Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-024-03954-x OPEN

Missing the mark? Identifying child sexual abuse material forum structure and key-players based on public replies and private messaging networks

Frederic M. Gnielka[1], Rebecca Reichel[1], Arjan Blokland[2], Anton Daser[3], Meike de Boer[4], Colm Gannon[5], Alexander F. Schmidt[3✉], Thomas Schäfer[1], Salla Huikuri[6], Katarzyna Staciwa[7] & Robert J. B. Lehmann[1]

Darknet forums dedicated to child sexual abuse material (CSAM) attract thousands of users interacting with each other through online communications. Given finite resources, law enforcement agencies seek ways to effectively prioritise their investigative efforts by identifying key-players that are central to the forum community. For the identification of such users, law enforcement agencies typically rely on the communication network that can be derived from messages posted on the public part of the forum. Many forums, however, also allow for private communications between members, raising the question to what extent relying on only a single mode of communication biases key-player identification. Using data on both public and private communications on two large-scale darknet CSAM forums, two communication networks are derived and their structures analysed. Measures of centrality robustness are applied to ascertain the level of bias introduced when determining key-players on only one of the available networks. Findings show only a minority of members to participate in forum communication, and limited overlap between participants active in public and private communications. Key-players emerging from combining the public and private communications resemble those from the public network only, suggesting that police prioritisation based on public postings only is still 'on mark'. Members who are central to the private communications network may nevertheless be of special law enforcement interest.

[1] Medical School Berlin, Berlin, Germany. [2] Netherlands Institute for the Study of Crime and Law Enforcement & Leiden University, Leiden, The Netherlands. [3] Johannes Gutenberg University Mainz, Mainz, Germany. [4] Netherlands Institute for the Study of Crime and Law Enforcement, Amsterdam, The Netherlands. [5] La Trobe University, Melbourne, Australia. [6] Save the Children Finland, Helsinki, Finland. [7] Polish Platform for Homeland Security, Poznań, Poland. ✉email: alexander.schmidt@uni-mainz.de

Introduction

Cloaking users from publicity by random paths of encrypted servers, the darknet currently hosts hundreds of hidden services that function as so-called virtual offender convergence settings, i.e. online forums where offenders congregate to exchange illegal goods and services, share information, and socialise with like-minded others (Soudijn and Zegers, 2012). These darknet forums may be dedicated to different crimes like the trade in controlled substances, firearms, stolen credit card details, or malicious software (e.g. Goonetilleke et al., 2023; Jiang et al., 2021; Kigerl, 2022). A sizable subset of these forums orbits around the exchange of child sexual abuse material (CSAM; Gannon et al., 2023; IWF, 2023). Unaffected by physical and geographical constraints, the number of users interacting in these virtual offender convergence settings can reach thousands, if not tens of thousands. This greatly challenges the resources available to law enforcement to combat these types of cyber-enabled crimes (Europol, 2023), and forces agencies to prioritise and focus their investigative efforts on those they deem key to the criminal exchange patterns (e.g. da Cunha et al., 2020).

Identifying key targets for prioritisation by law enforcement. Following research on offline criminal groups and, more recently, Clearnet social media, law enforcement agencies have turned to social network analysis to help identify *key-players* as targets for their investigative efforts (Burcher and Whelan, 2018; Duijn and Klerks, 2014; for a comprehensive introduction to and overview of social network analysis, see McLevey et al., 2023). Key-players are typically defined as those taking up central positions in a network (Fonhof et al., 2019). Network centrality is seen as an important proxy for social capital (Ganley and Lampe, 2009) and is taken to be indicative of the individual's level of activity, connectedness, influence, prestige, and popularity in a community. In addition, centrality may signal individuals' access to information, resources, and criminal opportunities (Morselli, 2010).

In the context of darknet criminal marketplaces, key-players are typically defined based on their position in the forums' public communications network (Huang et al., 2021; Pete et al., 2020). On darknet forums where illegal goods and services are sold, bought, or bartered, online exchanges take place between parties oblivious to each other's true identity. In the absence of offline ties between forum members, public forum communication is an important vehicle to establish trust and build a reputation in the community (Duxbury and Haynie, 2021; Yip et al., 2013). Still, prior research typically finds that only a small subset of forum members takes part in public forum communication and that even among those publicly communicating on the forum the extent of contributions is heavily skewed, with a small number of highly communicative members responsible for a disproportionate share of all public forum communication (e.g. Sun et al., 2014; Zamani et al., 2019). Prioritising investigative efforts on key-players appears to have some merit, as centrality based on public communications in darknet criminal marketplaces has been linked to risk indicators such as, for example, higher involvement in discussions of cybercriminal activities and identity security management (Pete et al., 2020) or vendor success (Boekhout et al., 2024).

In prior research on key-players in the context of online CSAM, two streams of research can be distinguished. A number of studies have focused on darknet CSAM forum websites as their unit of measurement. These studies have identified structural properties of darknet CSAM forums and the relationships between them, e.g. through the number of times different websites referred to one another, or the amount of CSAM available on these sites (e.g. Westlake and Bouchard, 2015; Westlake and Bouchard, 2016; Westlake and Frank, 2017). This approach might be beneficial for prioritising forums for law enforcement take-downs in an effort to distort the online infrastructure facilitating the exchange of CSAM. A second stream of research is focussed on analysing relationship between users within a particular darknet CSAM forum. Like studies on darknet criminal marketplaces, these studies identified key forum members based on their position in the public communications network (Fonhof et al., 2019; Fortin et al., 2021). This way of identifying key-players is especially relevant for prioritisation of investigative efforts aimed at the apprehension and prosecution of high-risk individuals. Prior research showed that the most central individuals in the public communication network were found to provide the CSAM that resulted in the most views by far (da Cunha et al., 2020). It is this second stream of research that is of particular relevance to the current study.

Private messaging as an underdeveloped research focus. Although for reasons of practicality, public forum communication has been the focus of research on darknet CSAM use so far, many darknet forums also offer members the possibility to communicate with other forum members privately. Public forum communications are visible to the entire forum membership, and all forum members can contribute to a particular thread. In contrast, private messaging occurs member-to-member, or in ad hoc groups of forum members, and is concealed from other forum members. Typically, only forum administrators have an overview of all private conversations taking place in the forum community. Data on private messages is therefore not available in forum scrapes (as, for example, described by Bergman and Popov 2023), as crawlers are only able to scrape the public part of darknet forums. The same is true for (covertly operating) law enforcement officers conducting live investigations on active forums who only have access to the public parts of other members' communication (e.g. da Cunha et al., 2020). Only forum take-downs are sometimes able to secure both public and private forum communications (e.g. Afroz et al., 2013; Motoyama et al., 2011; Smirnova et al., 2024). Depending on its volume and structure, neglecting private communications in determining individuals' network centrality may lead key-player assignment, and law enforcement prioritisations resulting from this, to be misjudged. To assess the extent to which key-player assignment is biased, data on both public and private forum communications are pivotal to understand the precision of law enforcement target identification (Smirnova et al., 2024).

More formally, due to the presence of different modes of communication—i.e. public and private –, darknet forum communications can be constructed within the framework of a multiplex network (Contractor et al., 2011; Halu et al., 2013; Magnani and Rossi, 2011); a network in which individuals (nodes) are connected through multiple types of ties (edges). One way to think of multiplex networks is to distinguish different layers, in which each layer represents edges of a different nature (Magnani et al., 2021). By starting a public thread, forum users address all other forum users simultaneously, exposing themselves to their potential replies and giving up control over the entailing discussion. Contrarily, private messages are directed at designated fellow forum users the sender chooses to communicate with and who are the only ones who will be able to read and reply to these messages. In other words, "as users can send messages directly to one another and at the same time participate in discussion groups within a forum, they can be regarded as embedded in two related online social networks" (Halu et al.,

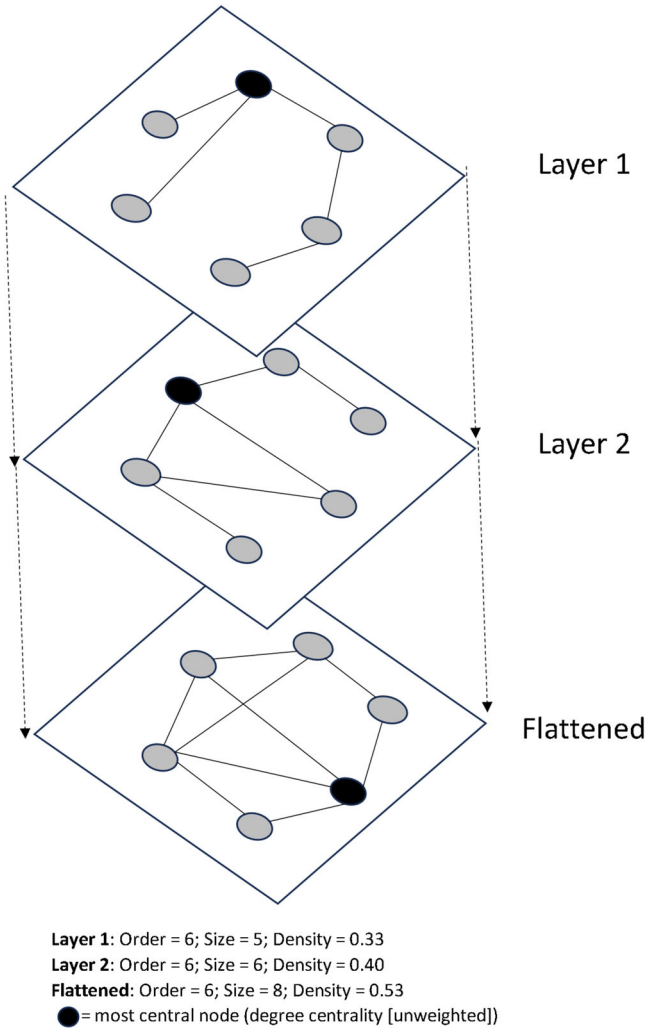

Fig. 1 Schematic of a fictitious multiplex network with two layers. A fictitious multiplex network with two layers and a flattened layer onto which the two layers are collapsed. Order, size, density, and the most central node per layer/network are displayed.

2013, p. 2). While multiplex networks are often used to model different types of relationships between individuals (e.g. public figures and friends; Magnani and Rossi, 2011), they have also successfully been used to represent different modes of communication, such as private and public communication on online forums (Halu et al., 2013) or even phone calls and text messages (Gaito et al., 2017).

Figure 1 depicts a (fictitious) darknet forum communication network consisting of two layers representing public replies and private messages respectively. The flattened network combines information from both layers in a single network representation. Figure 1 reports the order, size, and network density for both layers and for the flattened layer. The most central vertex in each of these networks is coloured black. What becomes evident from Fig. 1 is that for determining node centrality, the layer(s) taken into account could indeed matter; neither of the most central nodes in layer 1 and layer 2 is the key player in the exemplary flattened network combining public and private ties presented here.

To the best of our knowledge, prior research on darknet CSAM forums identified key-players based on members' public activity only (e.g. da Cunha et al., 2020; Fonhof et al., 2019; Fortin et al., 2021). Yet, there are a few studies on online criminal markets that were able to examine both public and private forum communications (Afroz et al., 2013; Motoyama et al., 2011; Overdorf et al., 2018; Smirnova et al., 2024). Those have identified discrepancies both in the number of members involved in each type of communicative behaviour, as well as in members' level of activity across layers (Motoyama et al., 2011; Overdorf et al., 2018; Smirnova et al., 2024). It has even been suggested that public and private communications might serve different purposes (Afroz et al., 2013; Smirnova et al., 2024) and that private messaging may represent relations between users most accurately (Yip et al., 2013). Due to the sensitive and criminal nature of the communications in a darknet CSAM forum, it is likely that, similar to other darknet forums, some individuals will choose to only communicate privately even within this hidden community of like-minded individuals (Smirnova et al., 2024). Disregarding the different forms of forum communications may therefore result in missing the mark when determining who key-players in the forum community are (see also Sun et al., 2019).

While access to all private messages from darknet forums is not impossible, it remains rare in research (Overdorf et al., 2018). Accordingly, authors of previous studies have called for replication of their results, especially on criminal forums in the darknet (Smirnova et al., 2024), and the development of measures that can reliably operate on subsets of data (e.g. only public posts without private messages) has been formulated as an explicit research goal in the field (Afroz et al., 2013). In other words, as the effective allocation of law enforcement resources depends heavily on the accuracy with which key-players can be identified, estimates of the level of bias introduced by only considering public forum communication are necessary. The current research aims to provide further insight in the accuracy of key-player assignment on darknet CSAM forums based on public communications data only by addressing the following research questions:

- How do the size and structure of communication networks underlying public and private parts of darknet CSAM forums compare?
- To what extent is key-player assessment on darknet CSAM forums biased by its reliance on public communication data only?

Methods

Data. To answer our research questions, we use data from two darknet CSAM forums made available to us by law enforcement agencies. While the main language on both forums was English, there were also posts and private messages in several other languages.

Forum A was online for approximately one and a half years, from April 2016 until the forum was seized by law enforcement in September 2017. It can be described as a general CSAM forum with the self-proclaimed goal "to provide a simple free access forum to the community, while simultaneously allowing a safe and secure place to talk and just be ourselves" (statement taken from one of the forum's introductory posts), offering a variety of topical subforums, such as softcore, hardcore, and fetishes, as well as the opportunity to engage in private conversations, which were also available in the seized data. The forum was run by two administrators, had 11 (sub-)forum moderators, and 60 members that were designated as either VIP (very important person; $n = 53$) or MVP (most valuable player; $n = 10$). Note that some users had several roles at the same time.

Forum B described itself as a "Boyloveronly [sic]" forum (statement taken from one of the forum's introductory posts) and was active for approximately one year from November 2021 to

Table 1 Forum descriptions.

Descriptions	Forum A	Forum B
Total registered users[a]	936,110	592,345
	Public posts[b]	
Total	194,551	656,824
Thread starters (percentage)	25,235 (13.0%)	29,995 (4.6%)
Md per day (*MAD*)	369 (78.5)	1715 (343)
	Private messages	
Total	58,484	61,992
Md per day (*MAD*)	112.5 (30.5)	163 (40)

MAD median absolute deviation.
[a]Excluding $n = 39$ users with faulty user IDs from forum A and $n = 1$ users with faulty user IDs from forum B.
[b]Excluding $n = 431$ posts from or replies to faulty user IDs and $n = 1236$ posts that replied to non-existing threads in forum A and $n = 4503$ posts from or to faulty user IDs and $n = 1026$ posts that were replies to non-existent threads on forum B.

November 2022. Even though it was smaller in terms of number of users than forum A, it had much more public posts and a higher median of messages posted per day. There were two (co-) administrators, 17 moderators, and five so-called "re-up doctors", a role that was described on the forum as an "employee who takes care of dead left. And makes backup copies of new files". Table 1 shows descriptive statistics for both forums.

When the forums were seized, law enforcement agencies secured data on the forums' histories in SQL (Structured Query Language) data dumps. As a result, we can no longer ascertain the reasons for any missing data (e.g. data might have been missing due to deleted posts, users who left the forum, or data loss during the take-down of the forum). The current analysis uses digital trace data on both the public replies and private messaging on these forums.[1]

Defining the networks. From the digital trace data available, we defined the forums' communication networks as multiplex, in the sense that each first layer pertained to public communications, each second layer to private messaging, and each flattened network represented the combination of the two layers (see also Halu et al. 2013 for a similar multiplex network). *Public communications* are organised in threads that consist of an initial post, which can be followed by one or more replies. While most replies are direct answers to the thread starter, users can also reply directly to another user's reply. From these posts and replies, we constructed the public replies networks; directed networks in which nodes represent forum members, in-coming edges represent replies received, and out-going edges represent replies sent. This meant that $n = 728$ users who only posted one or more original posts yet received no replies and also never replied to public contributions of others were excluded from the network of forum A, and $n = 268$ from forum B. With this definition of the network architecture, we captured all directed public communications on the forums, while purposefully ignoring undirected communications (i.e. thread starters). For a similar reason, all loops were removed for the social network analyses. Even though it seems intuitive that users can reply to their own posts, the goal of the analyses was to quantify users' interactions with others, and therefore loops would have biased the results insofar as the degrees would not have exclusively reflected social interactions with others.

In the *private messaging networks*, nodes again represent forum members and directed edges represent private messages sent. Again, loops (i.e. users replying to themselves) were removed from the private messaging network. Edges in both the public and private messaging network are weighted so that their weight reflects the respective number of replies and private messages exchanged between members. Note that nodes and edges between nodes may be present in both layers or in one layer only. Formally, the forum's communication networks are therefore modelled as multiplex networks in this study.

For the *flattened networks*, the two layers were collapsed into one, so that each node and each edge from each layer was present, respectively. The weights of the edges in the flattened network represent the summed number of all interactions between each two users in both layers.

Analytic strategy. All analyses were carried out in R (R Core Team, 2021). To analyse similarity between the different layers in the CSAM forum communication networks, we followed the approach taken by Bródka et al. (2018), where a multiplex network is represented as a property matrix that allows for simple comparisons across layers, e.g. by comparing the average degree across layers and the degree distribution across nodes in each layer, or by correlating node degree centralities between layers. Bródka et al. (2018) list a number of metrics that can be used to compare network structures across layers. Networks were constructed and analysed in R (R Core Team) with the igraph package (version 2.0.3; Csárdi et al., 2024).

Descriptive measures were calculated separately for all network layers. For the global transitivity index (see below), the network layers had to be transformed into undirected networks without multi-edges or weight attributes. Degrees were compared between layers via Spearman rank correlations, as recommended by Bródka et al. (2018) in the presence of severely skewed distributions (see below). In addition, we follow previous work by Borgatti et al. (2006) on the robustness of centrality measures in the face of random error to assess the comparability of node centrality rankings across layers. Several measures of centrality robustness, as recommended by Borgatti et al. (2006), were assessed. These measures are summarised in Table 4.

Results

Network layers. Table 2 gives an overview over a number of network-based measures for the different layers. Of the 936,110 members registered on forum A, 23,120 (2.47%) forum users were present in the public replies network, generating 106,779 unique edges. On forum B, 2.69% ($N = 15,491$) of the 592,345 registered users were in the public replies network, creating 174,192 edges.

The private messaging networks were smaller for both forums, with only 12,279 (1.31%) forum users sending at least one private message on forum A, and 3752 (0.63%) users on forum B. Together, users who were active on the private part of the forums generated 29,732 edges on forum A, and 19,089 edges on forum B.

Combining the public and the private messaging networks resulted in flattened networks with 28,364 nodes (3.03% of all forum members) for forum A, and 15,721 (2.65% of all forum members) on forum B. This indicates that, while many communicating forum members actively contribute to both public and private forum communication, part of the forum membership is active in one type of communication only. Because of the relatively small order of the private messaging network of forum B, the public replies network and the flattened network resembled each other more than on forum A.

In directed graphs, strong-connected components represent sets of nodes that are connected by reciprocal paths, while weak-connected components represent sets of nodes connected when ignoring edge direction. The high numbers of users in the main weak components of both the public replies and the private messaging layers in both forums indicated general high

Table 2 Descriptive measures per network.

Network	Order	Size	No. components		Users in main (weak) component		Density	Total degree		Centralisation		Avg. path length	Diameter	Global transitivity
			Strong	Weak	Total	Percentage		Geometric mean (*SD*)	Median (*MAD*)	In-degree	Out-degree			
Forum A														
Public replies	23,120	106,779	14,728	29	23,060	99.7	0.0002	2.72 (3.21)	2.00 (1.00)	0.05	0.05	4.06	67	0.04
Private messaging	12,279	29,732	6,877	233	11,778	95.9	0.0002	2.07 (2.57)	2.00 (1.00)	0.05	0.04	5.70	86	0.02
Flattened	28,364	130,939	16,543	62	28,234	99.5	0.0002	2.67 (3.18)	2.00 (1.00)	0.05	0.04	4.25	98	0.04
Forum B														
Public replies	15,491	174,192	7120	16	15,459	99.8	0.0007	4.57 (4.44)	3.00 (2.00)	0.11	0.06	3.73	49	0.11
Private messaging	3752	19,089	1349	40	3650	97.3	0.0014	3.67 (3.44)	3.00 (2.00)	0.13	0.12	5.58	142	0.06
Flattened	15,721	185,012	7074	15	15,691	99.8	0.0008	4.65 (4.54)	2.00 (2.00)	0.12	0.07	3.75	59	0.11

Order = number of nodes; Size = number of edges; No. components = number of components; *MAD* = median absolute deviation; Avg. path length = average path length; Diameter = longest distance between any two nodes. The global transitivity index was calculated with undirected networks for all networks and layers.

Table 3 Node and edge presence across layers.

	Forum A		Forum B	
	Public replies		Public replies	
Private messaging	Yes	No	Yes	No
	Nodes (Kulczynski = 0.67)		Nodes (Kulczynski = 0.71)	
Yes	7035	5244	3522	230
No	16,085	0	11,969	0
	Edges (Kulczynski = 0.96)		Edges (Kulczynski = 0.95)	
Yes	5572	24,160	8269	10,820
No	101,207	0	165,923	0

Kulczynski $= 1 - \left(\frac{p_1 \cap p_2}{p_1 p_2 + p_2 p_1}\right)$.

connectivity between all active forum members, meaning that some sort of path existed between almost any two nodes. Adding the private messaging network to the public replies network increased the number of components, while the percentage of members in the main component was relatively unchanged. This suggests that the small disconnected subsets observed in the private messaging network remained unconnected even when considering public communications between members.

The density of the public replies network, which can be interpreted as the probability that a pair of randomly chosen nodes are connected, on the other hand, was low for the public replies, the private messaging, and the flattened network, which is to be expected given the limited number of members engaging in communications. On average, each active member was connected to approximately two to three other members in all layers on forum A, and to two to five on forum B. The low centralisation for both in- and out-degrees on all layers indicates that these networks were not dominated by a single node, even though centralisation was higher for forum B, where each of the ten most active users had disproportionately higher numbers of posts than the most active user on forum A (see below). The diameter on all layers was relatively large compared to the respective average path length, suggesting the presence of hubs or highly connected nodes. This is a first indication that the distribution of degree centralities across nodes in the network was skewed. In the light of that, the high connectivity for the public replies network reported above could also be attributed to some highly influential hubs, rather than a balanced communication between all forum members. The average path length in both private messaging layers was even higher than that of the public replies layers, indicating lower cohesion.

Finally, global transitivity represents the proportion of triangles in the network and can be interpreted as the likelihood that two nodes have a common contact. In the public replies networks, transitivity was rather low, and even lower on the private messaging network, which again reflects its lower cohesion. However, the transitivity on forum B was generally higher than on forum A.

Layer comparison. Table 3 shows the number of nodes and edges that were present in the different layers respectively. Additionally, the Kulczynski distances in the network layers are displayed, i.e. the intersection of nodes and edges between each two of the layers in the multiplex network in relation to their symmetric difference. For the forums under scrutiny, we found that on forum A 30.43% ($n = 7035$) of all members present in the public replies network were also active in the private messaging part of the forum. On forum B, this number was slightly lower with 22.74% ($n = 3522$).

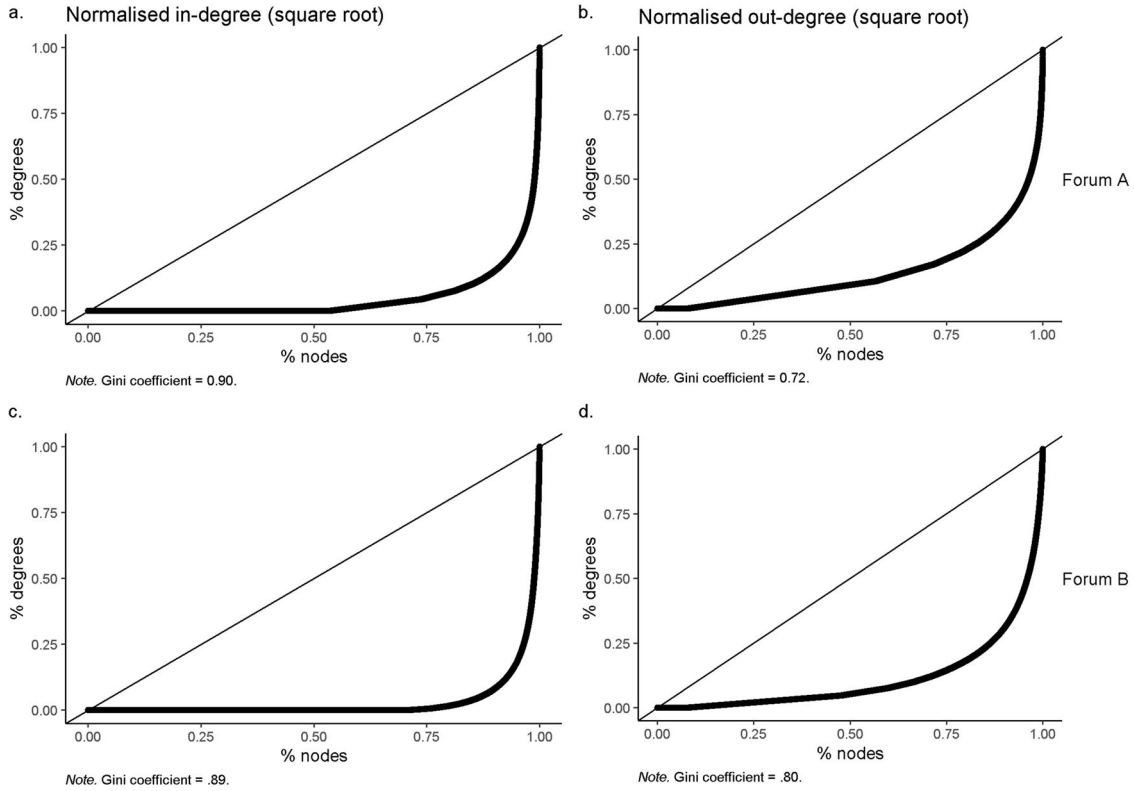

Fig. 2 Degree distributions of the public replies layer. Lorenz plots of in- (**a**, **c**) and out-degrees (**b**, **d**) of nodes in the public replies layer on forum A (**a**, **b**) and forum B (**c**, **d**). Square roots of the normalised in- and out-degrees were used due to the skewness of the distributions and high number of zeroes.

The majority (69.57% on forum A; 77.26% on forum B) of members sending or receiving public replies thus did not engage in private communications, suggesting that engaging in public conversation may represent a lower threshold for communicating. Still, we also found that 42.71% ($n = 5244$) of members active on the private part of the forum did not contribute any public replies on forum A. Again, this number was lower on forum B, with only 6.13% ($n = 230$) of members engaging in private communications while not engaging in public replies.

Also shown in Table 3 is the number of edges that were present in either network. Only 5.22% ($n = 5572$) of edges present in the public replies network were also present in the private messaging network on forum A, and 4.75% ($n = 8269$) on forum B. More interestingly, 81.26% ($n = 24{,}160$) of edges in the private messaging network of forum A, and 56.68% ($n = 10{,}820$) of forum B, were not mirrored by a similar edge in the public replies network. This indicates that forum members were connected by private messaging in ways that do not become apparent when only considering the public part of the forum, even though this difference was much more pronounced on forum A than on forum B. This was also reflected in the higher values of the Kulczynski distances corroborating higher dissimilarities for edges than for nodes on both forums.

The comparison of the network diameter to the average path length (Table 2) suggested the presence of hubs in both the public and private communication network. To provide a more detailed comparison of both networks, the degree distributions across layers were compared next. First, for the public replies network we found that of the 23,120 users posting at least one reply on the public part of forum A, 53.30% ($n = 12{,}322$) had an in-degree of 0, indicating that these members replied at least once on the public part of the forum but never elicited any reaction from the community (Fig. 2). This number was lower on forum B, with 38.91% ($n = 6028$) of the 15,491 members in the public replies network exhibiting an in-degree of 0. Only 7.86% ($n = 1817$) of members active in the public replies network on forum A, and 6.06% ($n = 938$) on forum B, showed an out-degree of 0, indicating that while they never replied to the public communications of others, others replied to theirs. On forum A, the highest number of replies a user sent was 3659 ($Md_{zeroes+} = MAD_{zeroes+} = 1.00$; $Md_{zeroes-} = 2.00$; $MAD_{zeroes-} = 1.00$), while the highest number of replies received was 4735 ($Md_{zeroes+} = MAD_{zeroes+} = 0.00$; $Md_{zeroes-} = 2.00$; $MAD_{zeroes-} = 1.00$).[2] The average number of replies each user sent per day ranged from 0 to 46 and from 0 to 129 for received replies per day. Note, however, that some of the high maximum values of daily sent or received replies (and, respectively, private messages) were partly caused by users who were highly active within only short periods of time. On forum B, the most active user sent with 14,115 replies almost four times as many as the most active user on forum A ($Md_{zeroes+} = Md_{zeroes-} = 2.00$; $MAD_{zeroes+} = MAD_{zeroes-} = 1.00$), while the highest number of received replies was with 62,700 more than 13 times as high as forum A ($Md_{zeroes+} = MAD_{zeroes+} = 1.00$; $Md_{zeroes-} = 3.00$; $MAD_{zeroes-} = 2.00$). On average, each user sent between 0 and 101 replies per day and received between 0 and 365.

The degree distributions for the private messaging layer are depicted in Fig. 3. We found that 39.32% ($n = 4828$) of members active on the private part of forum A did not receive a single private message from others. Given their presence in the private messaging network, this indicates that these members sent a private message to at least one other member but did not receive an answer. Again, this number was much lower on forum B, with 6.26% of users in the private messaging network having an out-degree of 0. Similarly, we found that 14.95% ($n = 1836$) of members involved in private messaging on forum A received at least one private message but never replied. This number was higher on forum B (28.68%; $n = 1076$). The highest number of

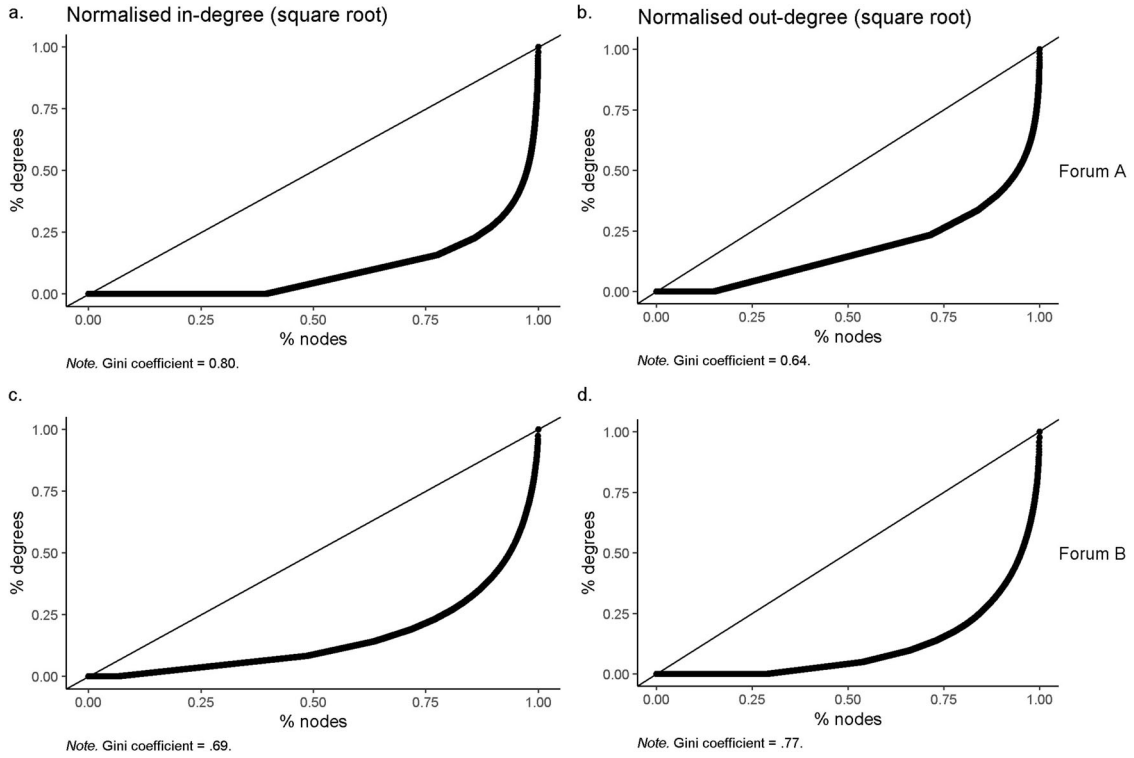

Fig. 3 Degree distributions of the private messaging layer. Lorenz plots of in- (**a**, **c**) and out-degrees (**b**, **d**) of nodes in the private messaging layer on forum A (**a**, **b**) and forum B (**c**, **d**). Square roots of the normalised in- and out-degrees were used due to the skewness of the distributions and high number of zeroes.

private messages sent by one user on forum A was 2429 ($Md_{zeroes+} = MAD_{zeroes+} = 0.00$; $Md_{zeroes-} = 1.00$; $MAD_{zeroes-} = 0.00$), while the highest number of private messages received was 3170 ($Md_{zeroes+} = MAD_{zeroes+} = 0.00$; $Md_{zeroes-} = MAD_{zeroes-} = 1.00$). On average, each user sent between 0 and 23 private messages per day and received between 0 and 31 private replies per day. On forum B, the user sending the most private messages sent with 2926 approximately as many as the one on forum A ($Md_{zeroes+} = MAD_{zeroes+} = 0.00$; $Md_{zeroes-} = 4.00$; $MAD_{zeroes-} = 3.00$), which was with 3646 also true for most private messages received ($Md_{zeroes+} = MAD_{zeroes+} = 0.00$; $Md_{zeroes-} = 3.00$; $MAD_{zeroes-} = 2.00$). The average number of private messages a user sent per day ranged from 0 to 17, and from 0 to 15 for private messages received.

Comparing Figs. 2 and 3, we found that both in- and out-degree distributions in the public networks were more unequally distributed than in the private messaging networks. This indicated that on the public part of the forum, a small number of highly active members were generating a disproportionate number of communications, more so than on the private part of the forum. For example, the Lorenz plot in panel c of Fig. 2 shows that roughly the top five percent of users on forum B were responsible for about three quarters of all in-degrees on the public replies network. Figure 4 depicts the distribution of local transitivity across nodes in each network, describing the tendency for a node's neighbours to be connected to each other. The presence of nodes with high local transitivity was indicative of densely linked subgroups within the overall network, which, in turn, may represent forum members interested in similar topics. Comparing the public and private communication networks, we found these clusters of densely interconnected nodes to be more prevalent in the public replies networks on both forums. These comparisons pertained to the overall distribution of degrees and local transitivity in the networks considered, irrespective of the position particular members take in these distributions. To examine the extent to which the same members took central positions across different layers, we turned to assessing the robustness of centrality measures across network layers.

Measures of centrality robustness. To assess centrality robustness across layers, several measures recommended by Borgatti et al. (2006) were assessed. The results are displayed in Table 4. For both the *Top 1* and the *Top 3* measure, the most central node/s in the public replies layer of forum A was/were not the same as the most central node/s in the private messaging layer (regarding both in- and outdegree) but was/were the same for the public replies layer and the flattened network. The *Top 10%* measure, on the other hand, showed that the most central node in the public replies layer of forum A was among the 10% most central nodes in both the private messaging layer and the flattened network. This was even more pronounced on forum B, where the *Top 1*, *Top 3*, and *Top 10%* measures all showed the same central nodes across layers. In addition to the Szymkiewicz-Simpson coefficient, we calculated the Jaccard's coefficient ($\frac{|A \cap B|}{|A \cup B|}$), which was 0.34 for in- and 0.26 for out-degrees (against a maximum of 0.53) on forum A, and even closer to its respective maximum on forum B (0.22 for in-, and 0.21 for out-degrees against a maximum of 0.24). For the public replies layer and the flattened network, Jaccard's coefficients equalled the maximum of 0.82 for both in- and out-degrees on forum A, while being a bit lower on forum B (0.94 for in- and 0.93 for out-degrees against a maximum of 0.99). Rank correlations for the weighted degrees (as recommended by Bródka et al. 2018, due to the high skewedness of the data) closely resembled those of the unweighted degrees for both forums and, similar to the other measures, showed that there was substantial overlap between the public replies and the flattened network, while this was less pronounced between the public replies and the

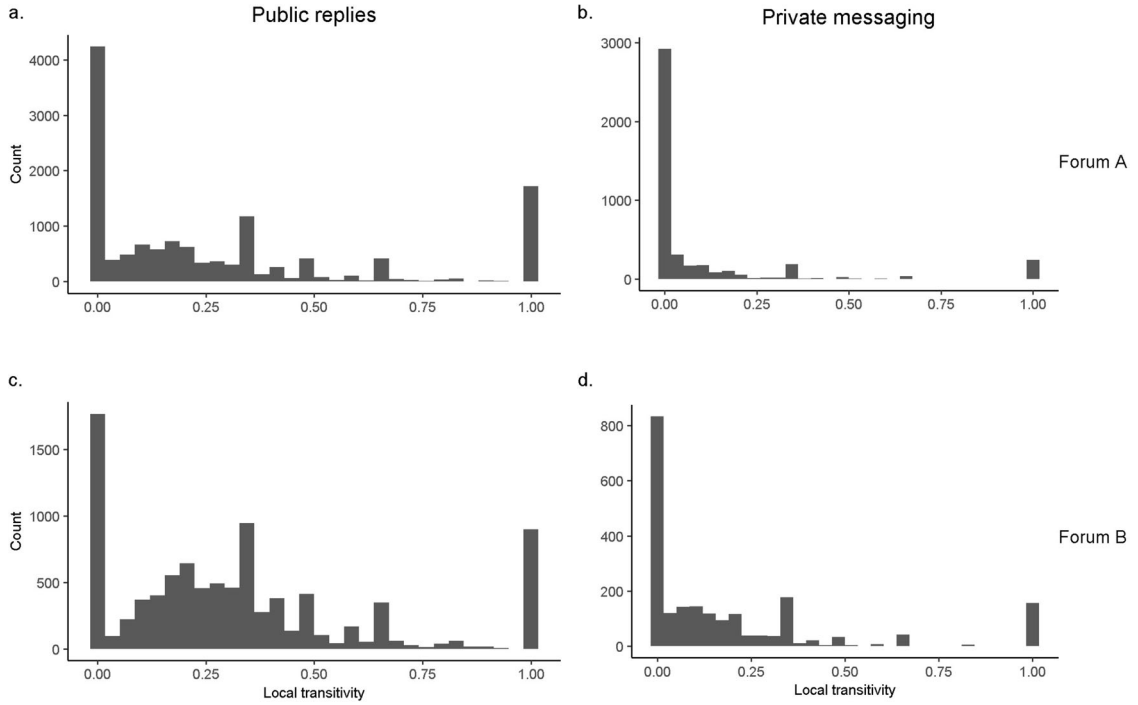

Fig. 4 Transitivity distributions of the multiplex network layers. Histograms of the transitivity index distributions for the public replies (**a**, **c**) and the private messaging (**b**, **d**) layer on forum A (**a**, **b**) and forum B (**c**, **d**). There are 9815 nodes with exactly one neighbour and therefore missing transitivity in the replies layer, and 7907 in the private messaging layer.

Table 4 Measures of centrality robustness.

Measure	Description	Comparison network	Forum A		Forum B	
			In-degree	Out-degree	In-degree	Out-degree
Top 1	Is the most central node in the public replies layer the same as the most central node in the private messaging layer/the flattened network?	Private messaging	No	No	Yes	Yes
		Flattened	Yes	Yes	Yes	Yes
Top 3	Is the most central node in the public replies layer in the top 3 of the most central nodes in the private messaging layer/the flattened network?	Private messaging	No	No	Yes	Yes
		Flattened	Yes	Yes	Yes	Yes
Top 10%	Is the most central node in the public replies layer among the 10% most central nodes in the private messaging layer/the flattened network?	Private messaging	Yes	Yes	Yes	Yes
		Flattened	Yes	Yes	Yes	Yes
Overlap (Szymkiewicz–Simpson coefficient)	What is the overlap between the 10% most central nodes in the public replies layer and the private messaging layer/the flattened network?	Private messaging	0.74	0.60	0.93	0.88
		Flattened	1	1	0.98	0.97
R	What is the correlation of the centralities between the public replies layer and the private messaging layer/flattened network (limited to actors who are in both layers)?	Private messaging	0.54	0.39	0.63	0.63
		Flattened	0.93	0.94	0.98	0.998

Table adapted from Borgatti et al. (2006). Szymkiewicz–Simpson coefficient $= \frac{|A \cap B|}{\min(|A|, |B|)}$.

private messaging network.[3] In addition, Table 5A, B shows the top five most central nodes for both in-degree and out-degree for each of the network layers. Across all layers, and considering in- as well as out-degree centrality, there were 10 distinct users among the five most central members of forum A, and 11 for forum B. In Table 5, these users are rank-ordered based on the total number of networks in which they were positioned among the five most central nodes.

On forum A, user 19406 was among the most central in all six networks considered, while user 32890 was central in four out of six networks. When it came to receiving messages, user 32890 was central only in the private messaging network, yet concerning sent messages this user was central in the public, private, and combined network. Table 5A shows that when considering in-degree, only two among the five most central users in the public replies network were also central to the private communications network, whereas for out-degree centrality this applied to three out of five users. While one of the administrators of the forum was among the topmost central across all networks (user 19406), the other administrator (user 1) was central in the private messaging network only. Of the 10 distinct users identified as most central, only one had no special role on the forum. These

Table 5 A Characteristics of the five most central actors across layers on forum A. B Characteristics of the five most central actors across layers on forum B.

ID	No. networks most central	In-degrees			Out-degrees			Roles				No. public posts	No. public posts with link	No. private messages	No. private messages with link
		Public replies	Private messaging	Flat-tened	Public replies	Private messaging	Flat-tened	Adminis-trator	Mode-rator	VIP	MVP				
A															
19406	6	1	1	1	1	1	1	1	1	1	1	4041	910	1646	232
32890	4	0	1	0	1	1	1	0	1	0	1	2181	935	1327	695
11615	3	1	1	1	0	0	0	0	0	0	0	312	146	274	31
6500	3	0	1	0	0	1	1	0	0	0	1	955	567	575	45
50229	3	0	0	0	1	1	1	0	1	0	1	2561	376	609	252
23406	2	1	0	1	0	0	0	0	0	0	1	3492	2835	283	117
13502	2	1	0	1	0	0	0	0	0	0	1	758	514	240	37
93933	2	1	0	1	1	0	0	0	0	0	1	1445	366	111	22
1	2	0	1	0	0	1	0	1	0	0	0	1118	161	2433	240
6306	2	0	0	0	1	0	1	0	1	0	0	3814	397	1776	223
												Geometric mean (*SD*)			
												2.31 (3.01)	2.13 (3.13)	1.91 (2.69)	1.76 (2.51)
												Median (*MAD*)			
												2.00 (1.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)

ID	No. networks most central	In-degrees			Out-degrees			Roles			No. public posts	No. public posts with links	No. private messages	No. private messages with link
		Public replies	Private messaging	Flat-tened	Public replies	Private messaging	Flat-tened	Adminis-trator	Mode-rator	"Reup doctor"				
B														
2	6	1	1	1	1	1	1	1	0	0	12,266	1844	2926	356
62260	2	0	1	0	0	1	0	0	0	0	230	27	757	96
85	2	0	1	0	0	1	0	0	1	0	6451	2625	1845	500
118502	2	0	1	0	0	1	0	0	1	0	1652	604	858	355
192	2	0	1	0	0	1	0	0	1	0	4459	2368	624	172
488	2	1	0	1	0	0	0	0	0	1	1870	442	179	37
54754	2	1	0	1	0	0	0	0	0	1	1243	787	39	10
678	2	1	0	1	0	0	0	0	0	0	816	343	86	35
8700	2	0	0	0	1	0	1	0	0	0	2640	794	214	47
72462	2	0	0	0	1	0	1	0	0	0	1567	241	56	12
19025	2	0	0	0	1	0	1	0	0	0	5746	1332	50	1
103	2	0	0	0	1	0	1	0	0	0	1423	89	108	11
3622	1	1	0	0	0	0	0	0	0	0	1778	1115	162	78
5222	1	0	0	1	0	0	0	0	0	0	2846	879	835	131
											Geometric mean (*SD*)			
											3.98 (4.92)	4.31 (6.14)	5.53 (4.30)	3.02 (3.40)
											Median (*MAD*)			
											2.00 (1.00)	2.00 (1.00)	4.00 (3.00)	2.00 (1.00)

Due to the skewed distributions, the last four rows show the geometric mean (excluding zeroes) and the median across all forum members who were active on the respective part of the forum, i.e. contributed at least one public post/reply (with link) or sent at least one private message (with link) for each particular column.

roles attribute certain special rights and/or responsibilities to selected members of the forum, and only roles that were explained on the forum are included here.

On forum B, only user 2 was central in more than two of the network layers, and this user was also the only one who was central in both the public replies and the private messaging layer when it came to both sending and receiving messages. The four users who are among the 11 most central users because of their centrality in the private messaging network were central for both sending and receiving private messages, while all users central in the public replies layer were central only for either their in- or their out-degree. Only six of the 11 most central users had a special user role, with only one of the two administrators being present among them. The two "re-up doctors" were both most central for the public replies they received, while the three moderators were only most central in the private messaging network.

The final four columns of Table 5A, B refer to the frequency and nature of members' forum communications. The most central users can be considered proliferate communicators, with posting frequencies high above the active users' average (shown in the last four rows of Table 5A, B). Still, posting frequency was not all-important, as was illustrated by users 6306 and 11615 on forum A. In terms of public communications, user 6306 trumped user 11615 by over factor ten. Yet, user 6306 was not among the most central when considering public replies received, whereas user 11615 was. This is because users were selected based on their unweighted (as opposed to weighted) degrees and therefore the numbers of sent replies and private messages might seem unintuitive at times. User 85 in Table 5B, for example, had contributed a high number of public posts, yet was not among the top 5 users with respect to in- or out-degree in the public replies network because these replies were distributed between fewer users than those of the users among the top 5.

Regarding the nature of the forum communications, Table 5A, B also show the number of public posts and private messages which contained a hyperlink. Previous research found that these hyperlinks are a good proxy for sharing CSAM (Blokland et al., 2024). User 23406 on forum A appeared to be sharing a disproportionately large amount of CSAM on the public part of the forum, potentially causing this user's high in-degree centrality, yet, judging by their out-degree centrality, this user did not seem very communicative towards others.

To better understand the relationship between communication and sharing of CSAM in the forums under scrutiny we computed Spearman rank correlations between in- and out-degrees and the number of hyperlinks posted in the communication networks. For forum A, the correlations on the replies network were of medium size (forum A: $r_{in,\ html} = 0.51$, $r_{out,\ html} = 0.29$). For forum B, the correlation value for in-degrees was similar, but higher for the correlation with out-degrees (forum B: $r_{in,\ html} = 0.56$, $r_{out,\ html} = 0.66$). Comparable values were obtained for the number of links shared in the private messaging network (forum A: $r_{in,\ html} = 0.25$, $r_{out,html} = 0.41$; forum B: $r_{in,html} = 0.56$, $r_{out,\ html} = 0.41$). Accordingly, the users identified to have the highest degrees in either the public or private part of the respective forum (Table 5A, B) only showed a moderate overlap with the users that shared the most links on each respective part of the forum. In forum A, three users of the ones with the highest degrees were also among the users who posted the most hyperlinks in their posts, and four of the most central users in the private part of the forum were among the ones that shared the most hyperlinks in their private messages. In forum B, on the other hand, none of the most central users on the public part of the forum were among the users who post the highest number of hyperlinks, while three of the most central users were among the five users who shared the most private messages with hyperlinks.

Figure 5A, B shows network plots of the 100 users with the highest total degrees in the respective flattened networks. The 100 users with the highest degrees were chosen because showing all users on the forums would have resulted in uninterpretable plots. All users were present in both respective public and the private messaging layers. Colours indicate whether users had special roles in the forum. Since some users had multiple roles, we interpreted the roles and their descriptions and ranked them from highest (administrator) to lowest (MVP on forum A; "reup doctor" on forum B). Accordingly, only each user's highest role is displayed in the plot. The 10 distinct users from forum A and the 11 from forum B that were present in the respective Top 5 lists across all layers are annotated with labels. It is noticeable from the plot that a large number of the displayed users had a special role in the forum, which was rare on both forums (only 0.3% of all users active on at least one of either the public replies or private messaging layer on forum A and 0.2% on forum B had one or more of these roles). Connections in the network plot show the summed number of all interactions between the users in the plot, i.e. replies in the public replies layer as well as private messages in the private messaging layer, with darker shades indicating more interactions. Because there were much more interactions between these users on forum B than on forum A, the shades indicate different numbers in the two plots. Again, for visibility reasons the interactions are shown as undirected edges. The plot for forum A shows that communication between users with the selected special roles appeared to be more interconnected and intense than that of the communication with users without special roles. For forum B, on the other hand, this was less pronounced.

Discussion

Using digital trace data from two large-scale darknet CSAM forums, the current study constructed distinct communication networks between forum members in which one layer represented members' public communications, and the other layer represented members' private messaging to assess the extent to which key-player identification is misjudged when only using public communication data. As in previous studies, findings showed that only a small minority of registered members were actively communicating on the forums (e.g. van der Bruggen and Blokland, 2022). Furthermore, the number of contacts per member was heavily skewed, indicating the presence of hubs or key-players. This again aligns with findings of prior studies describing darknet CSAM forum communications and identifying central users (Blokland et al., 2024; da Cunha et al., 2020; Fonhof et al., 2019). Comparisons of the public and private messaging networks, however, showed that not all communicating members were active in both layers, although the degree to which this was true varied across the forums. In forum B, only few users were active in the private messaging network and not in the public replies network. However, central players in the public network in either forum were not necessarily also hubs in the private messaging network. Likewise, key-players in the private messaging network were not always central to the public network. This is in line with previous research on comparing private and public messages, which has shown that, while there are similarities between both modes of communication, they are not the same (Afroz et al., 2013; Motoyama et al., 2011; Overdorf et al., 2018; Smirnova et al., 2024). As the private messaging network in both forums contained fewer members and fewer connections between members, combining both layers into a flattened network yielded results more similar to that of the public network alone.

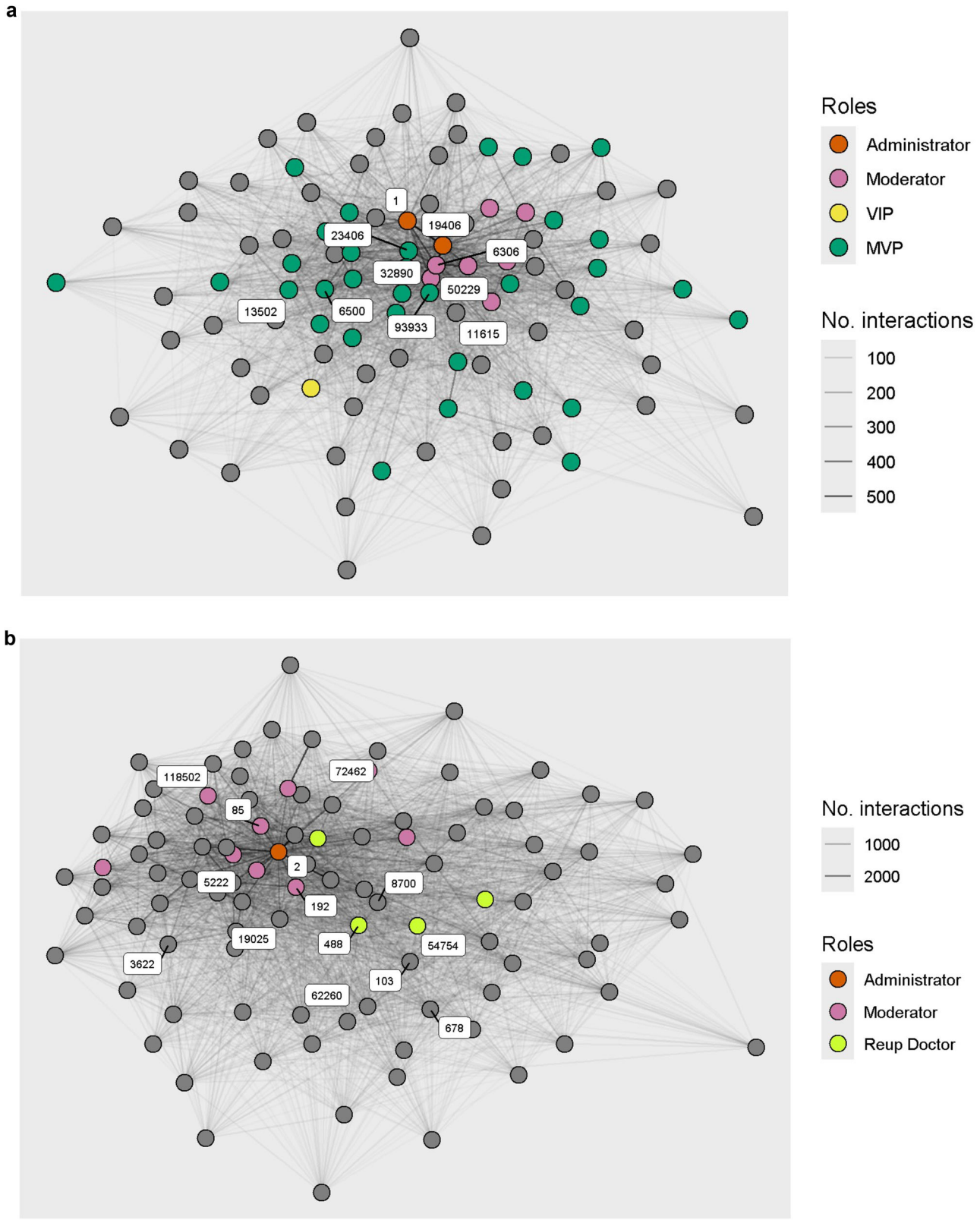

Fig. 5 Roles and interactions of the 100 users with the highest total degrees. a Roles and interactions of the 100 users with the highest total degrees on forum A. Network plot showing the 100 users with the highest total degrees in the flattened network of forum A, with colours representing their roles and opacity of their connections showing the number of interactions. **b** Roles and interactions of the 100 users with the highest total degrees on forum B. Network plot showing the 100 users with the highest total degrees in the flattened network of forum B, with colours representing their roles and opacity of their connections showing the number of interactions.

How do these results reflect on the main research question of to what extent key-player assessment on darknet CSAM forums is biased when relying on public communication data only? It seems that there are two possible answers to this question. One could argue that, given the number of members involved and the extent of their interactions, public communications are more important to the online community than private messages. Given that combining the public and private messaging layers into a flattened network yielded highly similar results when it comes to identifying key-players, the amount of bias introduced by relying solely on the public part of the forum seems limited. Targeting key-players in a forum's public replies network has been shown to significantly reduce the flow of CSAM through that forum (da Cunha et al., 2020), which is in accordance with Smirnova et al.'s (2024) assumption that the flow of information was similar within the private and the public part of the hacker forum they analysed. Additionally, publicly available user roles appeared to be a viable proxy for identifying central players in the flattened network on forum A, insofar that the overlap between users with high degrees and users with special roles was substantial. Even though this was less pronounced in forum B, it revealed another interesting dynamic, as all users with a special role, except the global administrator of the forum, were most central only in the private messaging network. That means that although user roles in forum B less reliably indicated a central position in the flattened network, they did so in that part of the forum that is usually not accessible to law enforcement agencies (i.e. the private messaging network) and could therefore provide additional valuable information for case prioritisation. In that sense, darknet CSAM forum users who have a special role can be said to reliably identify themselves as key-players in at least one communication layer in both researched forums. Preliminarily, law enforcement agencies can therefore be argued to be 'on the mark' when prioritising users central to public forum activity.

Contrarily, one could also argue that members who communicate privately might constitute a user subtype on their own, with their own dynamics and central figures—given that members active on the private part of forum A only partly overlapped with those active on the public part. Focussing only on the public part of the forum obscures the activities of this subgroup engaging in private interactions and would fail to identify a substantial part of those key to that particular subgroup. In prior analyses of underground forums, it has been similarly found that there was only partial overlap of users actively participating in the different modes of communication (Motoyama et al., 2011; Overdorf et al., 2018; Smirnova et al., 2024), leading to the assumption that private messages might be used for other operations than public ones (Afroz et al., 2013; Smirnova et al., 2024). This is less true for forum B, where only 230 users were active on the private part of the forum who were not also part of the public part of the forum. This shows that specific means of communications might not only serve different purposes (Afroz et al., 2013; Smirnova et al., 2024) but also develop differently across CSAM darknet forums; an assumption that may also be supported by the fact that one of the other forums considered for analysis used a special chat room instead of private messaging on the forum. Data on other aspects of member interactions, like the nature and novelty of the material shared, is needed to determine whether a separate focus on key-players in the private communication network is warranted (Westlake et al., 2011). To this end, future research should compare the content of material shared in public posts and in private messages.

The current analyses were based on directed networks derived from forum communications. On forum A we found that the distribution of degree centrality, i.e. the number of other members a user is connected to, is more skewed when considering in-degree than out-degree on both the public and private messaging network. In forum B, this pattern was reversed for private messages. For the public network this means that the original postings of some members are generating a disproportionate number of replies. While the distributions of replies sent were also skewed, they were yet more equally spread than those of replies received. It seems reasonable to assume that posts containing links to new, rare, and/or high-quality material elicit most reactions from the community, although this has yet to be studied empirically. High in-degree members, therefore, appear to be the most suitable targets for law enforcement prioritisation, while high out-degree members may play a more indirect role in the continuation of the community, creating a forum environment to which members are willing to contribute (Blokland et al., 2024). Thus, these members might play an instrumental role in sustaining the larger community of users across specific darknet CSAM forums. Future research will have to show if these users as well as members central to the private parts of the forums might be considered promising targets for law enforcement agencies. For forum A, the same pattern of less equally distributed out-degrees was found in the private communications network. For forum B, this pattern was again reversed, i.e. although both in- and out-degree distributions were very skewed (with more unique users trying to contact others than being contacted by private messages), in-degrees were more equally distributed than out-degrees. That might have been due to active users and users with administrative roles from the public replies layer of forum B using private messaging more often to reach out to other users they knew from the public part than on forum A.

In sum, the comparison of forums A and B speaks to the overall importance of relatively few key-players in the networks. While the skewed degree distributions of the two forums do not only resemble properties of generic internet forums but also each other, the current analysis also reveals a number of important differences between forums. The different styles of communication developing in different forums with otherwise similar forum architectures, for example, reveal a substantial amount of uniqueness.

The highly skewed distributions of the in- and out-degrees in the public and the private messaging networks might hint at network formation mechanisms at work on these forums. In the presence of skewed distributions, so-called preferential attachment is often proposed, stating that new users on the forum would tend to direct their communication disproportionately towards members that are already central (Albert and Barabási, 2002). Given that, to our experience, a lot of the communication in the network evolves around the praise of or requests for CSAM, this proposition makes intuitive sense and has already been suggested for other darknet forums (Me and Pesticcio, 2018). However, the analysis of longitudinal dynamics in the formation of the forums' communication networks would have been beyond the scope of this study and should be subject to future research.

Moreover, other sources of information besides users' positions in the social network structure might be considered important in the identification of key-players. For example, the content of the communication on the forum (publicly or privately) might highlight users with important social roles (L'Huillier et al., 2011). Basu and Sen (2021) even deem the monitoring of key-players solely based on their network measures an uneconomical practice. In fact, our results support that prioritisation utilising network degrees should be combined with other information to enable the identification of the most promising suspects, given that the overlap between top communicators and top sharers of CSAM was only moderate. Other authors have, for example, explored the combination of network measures with the severity of the material shared and find that it

might enhance key-player detection (Westlake et al., 2011). Future research into the content of the public and private messages and the CSAM exchanged, as well as identification strategies from other research areas, may provide more insights into finding users who are important for the structure of darknet CSAM forums. Still, social network measures are relatively easy to use as heuristics on large and diverse data sets, especially in comparison to more labour-intensive methods such as the manual annotation of post contents.

Given the size of the forums, and the breadth of genres and interests covered by its subforums – for instance in terms of victims' age and gender –, it also seems likely that some members will limit their interactions with other members to one or only a few subforums. While these members may not be among the most central when the total forum is considered, they might be key-players in subcommunities orbiting around specific sets of sexual or crime preferences. In the context of mainstream news forums, Forestier et al. (2012) refer to such forum members as contextual celebrities. Indeed, our current results on transitivity in both the public and the private messaging networks suggest the existence of densely connected subcommunities within the forum community as a whole. To the extent that these subcommunities focus on specific types of CSAM that can be considered especially harmful, despite not being most central to the overall community, contextual celebrities may also warrant law enforcement vigilance.

Limitations

Some limitations to the current study must be noted when considering its implications for law enforcement practice. First, given the context of the darknet, the information available on individual forum members is limited to their user handle. Hence, in the communication networks under scrutiny here, each user handle is considered a separate node. From law enforcement practice, we know that some individuals may at some point in time simultaneously operate under multiple user handles on the same forum. To the extent that this may have been the case on the forums analysed, we overestimate the number of nodes, and, to the extent that members interacted differently with other users depending on the user handle they had operated under, we underestimate members' network centrality.

Second, our analyses are limited to conversations taking place within single forums. As at any given moment multiple darknet CSAM forums are online, members who are not central in any particular forum may be key-players when activities across multiple forums are combined. Additionally, members may be in contact with each other outside of the forum environment, and continue their interactions on, for example, darknet chat sites or communication services such as Telegram. This too could lead to underestimating the centrality of individual members.

Third, although digital trace data provide a unique window at membership activity, we noticed that at least some posts had been deleted. Given that the reasons why and the exact number of posts deleted remain unknown, this introduces an unspecified level of inaccuracy to our analyses. Based on the information we do have, e.g. the number of replies to posts not in the data, we assume this level of inaccuracy to be low, however.

Lastly, but probably most importantly, our study describes only two large darknet CSAM forums. Despite these forums having thousands of members, since our units of analysis are forums, our sample size remains to be two. This sample size does not allow for any statistical tests and the generalisability of our findings to other CSAM darknet forums remains uncertain. However, this is a limitation that all current research studies dealing with (private) messages on CSAM darknet forums must acknowledge, as the number of seized forums a research team can gain access to is still limited.

Conclusion

The current study set out to assess the extent to which law enforcement agencies identifying key-players based on darknet CSAM forum members' public communications are missing the mark in terms of detecting those most central to the forum community. Given the rarity of full access to private communication, the results of this study offer the first analysis of different modes of communication between members of two darknet CSAM forums. Comparing the public and the private messaging networks from two large-scale darknet CSAM forums, we find notable differences in the size and structure of the public and private forum communication networks. The robustness of key-player identification across network layers can be considered moderate. As adding private communication data to those based on public posts and replies does not substantially alter users designated as central to the network, however, law enforcement practice of using public communications only to identify key-players does not seem to significantly suffer from intolerable bias. Yet, to the extent in which private messaging might include the exchange of new or extremely harmful CSAM, those central to the private messaging network only may still warrant (limited) law enforcement attention.

Received: 28 March 2024; Accepted: 15 October 2024;
Published online: 02 November 2024

Notes

1 During our analyses, we also considered the inclusion of data from two other seized darknet forums we had access to. However, an inspection of the private messages showed that we could not use these data because one of the forums had a separate chat room, completely replacing the function of the private messages, and the network created from the other forum was too small.

2 Median and median absolute deviation (*MAD*) values reported for all active users, i.e. those who sent/received at least one public message or sent/received at least one private message, respectively, once including users with zero values ($Md_{zeroes+}$ and $MAD_{zeroes+}$), and once excluding users with zero values ($Md_{zeroes-}$ and $MAD_{zeroes-}$).

3 Note, however, that only users present in both layers of the network were analysed via correlations. Including users present in only one layer and assigning them a degree value of zero would have resulted in lower correlation coefficients. The correlations reported here should therefore rather be considered an upper bound of this association.

References

Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74(1):47–97. https://doi.org/10.1103/RevModPhys.74.47

Afroz S, Garg V, McCoy D, Greenstadt R (2013) Honor among thieves: a common's analysis of cybercrime economies. In: 2013 APWG eCrime Researchers Summit, San Francisco, USA, 1–11. https://doi.org/10.1109/eCRS.2013.6805778

Basu K, Sen A (2021) Identifying individuals associated with organized criminal networks: a social network analysis. Soc Netw 64:42–54. https://doi.org/10.1016/j.socnet.2020.07.009

Bergman J, Popov OB (2023) Exploring dark web crawlers: a systematic literature review of dark web crawlers and their implementation. IEEE Access 11:35914–35933. https://doi.org/10.1109/ACCESS.2023.3255165

Boekhout HD, Blokland AAJ, Takes FW (2024) Early warning signals for predicting cryptomarket vendor success using dark net forum networks. Sci Rep 14(1):16336. https://doi.org/10.1038/s41598-024-67115-5

Blokland A, Daser A, de Boer M, Gannon C, Gnielka F, Huikuri S, Reichel R, Schäfer T, Schmidt AF, Staciwa K, Lehmann R (2024) Why do users continue to contribute to darknet CSAM forums? Examining social exchange, social capital, and social learning explanations using digital forensic artifacts. Child Abuse Negl 153(106815). https://doi.org/10.1016/j.chiabu.2024.106815

Borgatti SP, Carley KM, Krackhardt D (2006) On the robustness of centrality measures under conditions of imperfect data. Soc Netw 28(2):124–136. https://doi.org/10.1016/j.socnet.2005.05.001

Bródka P, Chmiel A, Magnani M, Ragozini G (2018) Quantifying layer similarity in multiplex networks: a systematic study. R Soc Open Sci 5(8):1–16. https://doi.org/10.1098/rsos.171747

Burcher M, Whelan C (2018) Social network analysis as a tool for criminal intelligence: understanding its potential from the perspectives of intelligence analysts. Trends Organ Crime. 21:278–294. https://doi.org/10.1007/s12117-017-9313-8

Contractor N, Monge P, Leonardi P (2011) Multidimensional networks and the dynamics of sociomateriality: bringing technology inside the network. Int J Commun 5:682–720

Csárdi G, Nepusz T, Traag V, Horvát S, Zanini F, Noom D, Müller K (2024) igraph: network analysis and visualization in R. https://doi.org/10.5281/zenodo.7682609

da Cunha BR, MacCarron P, Passold JF, dos Santos LW Jr, Oliveira KA, Gleeson JP (2020) Assessing police topological efficiency in a major sting operation on the dark web. Sci Rep 10(73). https://doi.org/10.1038/s41598-019-56704-4

Duijn PA, Klerks PP (2014) Social network analysis applied to criminal networks: recent developments in Dutch law enforcement. In: Masys AJ (ed) Networks and network analysis for defence and security. Springer International Publishing, p 121–159

Duxbury SW, Haynie DL (2021) Shining a light on the shadows: endogenous trade structure and the growth of an online illegal market. Am J Sociol 127(3):787–827. https://doi.org/10.1086/718197

Europol (2023) Internet organised crime threat assessment (IOCTA) 2023. https://www.europol.europa.eu/cms/sites/default/files/documents/IOCTA%202023%20-%20EN_0.pdf. Accessed 27 Mar 2023

Fonhof AM, van der Bruggen M, Takes FW (2019) Characterizing key-players in child exploitation networks on the dark net. In: Aiello LM, Cherifi C, Cherifi H, Lambiotte R, Lió P, Rocha LM (eds) Complex networks and their applications VII: volume 2 proceedings the 7th international conference on complex networks and their applications COMPLEX NETWORKS 2018. Springer International Publishing, p 412–423

Forestier M, Velcin J, Zighed DA (2012) Analyzing social roles using enriched social network on on-line sub-communities. In: ICDS 2012: The sixth international conference on digital society, 17–22

Fortin F, Paquette S, Gagné S (2021) Challenges and opportunities in investigations of online sexual exploitation of children: old networks, dark web, and proactive response. In: Deslauriers-Varin N, Bennell C (eds) Criminal investigations of sexual offenses: techniques and challenges. Springer International Publishing, p. 217–233. https://doi.org/10.1007/978-3-030-79968-7

Gaito S, Quadri C, Rossi GP, Zignani M (2017) Urban communications and social interactions through the lens of mobile phone data. Online Soc Netw Media, 1:70–81. https://doi.org/10.1016/j.osnem.2017.04.003

Ganley D, Lampe C (2009) The ties that bind: social network principles in online communities. Decis Support Syst 47(3):266–274. https://doi.org/10.1016/j.dss.2009.02.013

Gannon C, Blokland AA, Huikuri S, Babchishin KM, Lehmann RJ (2023) Child sexual abuse material on the darknet. Forens Psychiatr Psychol Kriminologie 17:353–365. https://doi.org/10.1007/s11757-023-00790-8

Goonetilleke P, Knorre A, Kuriksha A (2023) Hydra: lessons from the world's largest darknet market. Criminol Public Policy 22(4):735–777. https://doi.org/10.1111/1745-9133.12647

Halu A, Mondragón RJ, Panzarasa P, Bianconi G (2013) Multiplex PageRank. PLoS ONE 8(10):1–10. https://doi.org/10.1371/journal.pone.0078293

Huang C, Guo Y, Guo W, Li Y (2021) HackerRank: identifying key hackers in underground forums. Int J Distrib Sens Netw 17(5):15501477211015145. https://doi.org/10.1177/15501477211015145

IWF (2023) #BehindTheScreens. A deep dive into the digital and social emergency happening #BehindTheScreens, in children's bedrooms. The annual report 2022. https://annualreport2022.iwf.org.uk/wp-content/uploads/2023/04/IWF-Annual-Report-2022_FINAL.pdf. Accessed 10 Dec 2023

Jiang C, Foye J, Broadhurst R, Ball M (2021) Illicit firearms and other weapons on darknet markets. Trends Issues Crime Crim Justice 622:1–20

Kigerl A (2022) Behind the scenes of the underworld: hierarchical clustering of two leaked carding forum databases. Soc Sci Comput Rev 40(3):618–640. https://doi.org/10.1177/0894439320924735

L'Huillier G, Alvarez H, Ríos S, Aguilera F (2011) Topic-based social network analysis for virtual communities of interests in the dark web. ACM SIGKDD Explor Newsl 12:66–73. https://doi.org/10.1145/1964897.1964917

McLevey J, Scott J, Carrington PJ (eds) (2023) The Sage handbook of social network analysis. SAGE Publications Limited

Magnani M, Rossi L (2011) The ML-Model for multi-layer social networks. In: 2011 International Conference on Advances in Social Networks Analysis and Mining, 5–12. https://doi.org/10.1109/ASONAM.2011.114

Magnani M, Rossi L, Vega D (2021) Analysis of multiplex social networks with R. J Stat Softw 98:1–30. https://doi.org/10.18637/jss.v098.i08

Me G, Pesticcio L (2018) Tor black markets: economics, characterization and investigation technique. In: Jahankhani H (ed) Cyber Criminology. Springer, 119–140

Morselli C (2010) Assessing vulnerable and strategic positions in a criminal network. J Contemp Crim Justice 26(4):382–392. https://doi.org/10.1177/1043986210377105

Motoyama M, McCoy D, Levchenko K, Savage S, Voelker GM (2011) An analysis of underground forums. In: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference, 71–80. https://doi.org/10.1145/2068816.2068824

Overdorf R, Troncoso C, Greenstadt R, McCoy D (2018) Under the underground: Predicting private interactions in underground forums. https://doi.org/10.48550/ARXIV.1805.04494

Pete I, Hughes J, Chua YT, Bada M (2020) A social network analysis and comparison of six dark web forums. 2020 IEEE European symposium on security and privacy workshops, Genoa, Italy, 484–493. https://doi.org/10.1109/EuroSPW51379.2020.00071

R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Smirnova O, Hyslip TS, Holt TJ (2024) Are active users the most central to hacker social networks? A comparative analysis of public and private online network structures among hackers. Deviant Behav, 1–17. https://doi.org/10.1080/01639625.2024.2373356

Soudijn MR, Zegers BCT (2012) Cybercrime and virtual offender convergence settings. Trends Organ Crime. 15(2-3):111–129. https://doi.org/10.1007/s12117-012-9159-z

Sun N, Rau PPL, Ma L (2014) Understanding lurkers in online communities: a literature review. Comput Hum Behav 38:110–117. https://doi.org/10.1016/j.chb.2014.05.022

Sun Z, Rubio-Medrano CE, Zhao Z, Bao T, Doupé A, Ahn GJ (2019) Understanding and predicting private interactions in underground forums. In: Proceedings of the ninth ACM conference on data and application security and privacy, 303–314. https://doi.org/10.1145/3292006.3300036

van der Bruggen M, Blokland A (2022) Profiling darkweb child sexual exploitation material forum members using longitudinal posting history data. Soc Sci Comput Rev 40(4):865–891. https://doi.org/10.1177/0894439321994894

Westlake BG, Bouchard M (2015) Criminal careers in cyberspace: examining website failure within child exploitation networks. Justice Q 33(7):1154–1181. https://doi.org/10.1080/07418825.2015.1046393

Westlake BG, Bouchard M (2016) Liking and hyperlinking: community detection in online child sexual exploitation networks. Soc Sci Res 59:23–36. https://doi.org/10.1016/j.ssresearch.2016.04.010

Westlake BG, Bouchard M, Frank R (2011) Finding the key-players in online child exploitation networks. Policy Internet 3(2):104–135. https://doi.org/10.2202/1944-2866.1126

Westlake BG, Frank R (2017) Seeing the forest through the trees: identifying key players in online child sexual exploitation distribution networks. In: Holt T (ed) Cybercrime through an interdisciplinary lens. Routledge, 189–209

Yip M, Shadbolt N, Webber C (2013) Why forums? An empirical analysis into the facilitating factors of carding forums. In: Proceedings of the 5th annual ACM web science conference, Paris, France, 453–462. https://doi.org/10.1145/2464464.2464524

Yip M, Webber C, Shadbolt N (2013) Trust among cybercriminals? Carding forums, uncertainty and implications for policing. Policy Soc 23(4):516–539. https://doi.org/10.1080/10439463.2013.780227

Zamani M, Rabbani F, Horicsányi A, Zafeiris A, Vicsek T (2019) Differences in structure and dynamics of networks retrieved from dark and public web forums. Phys A 525:326–336. https://doi.org/10.1016/j.physa.2019.03.048

Author contributions

Frederic M. Gnielka, Rebecca Reichel: These authors contributed equally to this work. Alexander F. Schmidt, Thomas Schäfer, Salla Huikuri, Katarzyna Staciwa, Robert Lehmann: These authors jointly supervised this work.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

This research was funded by the European Union under Grant ISF-2021-TF1-AG-CYBER.

Ethical approval

This study was conducted retrospectively on pre-anonymised data. Therefore, ethical approval was not required. All research was performed in accordance with the ethical standards as laid down by the 1964 Declaration of Helsinki.

Informed consent

This study was conducted retrospectively on pre-anonymised data. Therefore, informed consent was not required.

Additional information

Correspondence and requests for materials should be addressed to Alexander F. Schmidt.