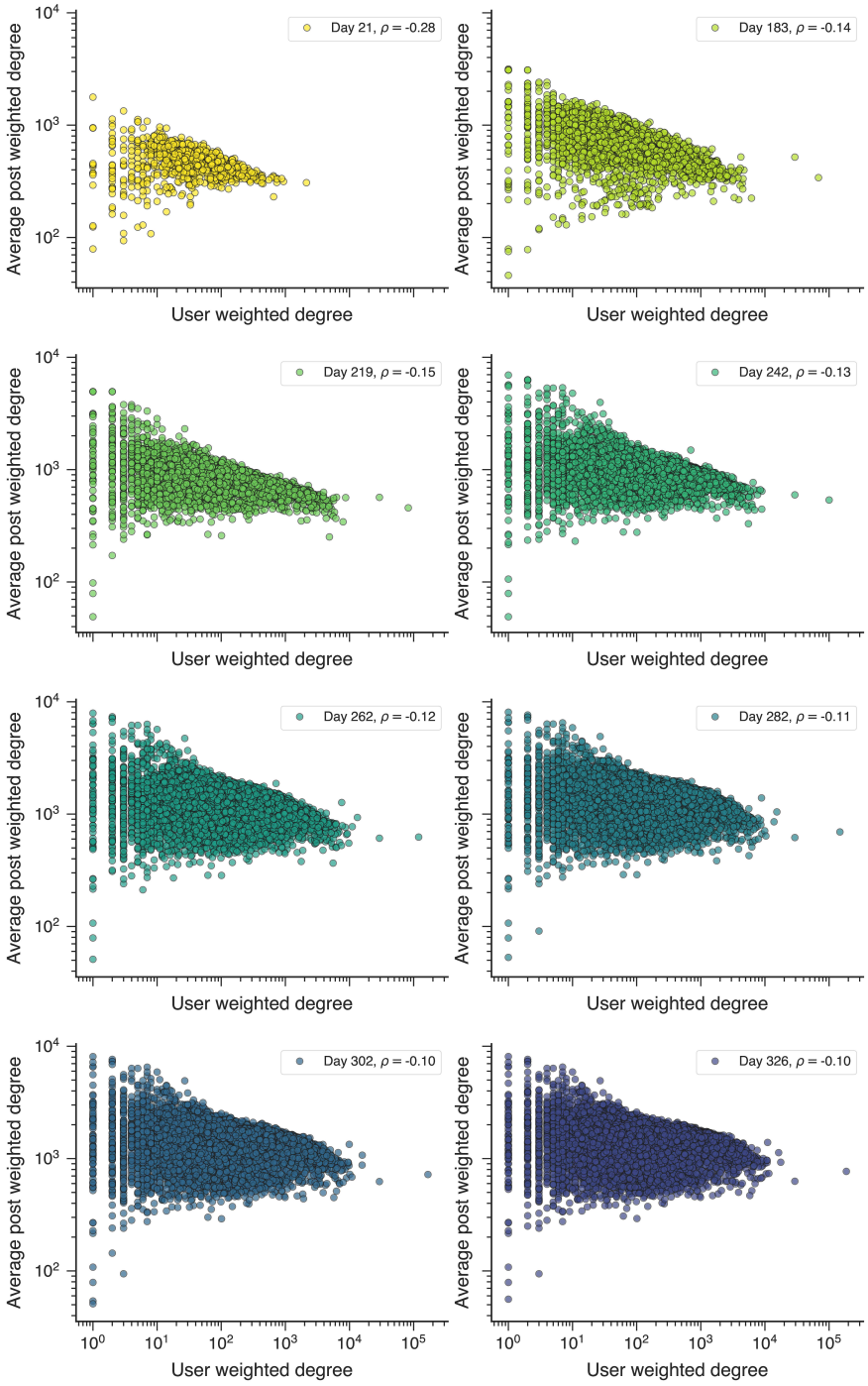

Figure 2.12: Correlations of weighted user and average post degrees across eight different days evenly spaced according to the percentiles of total forum accesses.

While the previous results provide insights into user engagement, Figure 2.13 focuses on post dynamics exhibiting the correlation between the weighted post and average user degrees, highlighting thus whether certain posts played a central role in shaping the forum's structure. In this case, high-degree posts were, on average, more frequently accessed by low-degree members. These users were likely newcomers to the forum, initially accessing its

most popular posts to familiarize themselves with the main content and discussions.

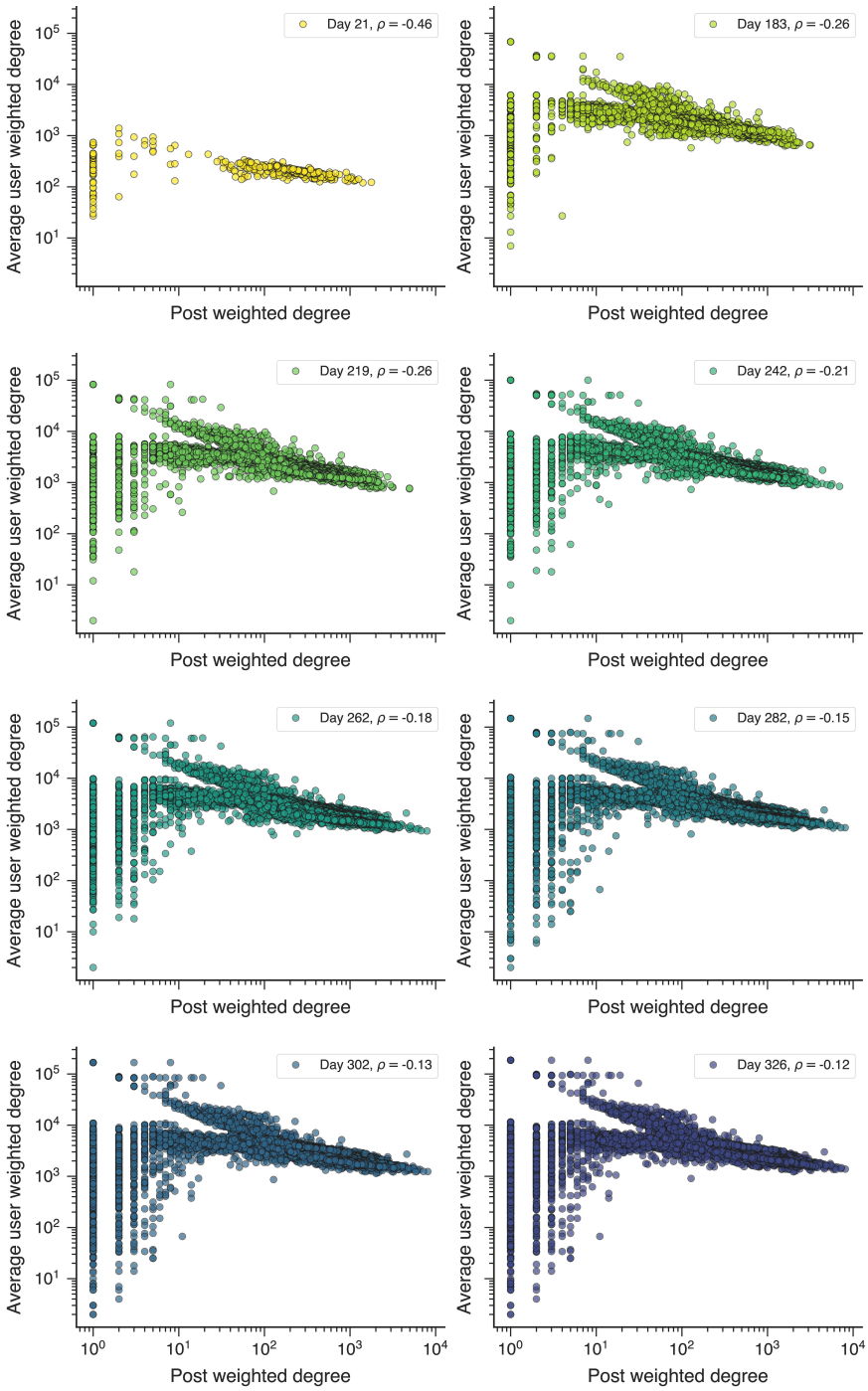

Figure 2.13: Correlations of weighted post and average user degrees across eight different days evenly spaced according to the percentiles of total forum accesses.

Despite both figures displaying negative correlations, the differing perspectives and values of the correlation coefficient reveal a key insight: post popularity had a stronger influence on user behavior than user behavior had on post popularity. In other words, the forum's overall interaction dynamics was more structured around content than users. Although some highly

active members were present, the stronger post-driven correlation suggests that the content itself was the primary force behind interactions. This finding implies that law enforcement efforts should focus not only on the most active users but also on the most structurally significant posts, as removing high-engagement content may be more effective in disrupting forum activity than targeting individual users.

To better understand how the average user degree correlated with the average post degree over the monitoring period, we analyzed this degree correlation in more detail. Figure 2.14(A) presents the correlation coefficient values for the user-post networks, and Figure 2.14(B) considers weighted degrees, accounting for repeated interactions. For comparison, the plot lines with cross markers in Panels (A) and (B) represent users with a degree greater than 10 and a weighted degree exceeding 50, respectively.

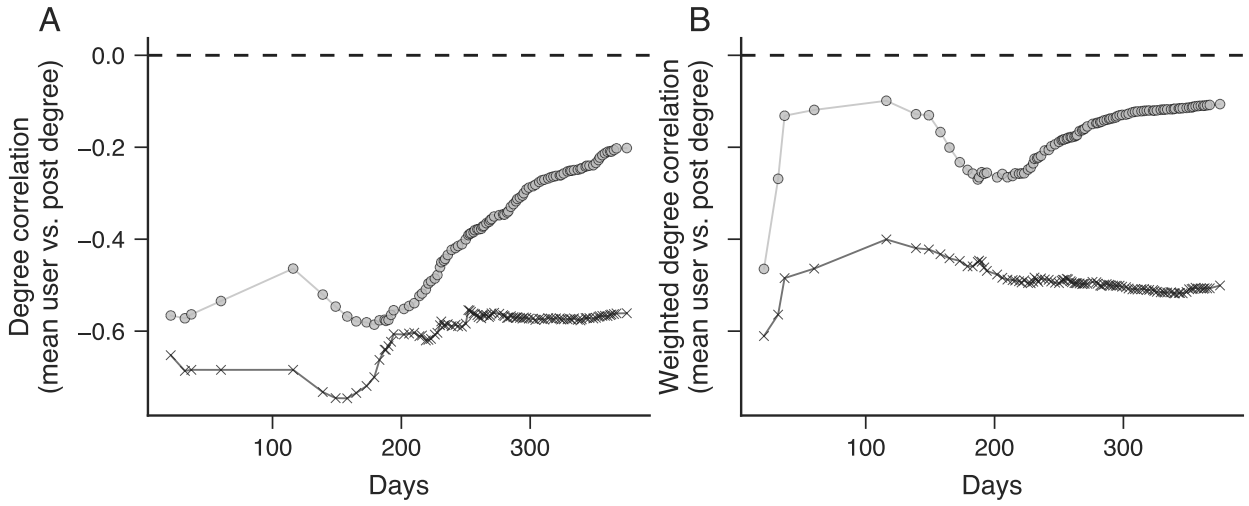

Figure 2.14: Temporal evolution of the (A) degree correlation and the (B) weighted degree correlation between the average user and post degrees.

As both correlations remained persistently negative, a fundamental misalignment exists between highly active users and highly accessed posts – they do not strongly overlap. This indicates that the forum exhibited strong disassortativity, where the most engaged users actively avoided popular posts. Such behavior could suggest deliberate evasion tactics, with key members attempting to stay under the radar. It also implies a fragmented discussion structure with hierarchical engagement, where different user groups interacted within separate spheres: high-degree users engaged in smaller networks, focusing on specific, less visible content, while casual users primarily interacted with popular posts, which may have functioned as entry points. Additionally, monitoring periods when the correlation shifts toward

zero could help identify moments of increased coordination, when users temporarily align their engagement patterns.

Even in terms of repeated engagement, Figure 2.14B reveals that highly active users did not concentrate on the most popular posts; however, it displays sharper fluctuations. This suggests that high-degree posts struggled to retain active users over multiple interactions. Instead, users not only engaged separately but did so repeatedly, forming isolated content hubs. The weaker negative values further indicate that some high-activity users occasionally interacted with popular content.

2.5 Profile and topological properties of law-enforcement-identified users

During Operation Darknet, law enforcement identified the IP addresses of 202 forum users by means of traps planted in the forum structure. Among these, about a hundred were actually identified and targeted or arrested by the Brazilian Federal Police. Analyzing the behavioral patterns of these individuals is particularly valuable for future investigations, as it provides critical insights into their level of engagement and potential influence within the ecosystem of online CSA content consumption. By assessing their characteristics, we aim to determine whether identified users exhibited distinct interaction patterns. It is important to highlight that 31.7% (64 individuals) of the identified users authored posts in the "Potential Offenders" category, while 57.9% (117 individuals) contributed to the "General topics, rules, suggestions" category. This distribution showcases that law enforcement successfully identified some of the most involved and potentially high-risk forum members, underscoring the effectiveness of the investigative strategy.

Thus, to better understand the structural role of these uses within the network, we created a modified version of the user-post network previously shown in Figure 2.3. This new version network, presented in Figure 2.15, displays the identified users and their respective edges highlighted in red, indicating that they were not uniformly distributed in the network.

The degree distribution of identified and unidentified users, representing the total number of posts they accessed, is presented in Figures 2.16A-B. Panel (A) shows that identified users had an overall higher degree than unidentified users, indicating greater content consumption.

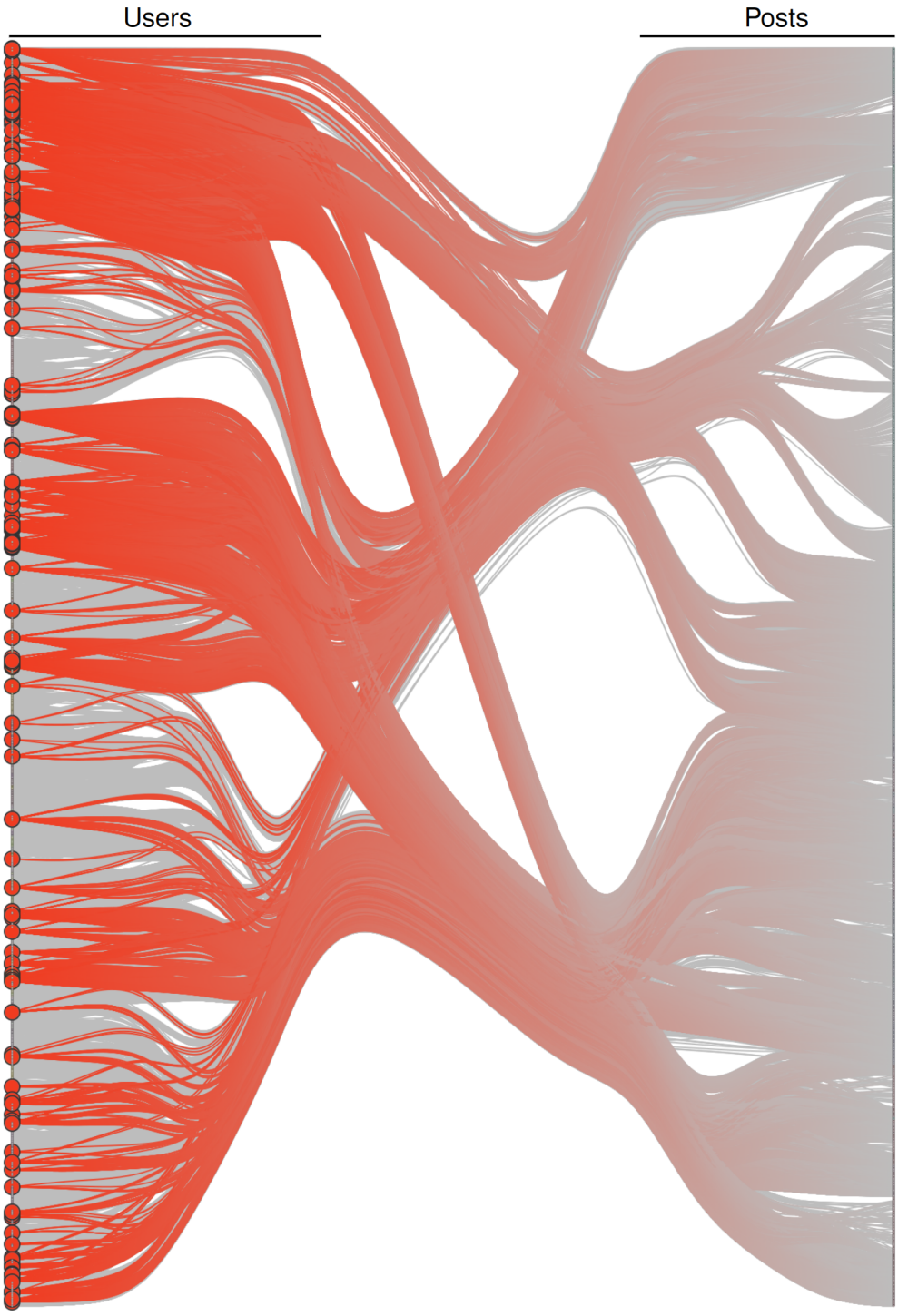

Figure 2.15: Visualization of law-enforcement-identified users in the user-post network after 375 days of monitoring. Red nodes and their corresponding edges represent the identified users, while gray nodes indicate unidentified users.

However, the presence of a few low-degree outliers suggests that even minimally engaged individuals could be detected. In addition, Panel (B) displays the weighted user degree, which accounts for access frequency and is essential for distinguishing habitual from casual users. In this case, identified users exhibited a much stronger tendency to revisit specific posts compared to unidentified users, suggesting a compulsive viewing of CSAM.

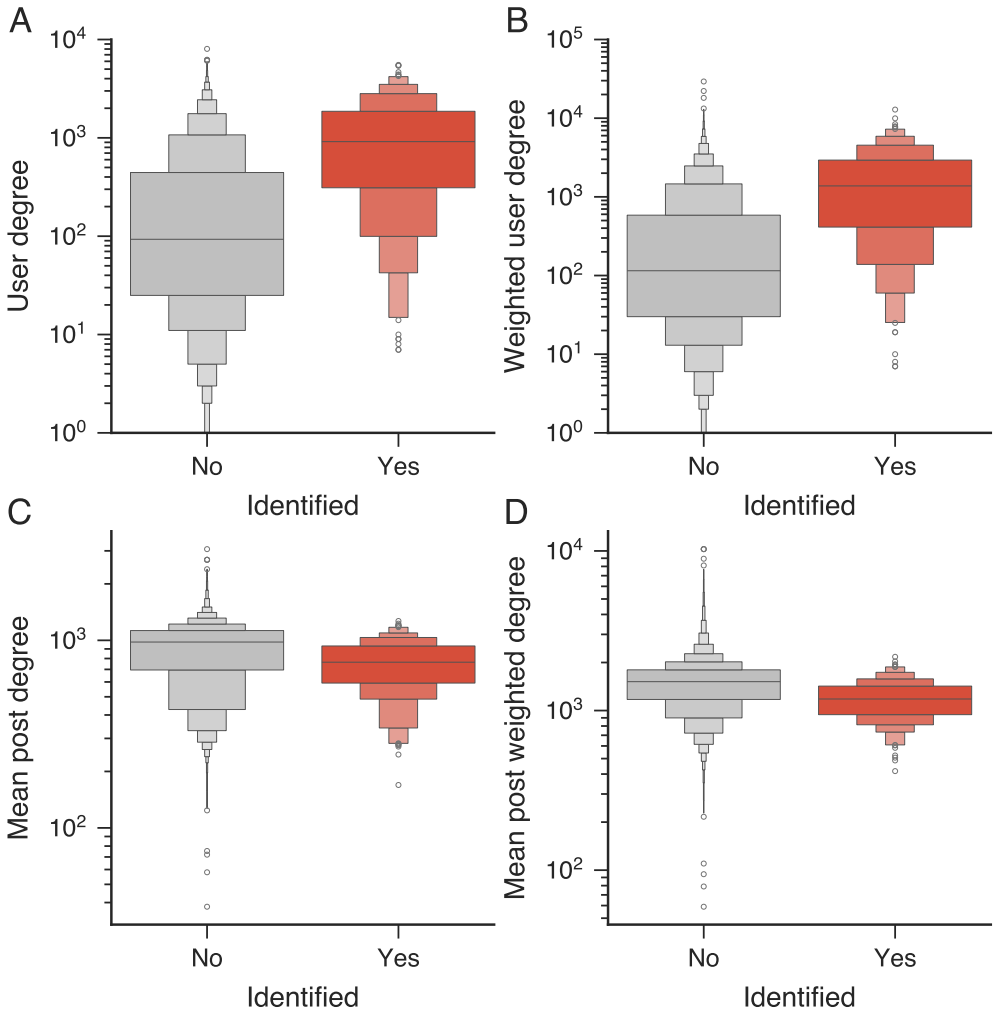

Figure 2.16: Distributions of (A) user degree, (B) weighted user degree, (C) mean post degree, and (D) mean post weighted degree in the user-post network of identified users.

To further examine whether identified users were more likely to access highly popular posts, we extended this analysis to the mean post degree, as shown in Figures 2.16C-D. In contrast to user degree, identified users tended to access posts with a lower mean degree compared to other users, meaning they engaged with less popular content. Similarly, their mean post-weighted degree was also generally lower, suggesting participation in more isolated clusters. Overall, their engagement patterns aligned with habitual browsing of more

specialized CSA content.

Figure 2.17 presents a modified version of Figure 1.9, where red markers represent identified users and gray markers denote unidentified users. The distribution of identified users is not uniform across all engagement levels – a concentration of these individuals appears among users with a higher number of total views. Fewer identified users are observed in the low-engagement region, suggesting that Operation Darknet was more prevalent among frequent forum users. Notably, Figure 2.17B shows red markers distributed across varying activity durations, with a significant concentration in the longer time span region. This implies that many of the identified users were active on the forum for extended periods between their first and last visit. These users were likely part of a core group of frequent participants, either actively interacting with the forum through liking or commenting on posts, sharing CSAM, or "lurking" with a fixated need for CSA content consumption.

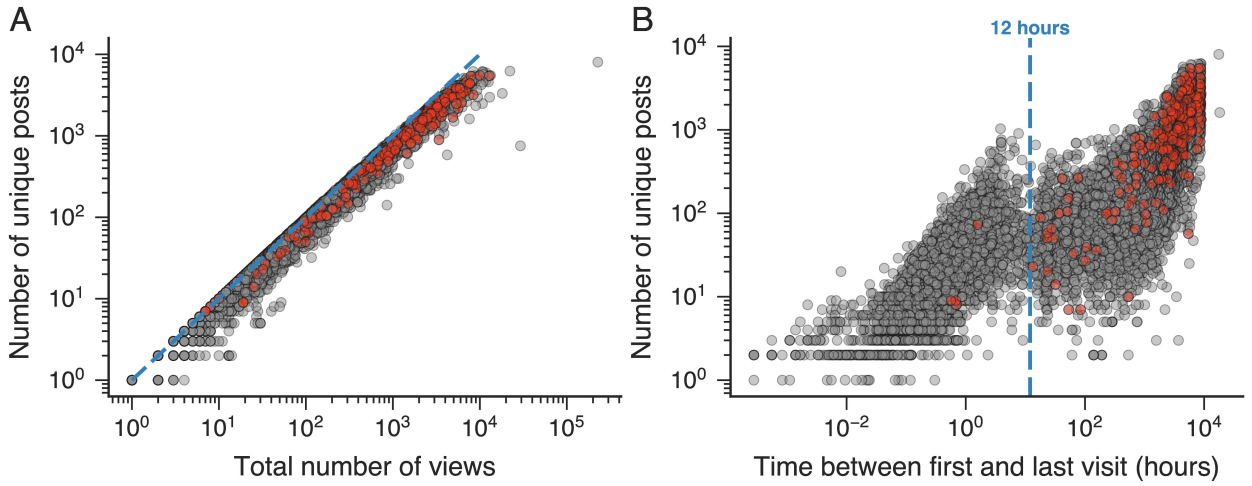

Figure 2.17: Number of unique posts as a function of (A) the total number of views per user, and (B) the time (in hours) between their first and last visit. The dashed blue line in Panel (A) represents a one-to-one linear relationship, while in Panel (B), it indicates the 12-hour mark. Red markers indicate identified users, while gray markers represent unidentified users.

Finally, we assessed the distribution of identified users across different hierarchical levels of the network depicted in Figure 2.15. This analysis provided additional insights into how these users were structurally embedded within the forum's interaction patterns, evaluating how their presence evolved across different depths of the hierarchy. The hierarchical nature of the network is particularly relevant in this context, as deeper levels offer a clearer perspective on the types of content accessed by identified users, capturing those who engaged with more

specialized or niche CSAM.

Figure 2.18 illustrates the number of identified users within each group at the third ($l = 3$) and fourth ($l = 4$) hierarchical levels (red bars) compared to their random expectation (gray bars), obtained through the permutation of identified user nodes. By contrasting the observed number of users with this randomized baseline, we determined whether identified users exhibited structural patterns that distinguished them from a null expectation. In both panels, we observed significant deviations in specific groups, where identified users are either overrepresented or underrepresented relative to the random expectation. This confirms that identified users were not randomly distributed but instead tended to concentrate in particular groups, demonstrating that this pattern is nontrivial and linked to intrinsic properties of the underlying network. As previously discussed, their navigation patterns were distinct, and Panel (A) indicates a tendency for identified users to cluster within specific subcommunities at deeper hierarchical levels, likely due to their engagement with certain content categories that were more frequently accessed by them.

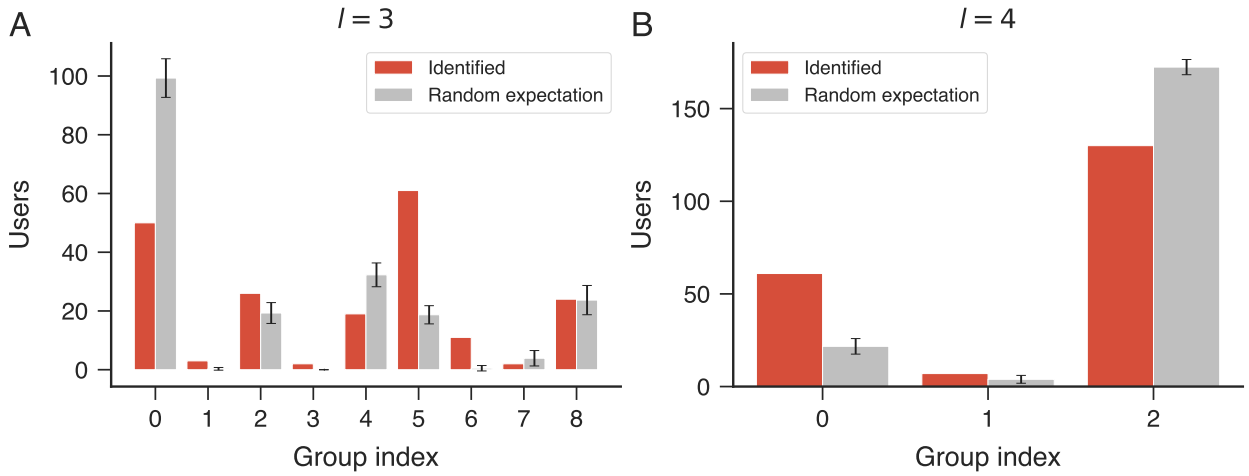

Figure 2.18: Distribution of users across groups in the (A) third and (B) fourth hierarchical levels. Red bars represent the observed number of identified users, while gray bars show the expected distribution under a random permutation of identified nodes. Error bars indicate the standard error of the mean.

Conclusions

In this dissertation, we investigated the structure and dynamics of a child sexual abuse (CSA) forum hosted on the dark web, applying tools from complex systems, network science, and insights from criminology literature. Leveraging privileged data collected during Operation Darknet – a landmark investigation conducted by the Brazilian Federal Police between 2013 and 2016 – we comprehensively analyzed user behaviors, content dissemination, and the evolving network topology of an illegal online community. We aimed to understand how such forums function, identify patterns that distinguish key users, and generate insights that could inform more effective law enforcement interventions.

Throughout our analysis, we consistently observed a highly asymmetric pattern of engagement. While the forum hosted nearly 10,000 registered members, most engaged passively, seldom or never contributing or interacting with others beyond consuming content. In contrast, a tiny core of users accounted for most posts, comments, and views. This disproportionate contribution, concentrated in less than one percent of users, reflected a distinct split between content producers and consumers. We also found evidence of compulsive behavior among some high-activity members who repeatedly accessed the same content over extended periods.

Despite the forum's illicit nature, it exhibited clear signs of internal organization. As the forum evolved, we observed the emergence of hierarchical and modular structures. Users interacted through posting and viewing and by validating each other's contributions with comments and likes. This revealed a social environment that, although deviant, mirrored le-

gitimate online communities in its mechanisms for rule enforcement and user validation. The increasing structural complexity of the network, reflected in the rise of hierarchical levels and modular communities, indicated that the forum was not static but continually adapting to user preferences and external pressures. Over time, subgroups formed around specific content categories, and members began to exhibit increasingly specialized engagement patterns.

We also analyzed the thematic distribution of posts and found that certain categories – particularly those suggesting real-world abuse – received elevated attention. Although posts in these categories represented a relatively small portion of the overall content, they consistently ranked among the most viewed and most liked. This suggests that the forum was a platform for content sharing and a space in which especially harmful material was promoted, reinforced, and rewarded.

From a network science perspective, we uncovered disassortative mixing in the forum's structure. Highly active users tended to avoid the most popular posts, and vice versa. This misalignment between user and post popularity may indicate strategic behavior – perhaps an attempt by key users to avoid detection or to engage within more isolated subcommunities. Additionally, the forum exhibited a high concentration of repeated interactions with a narrow subset of content, reinforcing that certain posts served as focal points of activity and were central to the network's cohesion.

In collaboration with law enforcement, we analyzed a subset of users whose IP addresses had been identified through investigative techniques. These users, many of whom were arrested or formally targeted, displayed markedly different behaviors compared to the rest of the forum population. They exhibited higher engagement levels, revisited posts more frequently, and were likelier to participate in niche or low-visibility content categories. Interestingly, they also tended to be structurally embedded within specific subcommunities of the network, particularly at deeper hierarchical levels. These patterns suggest that structural positioning within the network may be a useful predictor of offending behavior and that key actors are not randomly distributed but occupy identifiable regions of the interaction graph.

Through this work, we contributed to the growing literature on criminal network analysis by applying a data-driven approach to an especially sensitive domain. By modeling the forum as a bipartite network and using hierarchical stochastic block modeling, we gained insight into the formation and evolution of subcommunities, the development of thematic hierarchies, and the behavioral profiles of distinct user types. This methodological framework provides

a scalable and rigorous toolset for examining other online criminal platforms, whether they involve CSA content or other illicit activity.

We believe the implications of our findings extend beyond academic contributions. From a practical standpoint, our results underscore the importance of prioritizing individuals who display high-frequency engagement, sustained activity, and involvement with less popular but structurally significant content. These users are more likely to be central to the forum's persistence and resilience. Moreover, our results suggest that targeting high-impact content – posts that draw frequent revisits and reinforce user activity – may prove more disruptive than simply removing users. We also identified behavioral signatures that may serve as early warning indicators of risk, such as high-volume access in short time frames or repeat visits to specific types of content.

This study is, however, not without limitations. Our data reflect a single platform, and despite the considerable challenge of getting similar data from other sources, it remains an open question whether similar structures and dynamics are present in other CSA forums. Additionally, the ethical and legal constraints surrounding this type of data restrict our ability to directly analyze certain forms of content. Moving forward, we see several promising directions. One would be to examine cross-platform behavior to determine whether users display consistent patterns across multiple CSA forums. It would also be valuable to integrate other data types, such as textual information, into our investigation.

Finally, through the lens of complex systems and network science, we have shown that deviant behavior follows recognizable patterns even in hidden and illegal corners of the internet. These patterns can be measured, modeled, and ultimately disrupted. In doing so, we hope this work contributes to scientific understanding and the urgent societal effort to combat one of the most egregious forms of online crime.

Bibliography

[1] Jensen, H. J. *Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems* (Cambridge University Press, Cambridge, 1998).

[2] Mitchell, M. *Complexity: A Guided Tour* (Oxford University Press, New York, 2009).

[3] Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Reviews of Modern Physics* **81**, 591 (2009).

[4] Jusup, M. *et al.* Social physics. *Physics Reports* **948**, 1–148 (2022).

[5] Newman, M. E. J. *Networks: An Introduction* (Oxford University Press, New York, 2010).

[6] Barabási, A.-L. *Network Science* (Cambridge University Press, Cambridge, 2015).

[7] Estrada, E. What is a complex system, after all? *Foundations of Science* **29**, 1143–1170 (2024).

[8] Barabási, A.-L. Scale-free networks: A decade and beyond. *Science* **325**, 412–413 (2009).

[9] Euler, L. Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Petropolitanae* 128–140 (1741). English translation available at https://www.cantab.net/users/michael.behrend/repubs/maze_maths/pages/euler.html.

[10] Newman, M. The physics of networks. *Physics Today* **61**, 33–38 (2008).

[11] Barzel, B. & Barabási, A.-L. Universality in network dynamics. *Nature Physics* **9**, 673–681 (2013).

[12] White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society B* **314**, 1–340 (1986).

[13] Williams, R. J., Berlow, E. L., Dunne, J. A., Barabási, A.-L. & Martinez, N. D. Two degrees of separation in complex food webs. *Proceedings of the National Academy of Sciences* **99**, 12913–12916 (2002).

[14] Newman, M. E. J. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences* **98**, 404–409 (2001).

[15] Ribeiro, H. V., Alves, L. G. A., Martins, A. F., Lenzi, E. K. & Perc, M. The dynamical structure of political corruption networks. *Journal of Complex Networks* **6**, 989–1003 (2018).

[16] Martins, A. F., da Cunha, B. R., Hanley, Q. S., Gonçalves, S., Perc, M. & Ribeiro, H. V. Universality of political corruption networks. *Scientific Reports* **12**, 6858 (2022).

[17] Varese, F. The structure and the content of criminal connections: The Russian Mafia in Italy. *European Sociological Review* **29**, 899–909 (2013).

[18] Agreste, S., Catanese, S., Meo, P. D., Ferrara, E. & Fiumara, G. Network structure and resilience of mafia syndicates. *Information Sciences* **351**, 30–47 (2016).

[19] Cavallaro, L., Ficara, A., De Meo, P., Fiumara, G., Catanese, S., Bagdasar, O., Song, W. & Liotta, A. Disrupting resilient criminal networks through data analysis: The case of Sicilian Mafia. *PLOS ONE* **15**, e0236476 (2020).

[20] da Cunha, B. R. & Gonçalves, S. Topology, robustness, and structural controllability of the Brazilian Federal Police criminal intelligence network. *Applied Network Science* **2018**, 3–36 (2018).

[21] Duijn, P. A., Kashirin, V. & Sloot, P. M. The relative ineffectiveness of criminal network disruption. *Scientific Reports* **4**, 4238 (2014).

[22] Diviák, T., Dijkstra, J. K. & Snijders, T. A. Structure, multiplexity, and centrality in a corruption network: The Czech Rath affair. *Trends in Organized Crime* **22**, 274–297 (2019).

[23] Luna-Pla, I. & Nicolás-Carlock, J. R. Corruption and complexity: A scientific framework for the analysis of corruption networks. *Applied Network Science* **5**, 13 (2020).

[24] Granados, O. & Nicolás-Carlock, J. *Corruption Networks: Concepts and Applications*. Understanding Complex Systems (Springer, 2021).

[25] Smith, C. M. Exogenous shocks, the criminal elite, and increasing gender inequality in Chicago organized crime. *American Sociological Review* **85**, 895–923 (2020).

[26] da Cunha, B. R., MacCarron, P., Passold, J. F., dos Santos Jr, L. W., Oliveira, K. A. & Gleeson, J. P. Assessing police topological efficiency in a major sting operation on the dark web. *Scientific Reports* **10**, 73 (2020).

[27] ElBahrawy, A., Alessandretti, L., Rusnac, L., Goldsmith, D., Teytelboym, A. & Baronchelli, A. Collective dynamics of dark web marketplaces. *Scientific Reports* **10**, 18827 (2020).

[28] Bouchard, M. Collaboration and boundaries in organized crime: A network perspective. *Crime and Justice* **49**, 425–469 (2020).

[29] Chiang, Y.-S., Chang, Y.-C. & Weng, W. To blend in or hide out? a network analysis on maritime criminal co-voyages in taiwan. *Journal of Quantitative Criminology* **40**, 373–393 (2024).

[30] Lopes, D. D., Cunha, B. R. d., Martins, A. F., Gonçalves, S., Lenzi, E. K., Hanley, Q. S., Perc, M. & Ribeiro, H. V. Machine learning partners in criminal networks. *Scientific Reports* **12**, 15746 (2022).

[31] Ribeiro, H. V., Lopes, D. D., Pessa, A. A., Martins, A. F., da Cunha, B. R., Gonçalves, S., Lenzi, E. K., Hanley, Q. S. & Perc, M. Deep learning criminal networks. *Chaos, Solitons & Fractals* **172**, 113579 (2023).

[32] Toledo, A. S. O., Carpi, L. C., Atman, A. P. F. & Scarpelli, A. P. B. Multiplex key roles to disrupt criminal networks. *Social Network Analysis and Mining* **13**, 98 (2023).

[33] Toledo, A. S. O., Carpi, L. C., Atman, A. P. & Scarpelli, A. B. Outlier mining in criminal networks: the role of machine learning and outlier detection models. *Journal of Computational Social Science* **8**, 36 (2025).

[34] Divakarmurthy, P., Requião da Cunha, B., Passold, J. F., Oliveira, M. & Menezes, R. Unravelling the dynamics of child sexual exploitation material circulation on the dark web. *PLOS ONE* **19**, e0306516 (2024).

[35] Nearchou, N. *Combating Crime on the Dark Web: Learn how to access the dark web safely and not fall victim to cybercrime* (Packt, 2023).

[36] Gannon, C., Blokland, A. A. J., Huikuri, S., Babchishin, K. M. & Lehmann, R. J. B. Child sexual abuse material on the darknet. *Forensische Psychiatrie, Psychologie, Kriminologie* **17**, 353–365 (2023).

[37] IWF. Iwf annual report 2023. https://www.iwf.org.uk/annual-report-2023/trends-and-data/reports-analysis/ (2023). Accessed on 16 Mar 2025.

[38] Owen, G. & Savage, N. The tor dark net. Tech. Rep. 20, Centre for International Governance Innovation (2015).

[39] Nurmi, J., Paju, A., Brumley, B. B., Insoll, T., Ovaska, A. K., Soloveva, V., Vaaranen-Valkonen, N., Aaltonen, M. & Arroyo, D. Investigating child sexual abuse material availability, searches, and users on the anonymous tor network for a public health intervention strategy. *Scientific Reports* **14** (2024).

[40] Associated Press. Creator of 'world's largest kiddie-porn site' gets 30 years in prison as police arrest 900 suspected users and rescue hundreds of exploited children. https://www.dailymail.co.uk/news/article-4479240/Playpen-child-porn-king-Steven-Chase-gets-30-years-prison.html (2017). Accessed on 16 Mar 2025.

[41] Owens, J. N., Clapp, K., Craun, S. W., van der Bruggen, M., van Balen, I., van Bunningen, A. & Talens, P. Analysis of topic popularity within a child sexual exploitation tor hidden service. *Aggression and Violent Behavior* **68** (2023).

[42] da Cunha, B. R., MacCarron, P., Passold, J. F., dos Santos Jr., L. W., Oliveira, K. A. & Gleeson, J. P. Assessing police topological efficiency in a major sting operation on the dark web. *Scientific Reports* **10** (2020).

[43] Zamani, M., Rabbani, F., Horicsányi, A., Zafeiris, A. & Vicsek, T. Differences in structure and dynamics of networks retrieved from dark and public web forums. *Physica A: Statistical Mechanics and its Applications* **525**, 326–336 (2019).

[44] van der Bruggen, M. & Blokland, A. Profiling darkweb child sexual exploitation material forum members using longitudinal posting history data. *Social Science Computer Review* **40**, 865–891 (2021).

[45] van der Bruggen, M., van Balen, I., van Bunningen, A., Talens, P., Owens, J. N. & Clapp, K. Even "lurkers" download: The behavior and illegal activities of members on a child sexual exploitation tor hidden service. *Aggression and Violent Behavior* **67** (2022).

[46] Ngo, V. M., Gajula, R., Thorpe, C. & Mckeever, S. Discovering child sexual abuse material creators' behaviors and preferences on the dark web. *Child Abuse & Neglect* **147** (2024).

[47] van der Bruggen, M. & Blokland, A. A crime script analysis of child sexual exploitation material fora on the darkweb. *Sexual Abuse* **33**, 950–974 (2021).

[48] Blokland, A. *et al.* Why do users continue to contribute to darknet child sexual abuse material forums? examining social exchange, social capital, and social learning explanations using digital forensic artifacts. *Child Abuse & Neglect* **153** (2024).

[49] Kleinman, Z. Child sexual abuse: Four held in german-led raid on huge network. https://www.bbc.com/news/world-europe-56969414 (2021). Accessed on 16 Mar 2025.

[50] Ministério Público Federal. Operação darknet. https://www.mpf.mp.br/rs/sala-de-imprensa/docs/outros-documentos/operacao-darknet (2016). Accessed on 16 Mar 2025.

[51] BBC News. Brazil police crack 'darknet' in child pornography crackdown. https://www.bbc.com/news/world-latin-america-29639241 (2014). Accessed on 16 Mar 2025.

[52] Choi, K.-S. & Lee, H. The trend of online child sexual abuse and exploitations: A profile of online sexual offenders and criminal justice response. *Journal of Child Sexual Abuse* **33**, 804–823 (2024).

[53] Huikuri, S. Users of online child sexual abuse material. *Journal of Police and Criminal Psychology* **38**, 904–913 (2023).

[54] Kranenbarg, M. W. & Leukfeldt, R. *Cybercrime in Context: The human factor in victimization, offending, and policing.* Crime and Justice in Digital Society (Springer, 2021).

[55] Tagarelli, A. & Interdonato, R. "who's out there?": Identifying and ranking lurkers in social networks. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 215–222 (Association for Computing Machinery, 2013).

[56] Woodhams, J., Kloess, J. A., Jose, B. & Hamilton-Giachritsis, C. E. Characteristics and behaviors of anonymous users of dark web platforms suspected of child sexual offenses. *Frontiers in Psychology* **12** (2021).

[57] Latapy, M., Magnien, C. & Del Vecchio, N. Basic notions for the analysis of large two-mode networks. *Social Networks* **30**, 31–48 (2008).

[58] Gerlach, M., Peixoto, T. P. & Altmann, E. G. A network approach to topic models. *Science Advances* **4**, eaaq1360 (2018).

[59] Peixoto, T. P. Hierarchical block structures and high-resolution model selection in large networks. *Physical Review X* **4**, 011047 (2014).

[60] Peixoto, T. P. Model selection and hypothesis testing for large-scale network models with overlapping groups. *Physical Review X* **5**, 011033 (2015).

[61] Peixoto, T. P. Nonparametric bayesian inference of the microcanonical stochastic block model. *Physical Review E* **95**, 012317 (2017).

[62] Vaganov, D., Bardina, M. & Guleva, V. From generality to specificity: On matter of scale in social media topic communities. In *Computational Science – ICCS 2020*, vol. 12140, 305–318 (Springe, 2020).

[63] Rowe, M. & Strohmaier, M. The semantic evolution of online communities. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14 Companion, 433–438 (Association for Computing Machinery, 2014).

[64] Latapy, M., Magnien, C. & Vecchio, N. D. Basic notions for the analysis of large two-mode networks. *Social Networks* **30**, 31–48 (2008).

[65] Newman, M. E. J. The structure and function of complex networks. *SIAM Review* **45**, 167–256 (2003).

[66] Pishro-Nik, H. *Introduction to probability, statistics, and random processes* (Kappa Research LLC, 2014).

[67] Scipy.stats.pearsonr — SciPy v1.15.2 Manual. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html. Accessed on 16 Mar 2025.

[68] Moore, D. S., McCabe, G. P. & Craig, B. A. *Introduction to the Practice of Statistics* (W. H. Freeman and Company, 2017), 9th edn.