International Journal of Networking and Computing – www.ijnc.org, ISSN 2185-2847
Volume 11, Number 2, pages 354–382, July 2021

Graph visualization of dark web hyperlinks and their feature analysis

Taichi Aoki
Graduate School of Information Security, Institute of Information Security
2-14-1 Tsuruya-cho, Kanagawa-ku, Yokohama, Kanagawa 211-0835, Japan
dgs148101@iisec.ac.jp

Atsuhiro Goto
Graduate School of Information Security, Institute of Information Security
2-14-1 Tsuruya-cho, Kanagawa-ku, Yokohama, Kanagawa 211-0835, Japan
goto@iisec.ac.jp

Received: February 14, 2021
Revised: April 23, 2021
Accepted: June 2, 2021
Communicated by Toru Nakanishi

Abstract

Content regarding various illegal activities, such as weapon and drug trafficking, is shared on the dark web. Most of the illegal content is distributed on anonymous networks that cannot be directly accessed from the World Wide Web. A number of studies have been conducted on the network structure of the World Wide Web since its advent. Similar to the World Wide Web, the dark web is connected by hypertext transfer protocol (http); this makes it possible to use the methods developed for the web in the dark web. Many studies have investigated the dark web and its network structure. However, few studies have focused on the visualization of the dark web network structure, and there have been no studies investigating the temporal changes in the network structure.

In this study, to understand the hypertext markup language (html) network structure of the dark web, we created and visualized a graph of the html hyperlink relations of the Tor network, which is popular on the dark web. We then compared the insights gained from graph centrality metrics with those gained from visualizations. The analyzed dataset comprised 25,270,157 pages of html text files crawled from the Tor network by breadth-first search from June 1, 2018, to January 30, 2021. Subsequently, we acquired half-yearly snapshots from the collected data and investigated the change in the dark web network over time using a time-series graph. Then, we derived the centrality metrics from the created graph data and confirmed the differences between the centrality metrics and visualizations. The results obtained in this study provided new insights into the dark web. First, we found that the dark web fluctuated significantly; the structure of the dark web network was more strongly interconnected. Second, most of the nodes that had increased in the past two years may have disappeared rapidly after May 2020. Third, analysis of each snapshot revealed that the proportion of highly volatile domains increased from 40% to 75% during the observation period. Fourth, after calculating the network centrality metrics from each snapshot and comparing the transition of hub nodes in chronological order, we observed that the importance of link-collection sites as the main information retrieval method used in the dark web decreased. Finally, we estimated the size of the dark web based on our observed dark web measurements using the mark-recapture method. To the best of our

knowledge, this is the first study to use the mark-recapture method to estimate the size of the dark web network.

Keywords: Dark web, Tor, Crawling, Graph, Degree centrality, Mark recapture

1 Introduction

Content regarding various illegal activities, such as weapon and drug trafficking, is distributed on the dark web. These illegal activities are kept anonymous by highly developed anonymity software such as Tor [1] and I2P [2] network layers. Although anonymous networks cannot be accessed directly from the World Wide Web, they can be accessed using dedicated software. A proper understanding of the dark web can be helpful for various strategies in cybersecurity. For example, identifying the trends of threats on the dark web may help us to identify emerging threats against which cybersecurity needs to provide protective measures. Recognizing the state of the dark web network may help us in predicting its level of prosperity; it may also help us to understand the potential threats that are concentrated in the dark web, which we must be aware of. Further, observing the growth process of dark web networks may allow us to predict the inflow and outflow of illegal actors into networks other than the dark web.

During the period of this study, significant events occurred on the dark web. On April 30, 2019, Dream Market, which was the largest dark market of its time, was shut down [3]. On April 6, 2019, Gangsta's Paradise Forum, a large forum on the dark web, run by the admin Dread, was shut down [4]. On May 2, 2019, the German police took down the Wallstreet Market, a dark market [5]. On May 3, 2019, Europol took down the dark market called Valhalla Market [6]. On May 7, 2019, an international law enforcement agency took down DeepDotWeb, a dark web news site [7]. On May 9, 2019, DarkWebNews, a dark web news site, was shut down [8]. On May 22, 2019, the Dutch police took down Bestmixer.io, a cryptocurrency mixing site [9]. Did these actions succeed in making the dark web smaller?

This study aims to better understand the changes in the dark web network by visualizing them. To our knowledge, the state of the dark web is not easy to understand because the dark web network is huge, complex, and dynamic. Our visualizations help us to intuitively understand and gain insights into how large and complex networks have changed. In this paper, we present a comparison of the insights gained from graph centrality metrics and those gained from visualizations. We created a hyperlink topology of the dark web by crawling html from the popular Tor network. Our data were obtained by parsing a 25,270,157-page html text file crawled from the Tor network by breadth-first search from June 1, 2018, to January 30, 2021, and using the hyperlinks extracted from it as a data source. The crawled data collected during the period were divided into half-year periods and aggregated to create five topological graphs for comparison. This allowed us to visualize how the dark web had grown and evolved over the observed period. The results obtained in this study provided new insights into the dark web. Figure 1 and Figure 20 show a graph visualization of our crawled data for the entire period and an enlarged view of it, respectively. The size of the dark web was estimated using the mark-recapture method. To the best of our knowledge, this is the first study to use the mark-recapture method to estimate the size of the dark web network.

In Section 2, we show the differences between related work and our work and in Section 3, the dataset used in this analysis is described. We describe our analytical method in Section 4. In Section 5, we present the results. Here, we compare the results of the visualization of each snapshot with those of the centrality metrics. We discuss some of the details revealed by the analysis in Section 6. In Section 7, we summarize the results of our work. In the appendix, we present the top 10 hub nodes of the centrality metrics obtained from the analysis and an enlarged view of the visualization results.

2 Related Work

A number of studies have been conducted on the network configuration of the World Wide Web from the advent of the web up to the year 2000 [10, 11, 12, 13]. There have also been several works dealing

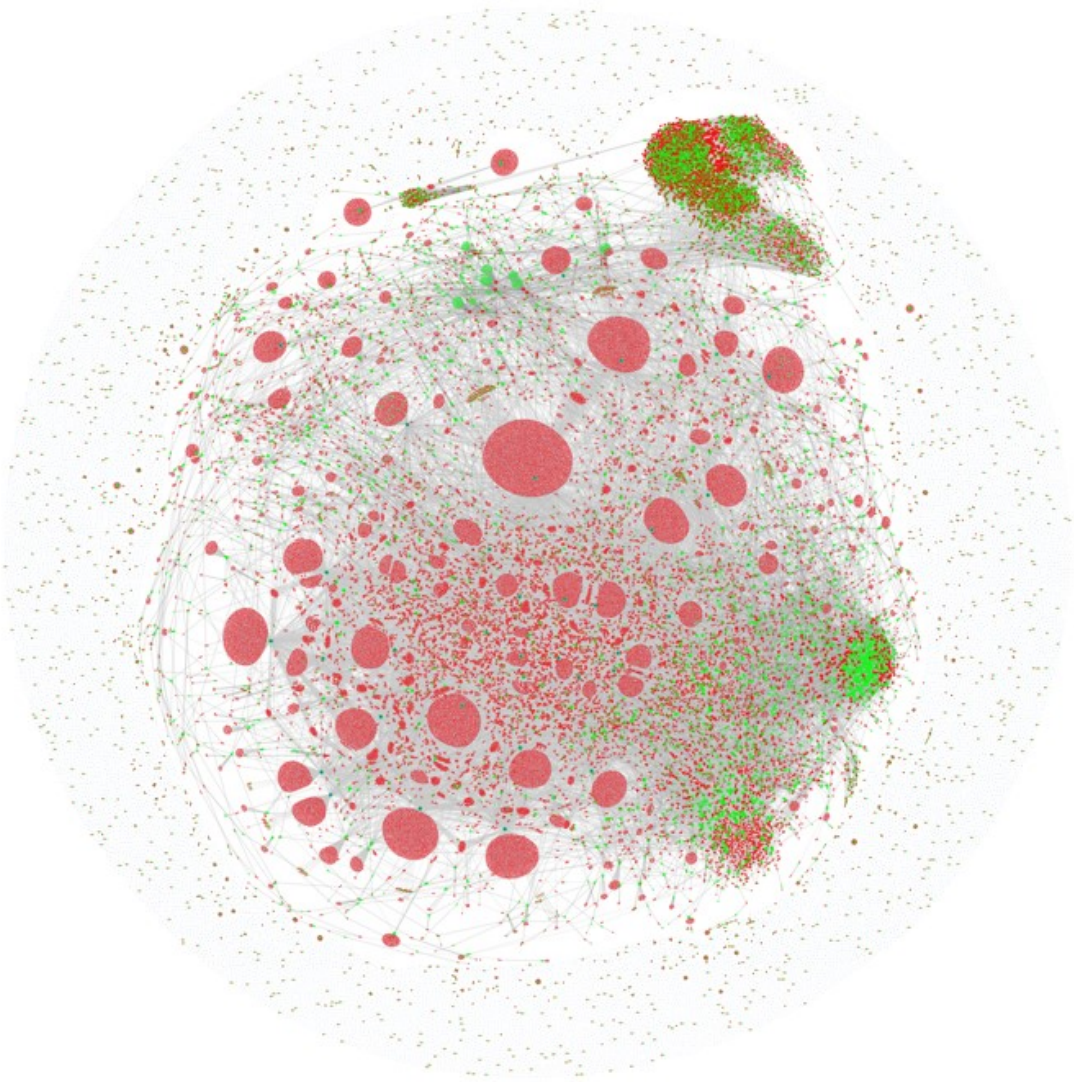

Figure 1: View of the dark web hyperlink network graph from June 1, 2018, to January 30, 2021.

Table 1: Related work and positioning of our study.

Works	Target	Dataset size	Relationship	Visualization	Time series change
Broder et al. [12]	WWW	1,000,000,000	Hyperlink	Conceptual diagram	No
Toyoda et al. [11]	WWW	2,830,000,000	Hyperlink	Graph	Yes
Heer et al. [17]	WWW	-	Social network	Graph	No
Fu et al. [17]	Tor	650	Social network	Graph	No
Hyperion et al. [25]	Tor	3,700	Contents structure	Graph	No
Domenico et al. [17]	Tor	5,535	Hyperlink	Graph	No
Griffith et al. [16]	Tor	13,117	Hyperlink	Chart	No
Brunner et al. [15]	Tor	34,714	Hyperlink	Graph	No
Cilleruelo et al. [26]	Tor and I2P	46,562	Hyperlink	Graph	No
Our study	Tor	185,989	Hyperlink	Graph	Yes
Yang et al. [16]	Tor	-	Social network	Graph	No
Sallaberry et al. [10]	Information search	-	Search query	Graph	No
Park et al. [19]	Dark web	-	-	Chart	No
Zhou et al. [18]	Dark web	-	Social network	Graph	No

with the visualization of the World Wide Web through network graphs [14]. Similar to the World Wide Web, the dark web is connected via http, which makes it possible to use the methods that apply to the web in this network. The dark web has been extensively researched. Popular research themes are the analysis of social and forum networks involved in cybercrime and analyses using the graph structure of the dark web. However, not many studies have focused on the visualization of the dark web network structure [15, 16, 17, 18, 19, 20, 21, 22, 23, 24].

Broder et al. [12] created a graph from 1 billion nodes collected from the World Wide Web and investigated its structure. They developed methods for graph analysis of the web. Toyoda et al. [11] developed a method to visualize the link structure of the World Wide Web using a time-series graph. Heer et al. [23] visualized the structure of the social networks graph. Moreover, they devised methods to combine pictures with graphs. However, none of these studies visualized the dark web. Brunner et al. [15] analyzed the network structure of the dark web in detail. They found 34,714 Tor domains from 67,296,302 links and analyzed the data collected by crawling approximately 10,000 running sites. In their research, 10% of the data they collected was visualized as the network structure of the dark web. From November 2016 to February 2017, Griffith et al. [15] investigated the network structure of the dark web, crawled 13,117 Tor domains, and found 7,178 active Tor domains. However, they did not visualize the graph structure. Domenico et al. [17] investigated the network structure of the dark web using three datasets for December 2013, May 2014, and January 2015 and analyzed 5,921, 4,953, and 5,535 nodes, respectively. Their study visualized the 5,535 nodes from the January 2015 dataset. Fu et al. [18] visualized the structure of forums collected from the dark web as a graph. However, they did not cover the entire dark web network. Sallaberry et al. [10] worked on a graph visualization for information retrieval on the dark web. They visualized the relationship between search queries and page graph. However, they did not visualize the structure of hyperlinks on the dark web. Park et al. [19] worked on the visualization of crime information on the dark web. However, they did not visualize the graph structure. Yang et al. [22] created and visualized a graph from the link relations of weblogs collected from the dark web. However, they did not cover the entire dark web network. Zhou et al. [24] analyzed the link relations collected from the dark web to identify terrorist groups and performed graph visualization of the terrorist communities. However, their research did not cover the Tor network. Hyperion Gray [25] presented an interesting visualization of the Tor network. The team created a visualization of the 3.7 K Tor onion service in 2019. Their visualization involved mapping a screen capture of a dark website on their graph structure. The graph in their work was related to the similarity of the site structure and was not a hyperlink-related graph. Cilleruelo et al. [26] created a graph structure of the dark web network to clarify the connection between Tor and the I2P network. They analyzed data from the Ahmia [27] and DUTA-10K datasets [28], dark web crawling, and a total of 46 K addresses obtained from hidden service descriptors. Although their work does not involve visualization, it is useful for comparison with our work.

After reviewing the above-mentioned studies, it was found that there is no research investigating the temporal change in the network structure of the dark web. We also did not find any study that used the mark-recapture method to estimate the dark web network size from observational data. We summarized the existing research in Table 1 to position our research and facilitate easy comparison.

3 Dataset

Our data were gathered by crawlers running a breadth-first algorithm known to be efficient in crawling the World Wide Web. Our crawling strategies were similar to the approach used by Brunner et al. [15]. However, we selectively crawled only text to improve network efficiency and to avoid the unnecessary collection of binary data. Our crawler was given some Tor pages as seeds; it then started collecting URLs with a top-level domain that ended with .onion. The hyperlink of that Tor page was extracted and pooled as the next crawling target. The pooled addresses for crawling were randomly selected. To prevent the concentration of access to a specific domain, each request was allowed to access the same domain with an interval of 10 s. To maintain the freshness of the collected data, a crawling target address that was not accessed within 14 days was deleted from the pool. This collection method was repeated from June 1, 2018, to January 30, 2021. The total number of pages collected during the entire period was 25,270,157, with 3,763,356 unique pages, 185,989 unique hosts, and 172,740 unique .onion top-level domains. The number of unique pages and unique hosts collected on the Tor network, summarized in 30-day intervals, and the total number of crawling pages are shown in Figure 2.

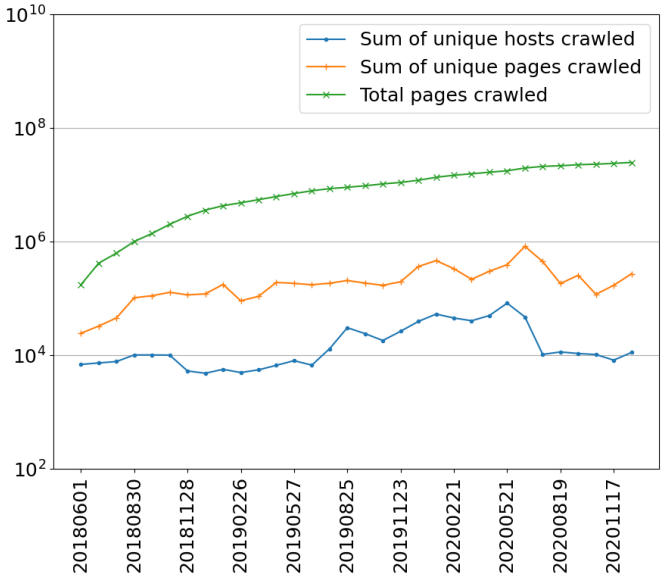

Figure 2: Plot of the total number of crawled pages (green) during our observation period, sum of unique pages per 30 days (orange), and sum of unique hosts per 30 days (blue).

4 Method

Our dataset collected by crawling was used to visualize the temporal changes in the dark web. The following sections explain the procedures used to create a graph from a dataset and visualize it.

4.1 Creating snapshots

The 2-year dataset was divided into six snapshots consisting of 180 days of crawled data. The dataset was divided according to the periods presented in Table 2 to create snapshots. The last snapshot, F, contained less than 180 days of data and was excluded from the comparison. Next, we extracted only the .onion links from the html files included in each snapshot using regular expressions.

Table 2: Number of unique domains per crawl dataset.

Label	Crawling period	Number of unique domains within the period
A	June 1, 2018 to November 27, 2018	48,360
B	November 28, 2018 to May 26, 2019	53,300
C	May 27, 2019 to November 22, 2019	70,684
D	November 23, 2019 to May 20, 2020	103,674
E	May 21, 2020 to November 16, 2020	54,509
F	November 17, 2020 to January 30, 2021	17,464
Total	June 1, 2018 to January 30, 2021	172,740

4.2 Visualization of snapshot differences

To check the domain differences between the snapshots, diff was applied to the domain text data pertaining to the snapshots, and the number of lost, new, and matched domains was counted (Table 3). Diff is a text comparison utility that computes and displays the differences between the contents of the texts. Next, to investigate the bias of the increasing and decreasing domains, the diff results were visualized using a method similar to the tools of WinMerge [29]. When comparing one snapshot with the next, a new entry was shown in red, entries present in both were shown in white, and entries that were no longer present were shown in gray (Figure 6). The vertical length of the two snapshots being compared was determined by the number of unique domains contained in each snapshot. A difference in length indicated a difference in the total number of unique domains in the two snapshots being compared.

4.3 Creating a graph

Similar to Toyoda et al.'s approach [11], a time series graph was created from the snapshots by using the following formula:

$$\{G_t := (V_t, E_t) | 1 \leq t \leq T\}, \tag{1}$$

where V_t is a domain node, E_t is the node relationship of V_t, and E_t applies relationship data that describe the hyperlink relationship between the domains, which are represented by table data with recorded input and output nodes. For visualization, five graphs, G, were created from snapshots A to E, that is, $T = 5$. To reduce the number of nodes in the graph, the domains were extracted from the URL, and graph data were created to show the hyperlink relationship between the domains. All the nodes contained in each snapshot were included in each graph.

4.4 Graph visualization

The graph data were visualized by applying a dynamic model. For visualization, D3 force-directed graph layout [30] was used. D3 is a library for producing data visualizations in web browsers.

Then, we adjusted the color of the graph so that the direction of the edge was represented visually. The source was marked in green, and the destination was marked in red (Figure 3). We set the spring model, Coulomb force, gravity, and collision detection as the D3 parameters. We used all the nodes in the snapshot and 5% randomly sampled edges for the automatic graph layout using the dynamic model.

4.5 Calculation of graph centrality metrics

We calculated the graph centrality metrics of each graph for degree centrality. The parameters of centrality were used to identify the important nodes in the network model. Degree centrality is the number of edges connected to the node. Because the edges are directional, the degree centrality can be calculated separately for the input and output. Therefore, in-degree is the number of ties that are

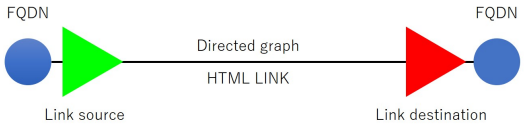

Figure 3: Graph legend and direction. Connect the two fully qualified domain names (FQDNs) with the green link source and red link destination arrows to create a graph structure.

directed to a node, and out-degree is the number of ties on a node that is directed to other nodes. We calculated the degree centrality, out-degree centrality, in-degree centrality, and PageRank [31] using NetworkX [32], which is a framework for network analysis. We obtained quantitative insights into the key nodes in the graph from the top 10 nodes for each centrality metric. The degree centrality C_d of a vertex v in the graph $G := (V, E)$ of the number of nodes n is defined as follows:

$$C_d(v) = \frac{deg(v)}{n-1}. \tag{2}$$

4.6 Extraction of the top 10 hub nodes

For each centrality metric, we extracted the top 10 domains with the highest score from each snapshot. The top 10 results of each centrality metric for each graph are presented in the tables included in the appendix.

4.7 Estimation of network node size

We estimated the Tor network size from the measured addresses. We used Peterson's mark-recapture method in population ecology to estimate the network node size. Crawling on a computer network is very similar to the collection of plants and animals in real life. Therefore, we can estimate the number of nodes in a network in the same manner that the number of plants and animals in real life can be estimated. The Petersen-type mark and recapture method [33] can be calculated using the following formula:

$$N = \frac{MC}{R}, \tag{3}$$

where M is the number of unique domains collected the first time, C is the number of unique domains collected the second time, R is the number of unique domains that overlap when comparing M and C, and N is the estimated total number of domains.

The Petersen model is a basic method for estimating the number of individuals that make up a population and is easy to calculate and handle. The requirements of this model are that it should be applied to a closed system and that the captured entities are well mixed before recapturing. When applying this model to network crawling, capture-target agitation can be achieved easily by uniformly and randomly selecting the crawl targets from the already collected target list. The conditions of the closed system can be approximated by shortening the recapturing cycle. We calculated this with a sampling width of 30 days of crawled data, and the marker was identified by the domain name. We aggregated the domains observed in the dataset and calculated the number of unique domains present in both the first capture and second capture sampling bins.

5 Result

5.1 Node size of each snapshot

We compared the node sizes of the five snapshots A-E (snapshot F was excluded). In snapshot A, the number of unique domains was 48,360. This increased steadily from snapshot B to D, where the

highest number of unique domains was recorded at 103,674. In snapshot E, it dropped significantly to 54,509 domains (Figure 4). This value is similar to the number of domains in snapshot B at 53,300 domains. This indicated that the growth in the number of domains over approximately one year disappeared rapidly within 180 days.

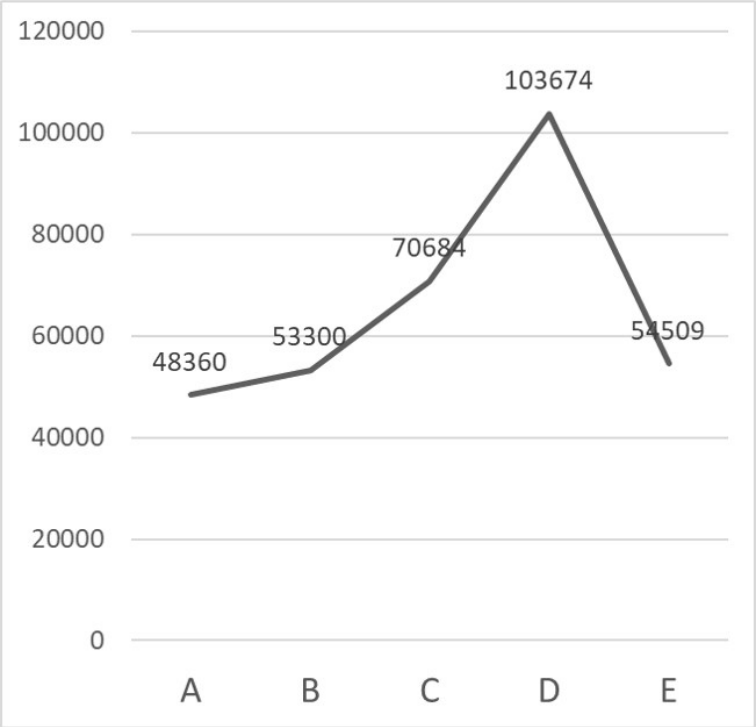

Figure 4: Number of unique domains in each snapshot.

5.2 Node volatility of each snapshot

For snapshots A-E, we compared each snapshot with the following snapshot in chronological order and counted the unique domains that disappeared, appeared, and existed in both (Table 3). Although the number of new domains increased from 14,966 to 54,747 from the comparison of snapshots A and B to the comparison of C and D, the number of matching domains remained relatively constant at 38,334 and 32,116 in the comparisons of A-B and D-E, respectively (Figure 5). The average number of matching domains over the total period was 39,398. The number of lost domains increased significantly from snapshot D to E. Further, the number of new domains decreased. Although the number of matching domains decreased, the rate of change was small when compared with the other two.

Next, we calculated the rate of domain change between two snapshots. For snapshots A-D, the average number of unique domains that disappeared was 17% of the number of unique domains that existed between the two snapshots being compared. However, in the comparison of snapshots D-E, the percentage of unique domains that disappeared was 57%. The percentage of unique domains that matched between snapshots dropped from 60% to 25% from snapshots A-B to D-E.

5.3 Node differences between snapshots

To investigate the bias in the increasing and decreasing domains, we visualized the differences between snapshots using the diff tool (Figure 6). A bias in either domain creation or disappearance is indicated in red or gray in the bar, respectively. In the comparisons of snapshots from A to D,

Table 3: Number of disappeared domains, new domains, and matching domains per snapshot comparison.

Label	Comparison snapshots	Number of disappeared domains	Number of appeared domains	Number of matched domains
Compare-AB	A and B	10,026	14,966	38,334
Compare-BC	B and C	15,084	32,468	38,216
Compare-CD	C and D	21,757	54,747	48,927
Compare-DE	D and E	71,558	22,393	32,116

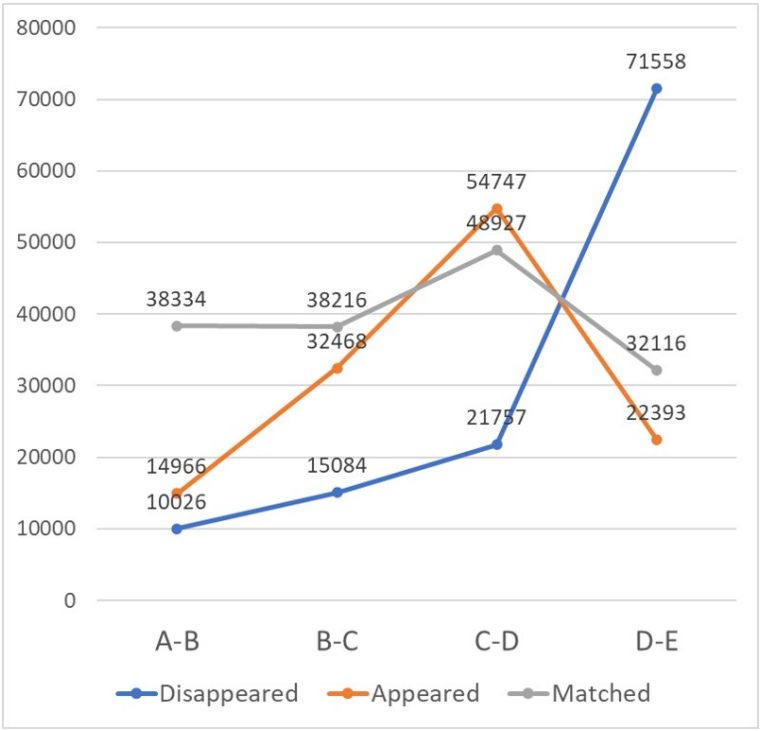

Figure 5: Number of disappeared domains, new domains, and matching domains per snapshot.

the number of new domains always exceeded the number of disappeared domains; therefore, the right bar of these comparison pairs were predominately red. In the comparison of snapshots D and E, the number of disappeared domains was higher; therefore, the gray color is more prominent in the right bar. Bias occurred when a large number of domains with similar domain names were intentionally created [34] on the Tor network. From the diff results, it was confirmed that there was no significant bias in domain generation and disappearance and the spectral results were generally evenly distributed.

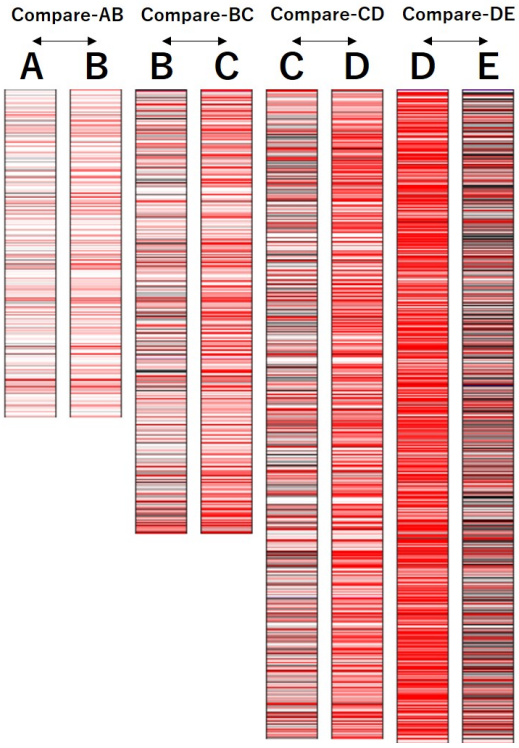

Figure 6: Diff visualization of disappeared and new domains between two snapshots. A new entry is shown in red, entries present in both are shown in white, and entries that are no longer present are shown in gray in each comparison. A difference in length indicates a difference in the total number of unique domains in the two snapshots being compared.

5.4 Estimated number of nodes

The number of unique hosts collected during our sampling period remained within the 10^4 range based on the aggregation of the 30-day sampling bins (Figure 2). Based on the number of unique domains observed in each sampling bin, the estimated number of domains was calculated using the mark-recapture method (Figure 12). Estimates showed values from 14,509 to 96,034. The average for the entire dataset was 40,848 domains. We plotted the recapture rate of the mark-recapture method in Figure 13. The recapture rate during the measurement period varied from 17% to 90%, and the average recapture rate was 56%. The recapture rate tended to decline gradually throughout the collection period.

5.5 Comparison of graph visualization with graph centrality metrics of each snapshot

Next, we compared the graph visualization results with the graph network centrality metrics. Figure 7 shows the visualization of snapshot A, where approximately eight nodes occupy a relatively large area. When these nodes were examined, we found link collections, search engines, and hosting services. In the graph, the red color represents the destination hyperlink, and the green color represents the hyperlink source. In snapshot A, the visualization shows that hub nodes are linked to many destination nodes. For comparison, we used the calculated degree centrality of snapshot A and identified the categories of the top 10 nodes (Figure 14). The categories identified were link collections and search engines, and the visual results matched this, with the exception of hosting. In addition, the categories of top 10 nodes of snapshot A for both out-degree and in-degree centrality were identified (Figures 15 and 16). The categories for out-degree and in-degree centrality were link collections and search engines, and link collections, markets, and forums, respectively. The visual results matched with the exception of the hosting, markets, and forum categories. Similarly, we investigated the category tendency of the top 10 nodes using PageRank (Figure 17). Similar to degree centrality and out-degree centrality, the categorical trends of PageRank were link collections and search engines.

Similarly, the categories visually extracted from snapshot B were link collections, search engines, and hosting services, as shown in Figure 8. We extracted the categories from each centrality metric. The categories obtained from degree-centrality and out-degree-centrality were link collections and search engines (Figures 14 and 15), and the categories obtained from PageRank were link collections, search engines, and hosting (Figure 15). These were all consistent with the visual results. The categories obtained from in-degree-centrality were link collections, markets, and hosting (Figure 16), and the visual results matched with the exception of markets.

The categories visually extracted in snapshot C were link collections, search engines, hosting services, and pornography (Figure 9). In the visualization of snapshot C, we observed a large green mass that was not observed in previous snapshots. We examined this area and found that it contained many nodes that provided pornographic content. Additionally, one of the visually emphasized nodes was a phishing site that mimicked the search engines that were popular on the dark web during this period. We extracted the categories from each centrality metric. The categorical trends obtained from degree-centrality, out-degree-centrality, and PageRank were link collections, search engines, and hosting. The visual results matched, except for pornography. Meanwhile, the categorical trends obtained from in-degree centrality were link collections, market, hosting, and pornography, and the visual results matched with the exceptions of the search engines and market categories (Figure 16).

The categories visually extracted in snapshot D were link collections, search engines, and pornography (Figure 10). Compared with the visualization of snapshot C, areas corresponding to pornography and phishing were even more prominent in snapshot D. In snapshot D, red masses, which are characteristic of nodes with many hyperlinks, also appeared sparse. Comparing snapshots A and B-D, snapshot D appears greener overall. The higher concentration of green in this visualization implies that nodes with hyperlink sources are densely packed together. This indicates that, unlike in the previous snapshots, not only are the nodes in the network of snapshot D unilaterally linked with the hub node but also more end nodes are linked with each other. We extracted the categories from each centrality metric. The categorical trends obtained from degree centrality, out-degree centrality, and PageRank were link collections, search engines, and hosting, and the visual results matched with the exception of the hosting and pornography categories (Figures 14, 15, and 17). The categorical trends obtained from in-degree centrality were markets, forums, and pornography. Only the pornography category showed a match with the visual results (Figure 16).

The categories visually extracted in snapshot E were search engines and pornography (Figure 11). The visual results of snapshot E clarified the direction of the changes that occurred between snapshots C and D. Compared with the visualizations of snapshots C and D, pornography and phishing colonies are even more prominent in snapshot E, and the red masses are even less noticeable. The number of hub nodes with many links appears to decline overall. Compared with the previous snapshots, the green areas, which denote the link destinations of the network, and mixed red and green areas,

which represent reciprocal links, are expanded. These mixed regions are easier to visually recognize as a mass of nodes, where individual nodes are less noticeable. After sampling and examining several nodes in the mixed area, we found that the area was dominated by pornographic content and phishing nodes. We extracted the categories from each centrality metric. The categorical trends obtained from the degree centrality were link collections, search engines, markets, and forums, and the visual results did not match, except for the search engine category (Figure 14). The categorical trends obtained from out-degree centrality and PageRank were link collections, search engines, and forums, and the visual results did not match, except for the search engine category (Figures 15 and 17). The categorical trends obtained from in-degree centrality were markets, forums, and pornography, and the visual results only matched for the pornography category (Figure 16).

By comparison, we found that the results of some centrality metrics were consistent with those of the visual representation. In the extraction by visualization, it was possible to easily recognize colonies that were difficult to recognize with only the centrality metrics. Hence, it was easy to capture the significant changes that occurred throughout the network.

The results of each network centrality metric showed that the number of link collections in the top 10 hub nodes decreased over time. On the other hand, the number of search engines in the top 10 hub nodes did not appear to change significantly during the measured period. Each centrality index showed that the influence of the search engine increased relative to the decrease in the influence of the link collections in the network.

Finally, we created a graph containing all five snapshots from A to F (Figure 1). This figure shows the entire dark web network observed during the crawling period.

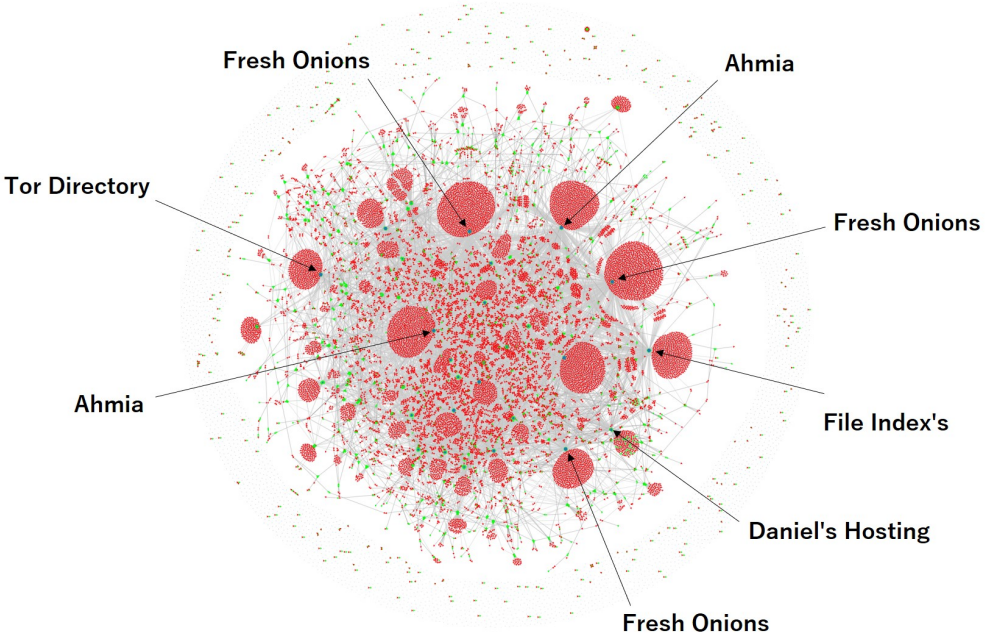

Figure 7: Graph of snapshot A (June 1, 2018, to November 27, 2018).

6 Discussion

6.1 Observed dark web size

During the two years since we started observing the dark web, the number of nodes in the dark web network continued to grow, and from June 2018 to May 2020, the dark web became more diverse

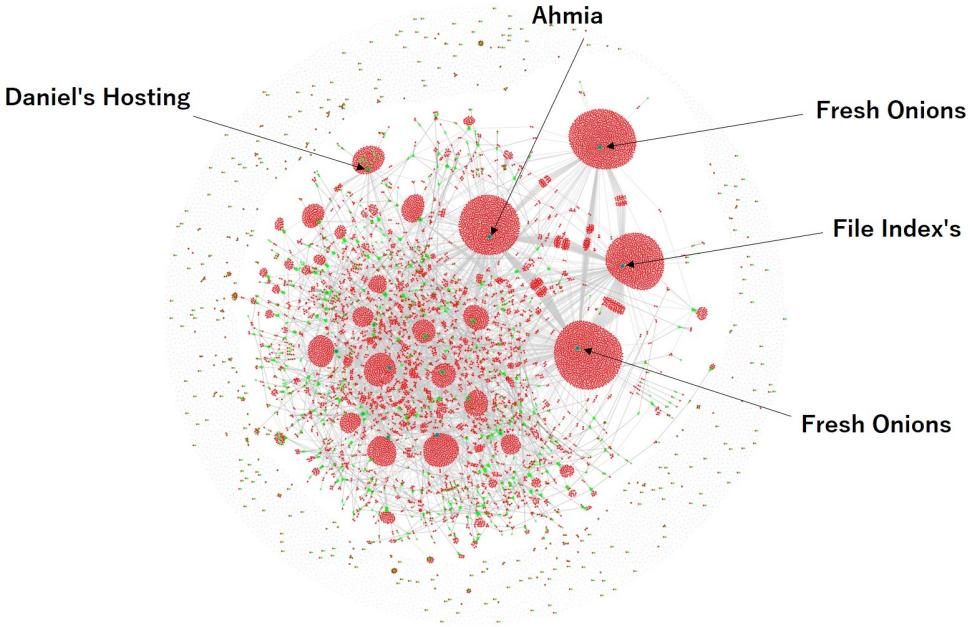

Figure 8: Graph of snapshot B (November 28, 2018, to May 26, 2019).

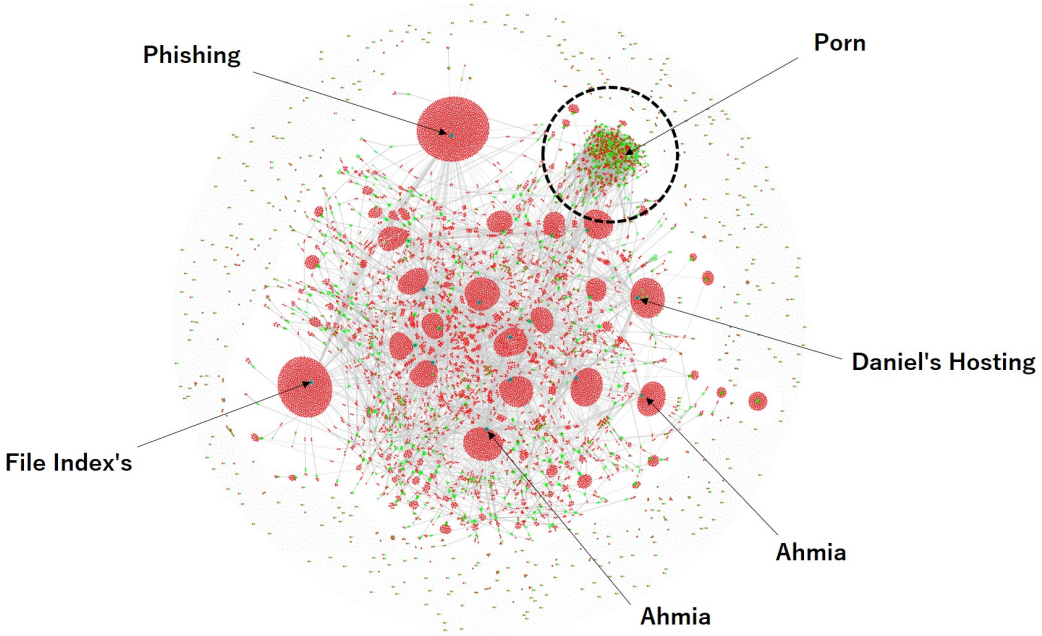

Figure 9: Graph of snapshot C (May 27, 2019, to November 22, 2019).

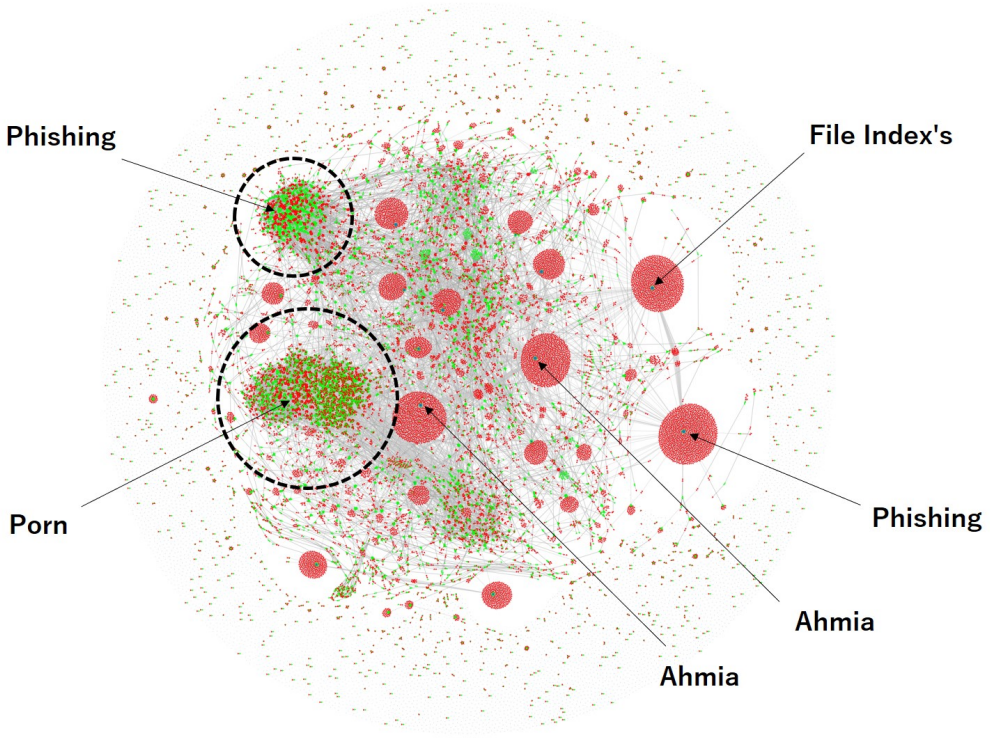

Figure 10: Graph of snapshot D (November 23, 2019, to May 20, 2020).

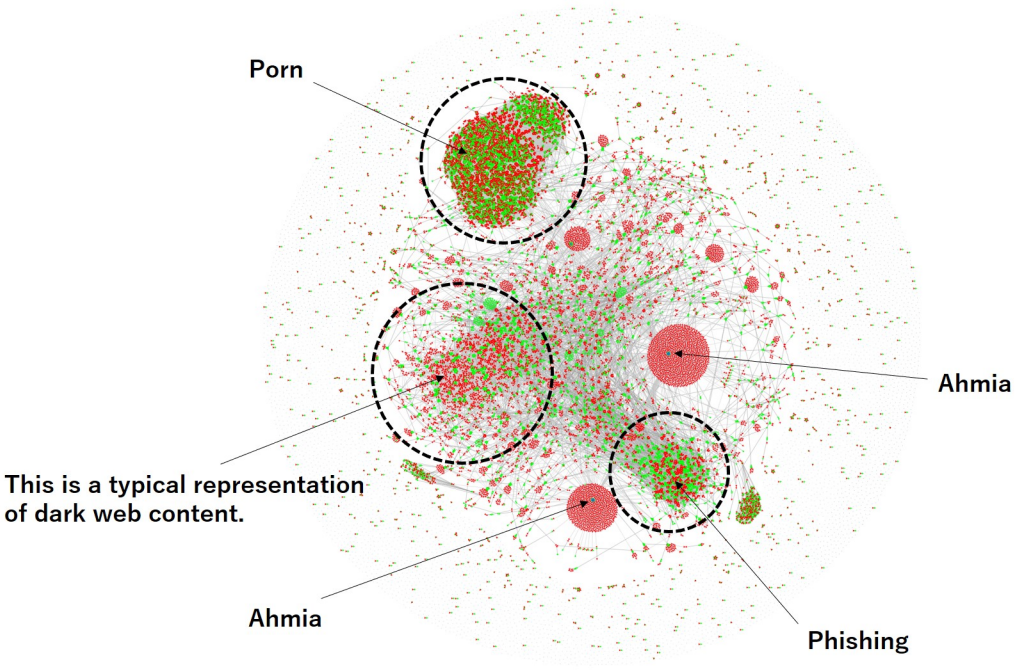

Figure 11: Graph of snapshot E (May 21, 2020, to November 16, 2020).

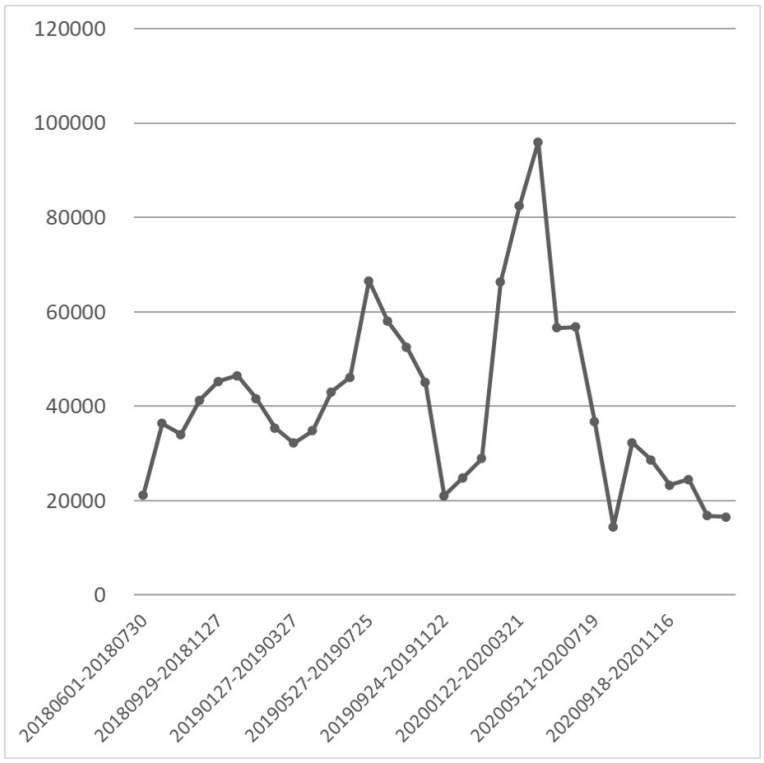

Figure 12: Number of .onion domains distributed on the dark web estimated per 30-day-bin sampling using mark-recapture method.

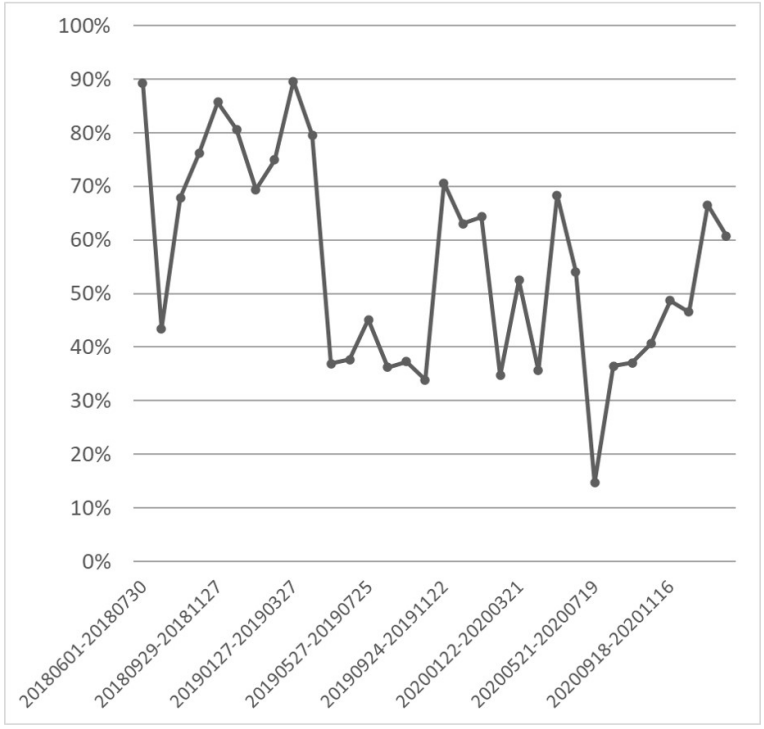

Figure 13: Percentage of recaptured by the mark-recapture method per sampling bin.

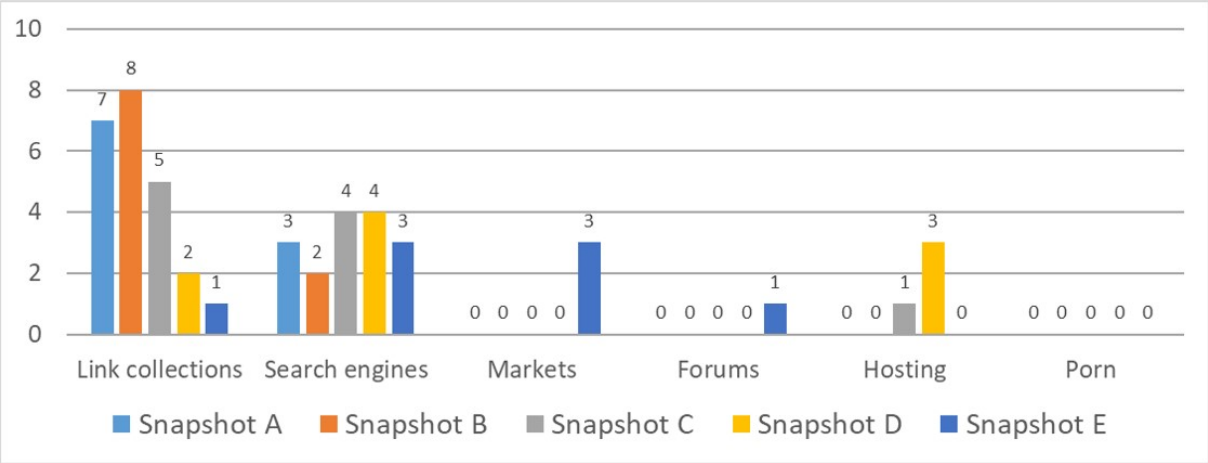

Figure 14: Number of domains in each category in the top 10 dark web hub nodes of degree-centrality (Table 4, 5, 6, 7, 8). (link collection / search engine / markets / forums / hosting / porn)

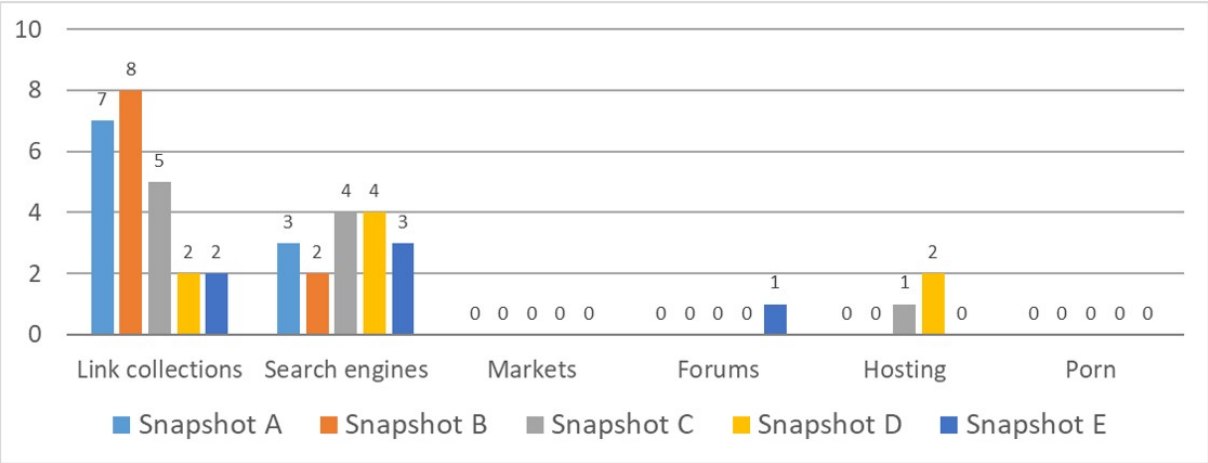

Figure 15: Number of domains in each category in the top 10 dark web hub nodes of out-degree-centrality (Table 10, 11, 12, 13, 14). (link collection / search engine / markets / forums / hosting / porn)

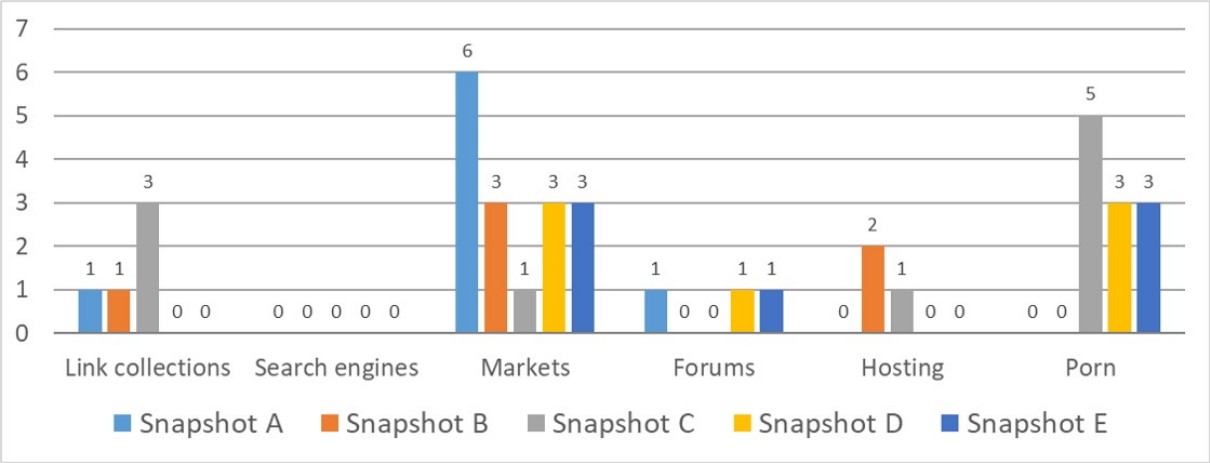

Figure 16: Number of domains in each category in the top 10 dark web hub nodes of in-degree-centrality (Table 16, 17, 18, 19, 20). (link collection / search engine / markets / forums / hosting / porn)

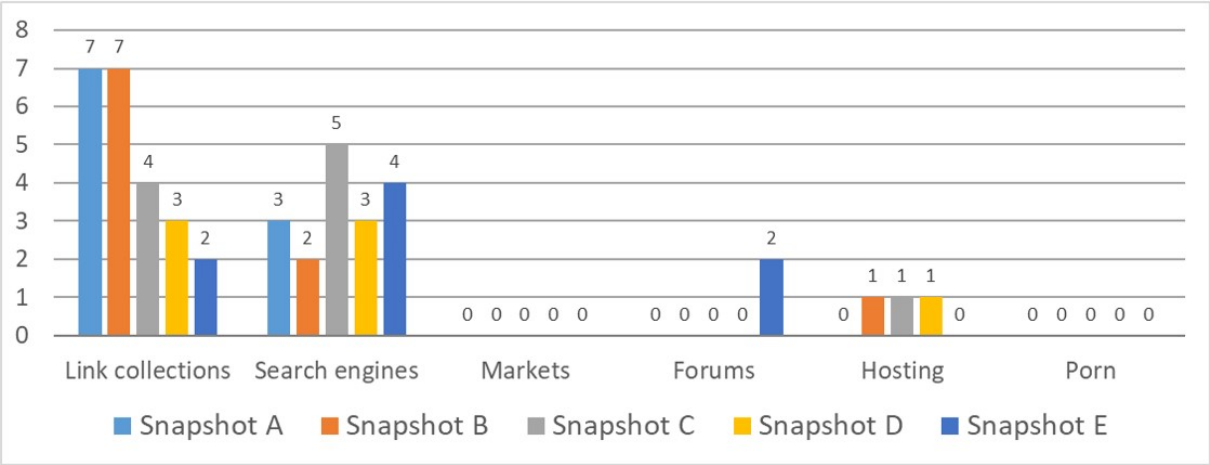

Figure 17: Number of domains in each category in the top 10 dark web hub nodes of PageRank (Table 22, 23, 24, 25, 26). (link collection / search engine / markets / forums / hosting / porn)

and complex. In 2019, law enforcement agencies successfully removed a number of illegal sites on the dark web. Despite these cybercrime measures, there has been no significant impact on the size of the dark web. We believe several factors have promoted the growth of the dark web; one of the most significant factors is that criminal business has become established and larger in scale. According to our observations, since May 2020, the number of nodes decreased significantly over 180 days, equal to the growth experienced over the past two years (Figure 5). On September 22, 2020, the FBI arrested 179 drug traffickers on the dark web [35]. On January 12, 2021, the Dark Market, known as the world's largest dark web marketplace, was taken down by Europol [36]. Did these succeed in making the dark web smaller? Records by TorProject suggest that these efforts have not been successful—the dark web has not become smaller. Figure 18 shows the number of v2 onion addresses measured by the Tor Project. It was recorded that the number of active addresses increased sharply from approximately 100,000 to over 200,000 during the same period, and immediately after, it decreased to approximately 175,000. According to their graph, although there was a sharp increase and then a decrease, the number of active addresses did not seem to have decreased to the number it was before the increase. From a comparison of the number of domains observed in each snapshot, the average number of matching domains is constant at 39,398 (Figure 5). Our observational data suggest that the proportion of highly volatile nodes in the dark web may have increased from 40% to 75% (Figure 19). Our work [37] has shown that by applying diff to the observed domains, there is no bias in the distribution of dark web domains. This result suggests that the proportion of highly volatile nodes may be gradually increasing in the dark web.

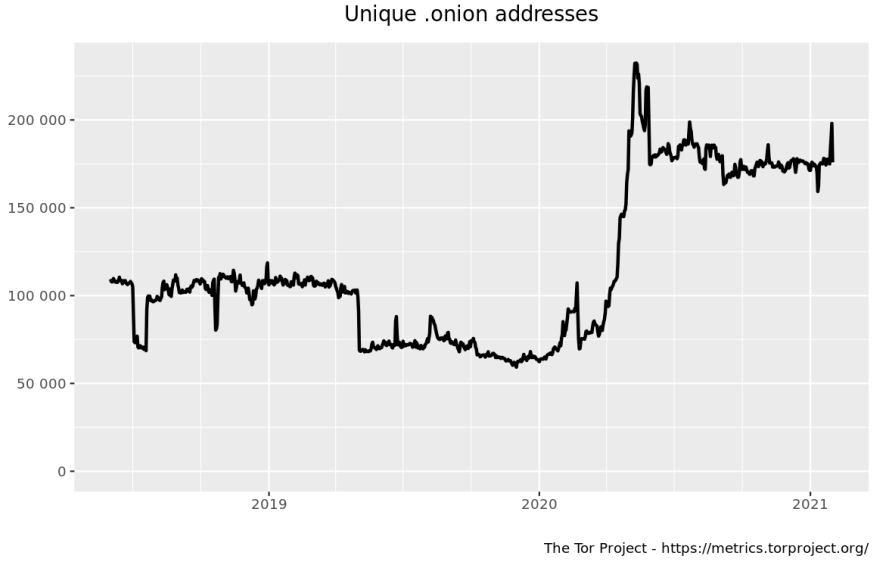

Figure 18: Number of onion services (v2 address only) from June 1, 2018, to January 31, 2021. Source: metrics.torproject.org (under Creative Commons Attribution 3.0 License - CC BY 3.0 US)

6.2 Estimated dark web size

Measuring the size of the dark web is considered difficult [26]. We estimated the number of dark web nodes from the observational data using the popular mark-recapture method in population ecology (Figure 12). Our estimates showed values from 14,509 to 96,034 over 32 measurements at different times. The average for the entire dataset was 40,848. Previous studies on the practicality of the mark-recapture method have shown that with a marked rate of 10%, the relative error is in the ± 0.5 range [38]. In our observation data, the average recapture rate was 56%, and the high recapture rate was maintained continuously during this period (Figure 13). The data we observed were unbiased; the crawl targets were uniformly and randomly selected from all the collected candidate targets

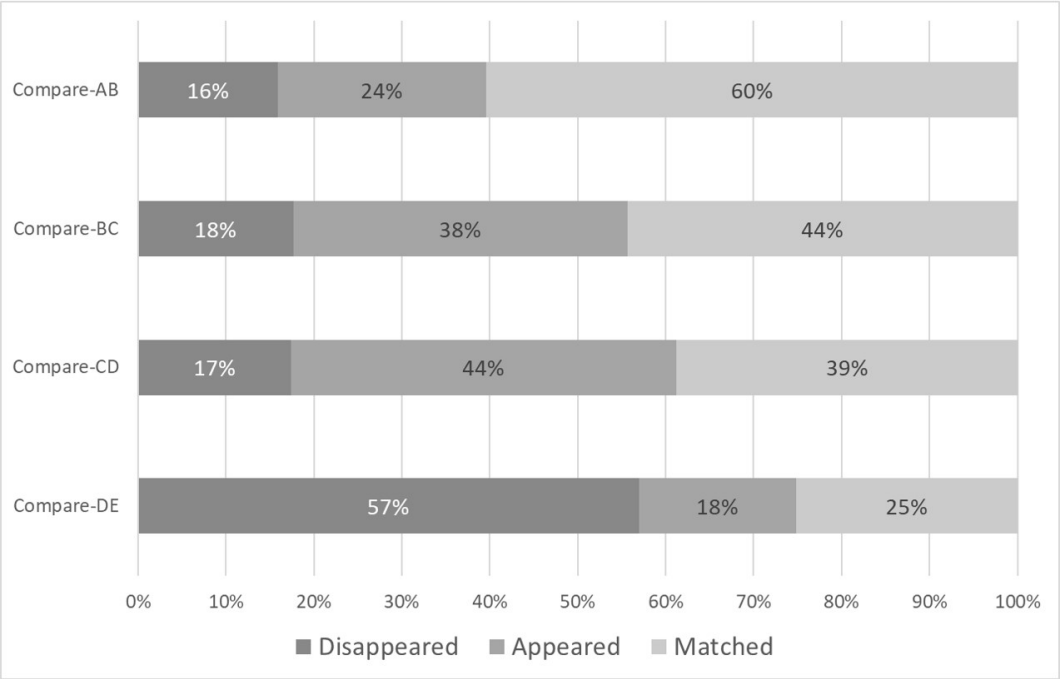

Figure 19: Domain change ratio between two snapshots in the order of disappeared/appeared/matched.

(Figure 6). The recapturing period was set to 30-day intervals to approximate a closed system. Additionally, the domain names used as markers were not intentionally reacquired by an attacker because of computational difficulty using hash; therefore, the domain names functioned correctly as markers.

6.3 Fraudulent network on the dark web

The major change during the observation period that was captured by visualization was that link collections on the dark web became less important and search engines became relatively more important. As link collection sites typically have many links, plotting a hyperlink-related graph will present them as nodes with a high out degree. In our visualization, link collections appeared as a large and noticeable red area. In the visualizations obtained from the first snapshot, many red areas were visible (Figure 7). This result was similar to the analysis results reported in 2017 by Griffith et al., who stated that many nodes on the dark web do not have out links [16]. As we progressed with the snapshots, some nodes with large red areas became dominant in the dark web (Figure 8). A large colony of pornographic content appeared in May 2019 on the dark web (Figure 9). Changes in the graph structure of the dark web also occurred around this time. Colonies of pornographic content were interconnected and did not have as large an out degree as a link collection. This is visually represented by showing the direction of the edges in red and green colors. Colonies of pornographic content continued to grow unabated during the latter part of our observation period (Figure 10). While examining some distinctive nodes in the visualized snapshot D, we observed a large fraudulent network on the dark web during this period (Figure 10). Similar to the colony of pornographic content, the structure of this malicious network did not have a central hub node. In other words, it would have been difficult to find it with a simple network degree centrality. By contrast, in our visualization, we were able to visually recognize these rogue colonies as green areas. The presence of such rogue colonies is a negative characteristic of today's dark web, which allows malicious nodes to exploit high anonymity to stay on the dark web for extended periods of time;

therefore, defense mechanisms that can track and protect against these malicious networks on the dark web are required.

7 Conclusion

In this study, we created snapshots from dark web crawling data divided into half-year time spans. We used a time series graph to investigate the variation in the dark web network over time. First, our visualization showed that the dark web fluctuated significantly during the observation period, making the dark web hyperlink network structure more strongly interconnected. Second, analysis of our dataset suggested that most of the nodes that grew over the last two years might have disappeared rapidly since May 2020. Third, in each snapshot, the difference between the increase and decrease in domains was visualized, confirming that the domain change was not biased throughout the observation period. Fourth, analysis of each snapshot revealed that the proportion of highly volatile domains throughout the network increased from 40% to 75% during the observation period. Fifth, after calculating the network centrality metrics for each snapshot and comparing the transition of hub nodes in chronological order, it became clear that the importance of link collection sites as the main information retrieval method used in the dark web decreased. Moreover, it was confirmed that search engines were becoming increasingly important on the dark web. This is an interesting result because it can overturn the myth that the dark web is a difficult place to search [39]. Sixth, we estimated the size of the dark web from our observed dark web measurements using the mark-recapture method. Estimates showed that the size of the dark web during the observation period fluctuated between 14,509 and 96,034 domains, with an average size of 40,848 domains during the period. To the best of our knowledge, this is the first study to use the mark-recapture method to estimate the size of the dark web network. Finally, after comparing the characteristic nodes of the dark web obtained from the visualized graph with the nodes obtained from the network centrality metrics, it was confirmed that the nodes with high network order match the visual extraction results. Additionally, nodes that did not have a major hub node, such as phishing colonies, can now be intuitively recognized through visualization. Such nodes could not be differentiated using network centrality metrics and were difficult to recognize.

The contribution of this study is that it shows how the dark web changed over time through intuitive and easy-to-understand visualizations.

Our data only reflected those areas that could be traced from the hyperlinks in the Tor network by crawling. In our future work, the inclusion of networks other than those of the http protocol or nodes that cannot be followed from hyperlinks may lead to new findings. Our analysis used snapshots divided into 180 days of crawled data. By increasing or decreasing this time span, different insights could be gained in the future. Our visualization only depicted the domain relationships to reduce computation complexity. We recommend gaining further insight through visualization on a page-by-page basis.

To our knowledge, the network of the dark web has never been easy to understand because it is huge, complex, and dynamic. Our efforts have provided an intuitive understanding and insight into the growth of dark web networks. An accurate understanding of the dark web can be useful to devise various strategies related to cybersecurity.

Acknowledgement

We would like to thank the staff at the Japan Cybercrime Control Center for providing the opportunity and place to conduct the experiment. We would like to thank Ms. Miwako Yamaguchi, President Office of IISEC, and Editage (www.editage.com) for English language editing.

References

[1] Tor. The tor netowrk. `http://www.torproject.org/`.

[2] I2P. The i2p netowrk. http://www.i2p2.de/.

[3] Wikipedia. Dream Market. https://en.wikipedia.org/wiki/Dream_Market.

[4] Darknetlive. Gangsta's Paradise Forum Shut Down by Dread. https://darknetlive.com/post/gangsta-s-paradise-forum-shut-down-by-dread/.

[5] Department of Justice. Three Germans Who Allegedly Operated Dark Web Marketplace with Over 1 Million Users Face U.S. Narcotics and Money Laundering Charges. https://www.justice.gov/opa/pr/three-germans-who-allegedly-operated-dark-web-marketplace-over-1\hyphenmillion-users-face-us.

[6] SC. Europol announces takedown of Wall Street Market and Valhalla dark web markets. https://www.scmagazine.com/home/security-news/government-and-defense/europol-today-announced-it\hyphen-dealt-a-double-blow-to-dark-web-marketplaces-after-taking-down-both\hyphen-the-wall-street-market-and-silkkitie-a-k-a-the-valhalla-marketplace/.

[7] EUROPOL. DEEPDOTWEB SHUT DOWN: ADMINISTRATORS SUSPECTED OF RECEIVING MILLIONS OF KICKBACKS FROM ILLEGAL DARK WEB PROCEEDS. https://www.europol.europa.eu/newsroom/news/deepdotweb-shut-down-administrators-suspected-of-\hyphenreceiving-millions-of-kickbacks-illegal-dark-web-proceeds.

[8] TechNadu. "Dark Web News" Discontinued in Fear of FBI and Europol Arrests. https://www.technadu.com/dark-web-news-discontinued-fear-fbi-europol-arrests/67252/.

[9] EUROPOL. MULTI-MILLION EURO CRYPTOCURRENCY LAUNDERING SERVICE BESTMIXER.IO TAKEN DOWN. https://www.europol.europa.eu/newsroom/news/multi-million-euro-cryptocurrency-laundering-service-bestmixerio-taken-down.

[10] Arnaud Sallaberry, Faraz Zaidi, Christian Pich, and Guy Melançon. Interactive visualization and navigation of web search results revealing community structures and bridges. In *Proceedings of Graphics Interface*, pages 105–112, 2010.

[11] Masashi Toyoda and Masaru Kitsuregawa. A system for visualizing and analyzing the evolution of the web with a time series of graphs. In *Proceedings of the sixteenth ACM conference on Hypertext and hypermedia*, pages 151–160, 2005.

[12] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Computer networks*, 33(1-6):309–320, 2000.

[13] Jon M Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew S Tomkins. The web as a graph: Measurements, models, and methods. In *International Computing and Combinatorics Conference*, pages 1–17. Springer, 1999.

[14] Ruslan Enikeev. The Internet map. https://internet-map.net/.

[15] Georgia Avarikioti, Roman Brunner, Aggelos Kiayias, Roger Wattenhofer, and Dionysis Zindros. Structure and content of the visible darknet. *arXiv preprint arXiv:1811.01348*, 2018.

[16] Virgil Griffith, Yang Xu, and Carlo Ratti. Graph theoretic properties of the darkweb. *arXiv preprint arXiv:1704.07525*, 2017.

[17] Manlio De Domenico and Alex Arenas. Modeling structure and resilience of the dark network. *Physical Review E*, 95(2):022313, 2017.

[18] Tianjun Fu, Ahmed Abbasi, and Hsinchun Chen. A focused crawler for dark web forums. *Journal of the American Society for Information Science and Technology*, 61(6):1213–1231, 2010.

[19] WooHyun Park. A study on analytical visualization of deep web. In *2020 22nd International Conference on Advanced Communication Technology (ICACT)*, pages 81–83. IEEE, 2020.

[20] Gastón L'huillier, Hector Alvarez, Sebastián A Ríos, and Felipe Aguilera. Topic-based social network analysis for virtual communities of interests in the dark web. *ACM SIGKDD Explorations Newsletter*, 12(2):66–73, 2011.

[21] Christopher C Yang and Tobun D Ng. Terrorism and crime related weblog social network: Link, content analysis and information visualization. In *2007 IEEE Intelligence and Security Informatics*, pages 55–58. IEEE, 2007.

[22] Christopher C Yang, Nan Liu, and Marc Sageman. Analyzing the terrorist social networks with visualization tools. In *International Conference on Intelligence and Security Informatics*, pages 331–342. Springer, 2006.

[23] Jeffrey Heer and Danah Boyd. Vizster: Visualizing online social networks. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 32–39. IEEE, 2005.

[24] Yilu Zhou, Edna Reid, Jialun Qin, Hsinchun Chen, and Guanpi Lai. Us domestic extremist groups on the web: link and content analysis. *IEEE intelligent systems*, 20(5):44–51, 2005.

[25] Hyperion Gray. Dark Web Map v2. `https://www.hyperiongray.com/dark-web-map/`.

[26] Carlos Cilleruelo, Luis De-Marcos, Javier Junquera-Sanchez, and Jose-Javier Martinez-Herraiz. Interconnection between darknets. *IEEE Internet Computing*, 2020.

[27] Juha Nurmi. Ahmia. `https://ahmia.fi/`.

[28] Mhd Wesam Al-Nabki, Eduardo Fidalgo, Enrique Alegre, and Laura Fernández-Robles. Torank: Identifying the most influential suspicious domains in the tor network. *Expert Systems with Applications*, 123:212–226, 2019.

[29] DIFF. WinMerge. `https://winmerge.org/`.

[30] D3. Force-directed graph layout. `https://github.com/d3/d3-force/`.

[31] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998.

[32] NetworkX. Network Analysis in Python. `https://networkx.org/`.

[33] PB Best and RW Rand. Results of a pup-tagging experiment on the arctocephalus pusillus rookery at seal island, false bay, south africa. *Rapports et Proces-Verbaux des Reunions (Denmark)*, 1975.

[34] Unperson Hiro. Eschalot. `https://github.com/ReclaimYourPrivacy/eschalot`.

[35] FBI. JCODE Actions in Los Angeles Shut Down Major Darknet Drug Vendor. `https://www.fbi.gov/news/stories/operation-disruptor-jcode-shuts-down-darknet-drug-vendor-092220`.

[36] EUROPOL. DARKMARKET: WORLD'S LARGEST ILLEGAL DARK WEB MARKETPLACE TAKEN DOWN. `https://www.europol.europa.eu/newsroom/news/darkmarket-worlds-largest-illegal-dark-web-marketplace-taken-down`.

[37] Taichi Aoki and Atsuhiro Goto. Graph visualization of the dark web hyperlink. In *2020 The Eighth International Symposium on Computing and Networking (CANDAR)*. IEEE, 2020.

[38] Shuichi Kitada, Sachio Sekiya, and Masashi Yokota. An evaluation of petersen method through experiments in a rearing tank. *Journals of Japanese Society of Fisheries Science*, 67(2):203–208, 2001.

[39] Andy (26 November 2009) Beckett. The dark side of the internet. `https://web.archive.org/web/20130908073158/http://www.theguardian.com/technology/2009/nov/26/dark-side-internet-freenet`.

Appendix

We show the network centrality metrics of each snapshot used in the analysis of this paper.

Table 4: Top 10 nodes by degree centrality in snapshot A. (n=48,360)

rank	node	degree centrality
1	56wr4dvq3abd2ivkf5z36nortvu7dgona55zqsihfaqo2aeg5er4moid.onion	0.602
2	zlal32teyptf4tvi.onion	0.532
3	msydqstlz2kzerdg.onion	0.463
4	msydqstjd6vtexpg.onion	0.445
5	msydqstlzbk3tr5q.onion	0.434
6	wrrkz262g55scqoj.onion	0.351
7	4doqhu4gw5xoddmn.onion	0.296
8	directoryvi6plzm.onion	0.227
9	2xyqdwad2laqcd3v.onion	0.172
10	acjhxk5yqwnw2jdu.onion	0.167

Table 5: Top 10 nodes by degree centrality in snapshot B. (n=53,300)

rank	node	degree centrality
1	56wr4dvq3abd2ivkf5z36nortvu7dgona55zqsihfaqo2aeg5er4moid.onion	0.532
2	msydqstlz2kzerdg.onion	0.449
3	zlal32teyptf4tvi.onion	0.408
4	pejjyyh7rhv5ctyu.onion	0.387
5	2xyqdwad2laqcd3v.onion	0.173
6	acjhxk5yqwnw2jdu.onion	0.172
7	darkweb2zz7etehx.onion	0.128
8	cb3robuo3hobodw6.onion	0.104
9	darkjmgsio2e3cq3.onion	0.103
10	3bbaaaccczcbdddz.onion	0.103

Table 6: Top 10 nodes by degree centrality in snapshot C. (n=70,684)

rank	node	degree centrality
1	visiwnqyii4r5f5l.onion	0.354
2	pejjyyh7rhv5ctyu.onion	0.280
3	msydqstlz2kzerdg.onion	0.158
4	2xyqdwad2laqcd3v.onion	0.150
5	3bbad7fauom4d6sgppalyqddsqbf5u5p56b5k5uk2zxsy3d6ey2jobad.onion	0.147
6	acjhxk5yqwnw2jdu.onion	0.143
7	dhosting4xxoydyaivckq7tsmtgi4wfs3flpeyitekkmqwu4v4r46syd.onion	0.136
8	3bbaaaccczcbdddz.onion	0.116
9	cb3rob5vwac2dtyc.onion	0.112
10	cb3robuo3hobodw6.onion	0.099

Table 7: Top 10 nodes by degree centrality in snapshot D. (n=103,674)

rank	node	degree centrality
1	msydqstlz2kzerdg.onion	0.246
2	visicgxrfb443cqh.onion	0.237
3	msyd6emf7clejhld.onion	0.212
4	pejjyyh7rhv5ctyu.onion	0.204
5	42yn43ahvsm7tonn.onion	0.084
6	dhosting4xxoydyaivckq7tsmtgi4wfs3flpeyitekkmqwu4v4r46syd.onion	0.080
7	3bbaaaccczcbdddz.onion	0.074
8	3bbad7fauom4d6sgppalyqddsqbf5u5p56b5k5uk2zxsy3d6ey2jobad.onion	0.066
9	dhosting4okcs22v.onion	0.057
10	kj7fx3wribkd4ara.onion	0.055

Table 8: Top 10 nodes by degree centrality in snapshot E. (n=54,509)

rank	node	degree centrality
1	msydqstlz2kzerdg.onion	0.478
2	msyd6emf7clejhld.onion	0.272
3	3bbaaaccczcbdddz.onion	0.094
4	jo7np7o7hhwiksse.onion	0.043
5	oa7afrpjhd6ffocc.onion	0.035
6	m6kshkn2vcovss35.onion	0.030
7	visicgxrfb443cqh.onion	0.030
8	qyh35wx75lkf6z55.onion	0.025
9	2pneiouz2aj27kjs.onion	0.024
10	sk3w2x7g2gksovk6.onion	0.024

Table 9: Top 10 nodes by degree centrality in snapshots A-F. (n=172,740)

rank	node	degree centrality
1	msydqstlz2kzerdg.onion	0.389
2	56wr4dvq3abd2ivkf5z36nortvu7dgona55zqsihfaqo2aeg5er4moid.onion	0.188
3	msyd6emf7clejhld.onion	0.180
4	zlal32teyptf4tvi.onion	0.171
5	visiwnqyii4r5f5l.onion	0.145
6	visicgxrfb443cqh.onion	0.142
7	3bbaaaccczcbdddz.onion	0.133
8	msydqstjd6vtexpg.onion	0.125
9	pejjyyh7rhv5ctyu.onion	0.123
10	msydqstlzbk3tr5q.onion	0.122

Table 10: Top 10 nodes by out-degree centrality in snapshot A. (n=48,360)

rank	node	out-degree centrality
1	56wr4dvq3abd2ivkf5z36nortvu7dgona55zqsihfaqo2aeg5er4moid.onion	0.602
2	zlal32teyptf4tvi.onion	0.532
3	msydqstlz2kzerdg.onion	0.462
4	msydqstjd6vtexpg.onion	0.444
5	msydqstlzbk3tr5q.onion	0.434
6	wrrkz262g55scqoj.onion	0.351
7	4doqhu4gw5xoddmn.onion	0.296
8	directoryvi6plzm.onion	0.226
9	2xyqdwad2laqcd3v.onion	0.172
10	acjhxk5yqwnw2jdu.onion	0.167

Table 11: Top 10 nodes by out-degree centrality in snapshot B. (n=53,300)

rank	node	out-degree centrality
1	56wr4dvq3abd2ivkf5z36nortvu7dgona55zqsihfaqo2aeg5er4moid.onion	0.532
2	msydqstlz2kzerdg.onion	0.448
3	zlal32teyptf4tvi.onion	0.408
4	pejjyyh7rhv5ctyu.onion	0.387
5	2xyqdwad2laqcd3v.onion	0.173
6	acjhxk5yqwnw2jdu.onion	0.172
7	darkweb2zz7etehx.onion	0.128
8	cb3robuo3hobodw6.onion	0.104
9	darkjmgsio2e3cq3.onion	0.103
10	3bbaaaccczcbdddz.onion	0.102

Table 12: Top 10 nodes by out-degree centrality in snapshot C. (n=70,684)

rank	node	out-degree centrality
1	visiwnqyii4r5f5l.onion	0.354
2	pejjyyh7rhv5ctyu.onion	0.280
3	msydqstlz2kzerdg.onion	0.158
4	2xyqdwad2laqcd3v.onion	0.150
5	3bbad7fauom4d6sgppalyqddsqbf5u5p56b5k5uk2zxsy3d6ey2jobad.onion	0.147
6	acjhxk5yqwnw2jdu.onion	0.143
7	dhosting4xxoydyaivckq7tsmtgi4wfs3flpeyitekkmqwu4v4r46syd.onion	0.136
8	3bbaaaccczcbdddz.onion	0.116
9	cb3rob5vwac2dtyc.onion	0.112
10	cb3robuo3hobodw6.onion	0.099

Table 13: Top 10 nodes by out-degree centrality in snapshot D. (n=103,674)

rank	node	out-degree centrality
1	msydqstlz2kzerdg.onion	0.245
2	visicgxrfb443cqh.onion	0.236
3	msyd6emf7clejhld.onion	0.212
4	pejjyyh7rhv5ctyu.onion	0.204
5	42yn43ahvsm7tonn.onion	0.084
6	dhosting4xxoydyaivckq7tsmtgi4wfs3flpeyitekkmqwu4v4r46syd.onion	0.079
7	3bbaaaccczcbdddz.onion	0.074
8	3bbad7fauom4d6sgppalyqddsqbf5u5p56b5k5uk2zxsy3d6ey2jobad.onion	0.066
9	dhosting4okcs22v.onion	0.056
10	kj7fx3wribkd4ara.onion	0.055

Table 14: Top 10 nodes by out-degree centrality in snapshot E. (n=54,509)

rank	node	out-degree centrality
1	msydqstlz2kzerdg.onion	0.476
2	msyd6emf7clejhld.onion	0.272
3	3bbaaaccczcbdddz.onion	0.093
4	jo7np7o7hhwiksse.onion	0.043
5	visicgxrfb443cqh.onion	0.030
6	runionwe25lxr3is.onion	0.019
7	onioofr4na4kvarv.onion	0.018
8	wikilink77h7lrbi.onion	0.018
9	oniot2zvfczp4lpc.onion	0.017
10	h7xxo7pbd3fjlru6.onion	0.016

Table 15: Top 10 nodes by out-degree centrality in snapshots A-F. (n=172,740)

rank	node	out-degree centrality
1	msydqstlz2kzerdg.onion	0.387
2	56wr4dvq3abd2ivkf5z36nortvu7dgona55zqsihfaqo2aeg5er4moid.onion	0.188
3	msyd6emf7clejhld.onion	0.180
4	zlal32teyptf4tvi.onion	0.170
5	visiwnqyii4r5f5l.onion	0.145
6	visicgxrfb443cqh.onion	0.142
7	3bbaaaccczcbdddz.onion	0.132
8	msydqstjd6vtexpg.onion	0.124
9	pejjyyh7rhv5ctyu.onion	0.123
10	msydqstlzbk3tr5q.onion	0.121

Table 16: Top 10 nodes by in-degree centrality in snapshot A. (n=48,360)

rank	node	in-degree centrality
1	dhosting4okcs22v.onion	0.015
2	blockchainbdgpzk.onion	0.005
3	lchudifyeqm4ldjj.onion	0.004
4	tmskhzavkycdupbr.onion	0.003
5	jd6yhuwcivehvdt4.onion	0.003
6	blockchainbdgpz.onion	0.003
7	6qlocfg6zq2kyacl.onion	0.003
8	k3pd243s57fttnpa.onion	0.003
9	7ep7acrkunzdcw3l.onion	0.003
10	xytjqcfendzeby22.onion	0.003

Table 17: Top 10 nodes by in-degree centrality in snapshot B. (n=53,300)

rank	node	in-degree centrality
1	dhosting4xxoydyaivckq7tsmtgi4wfs3flpeyitekkmqwu4v4r46syd.onion	0.006
2	blockchainbdgpzk.onion	0.003
3	dc72oitx4plhel3r.onion	0.003
4	torpress2sarn7xw.onion	0.003
5	z5taguvepvuevyp3.onion	0.002
6	au54gdcynh7jvet6.onion	0.002
7	xpotbpgfnliiudm.onion	0.002
8	torbox3uiot6wchz.onion	0.002
9	blockchainbdgpz.onion	0.002
10	lchudifyeqm4ldjj.onion	0.002

Table 18: Top 10 nodes by in-degree centrality in snapshot C. (n=70,684)

rank	node	in-degree centrality
1	cashvd5pznwrucio.onion	0.003
2	r72jjnwl6nh47zaks2loja4i34gh4soytlplk2tzsnd2s4rjptxf6ayd.onion	0.003
3	3h4ctuedj6meudw2k2u3td4j472rwqxekvqnkou7bvfrcwiztt66gfid.onion	0.003
4	dhosting4xxoydyaivckq7tsmtgi4wfs3flpeyitekkmqwu4v4r46syd.onion	0.003
5	oqqsn7yplu24dfirwerruz6bdupi6gqxzdxoxoacteyzoyfnyhjxn7qd.onion	0.003
6	nj3xpw7vvzzca3mvvaagzeri6nqclvdrdc5jdcc6t34q2hsiiq7dchad.onion	0.003
7	xq4qjtl354vrst67ko3dait6mmw3r4nbzry332lqrivyz4jwvsv3ogad.onion	0.003
8	uyqmzqvu252txhkqfy54a5bn7qk2z3v4vargjvqzm6wwrvgn6lod3fqd.onion	0.003
9	hueonvufzh2pps3gfsistijpj5u2v5hrohci4gc27venko2thuv63yad.onion	0.003
10	i3eauqwhnya2tdkxmp4enn4wi6h4mhu5nivkruwtet7qhd7mffi4crad.onion	0.003

Table 19: Top 10 nodes by in-degree centrality in snapshot D. (n=103,674)

rank	node	in-degree centrality
1	oa7afrpjhd6ffocc.onion	0.014
2	m6kshkn2vcovss35.onion	0.012
3	z5taguvepvuevyp3.onion	0.010
4	2pneiouz2aj27kjs.onion	0.009
5	qyh35wx75lkf6z55.onion	0.009
6	sk3w2x7g2gksovk6.onion	0.009
7	ahhq2dssfo7e77ui.onion	0.006
8	lolipornp7zsenhy3k3khowbdhyn4afshaj6axw7lakg5xgguezcinqd.onion	0.006
9	lolipornngyulcxe3goplk3cjao65lpxfhgkph6fdompotzjxlkplmid.onion	0.006
10	loliporns5lm2t3lvojm6k7jjhduhpsty43g7izoerx475jtuticpiqd.onion	0.006

Table 20: Top 10 nodes by in-degree centrality in snapshot E. (n=54,509)

rank	node	in-degree centrality
1	oa7afrpjhd6ffocc.onion	0.035
2	m6kshkn2vcovss35.onion	0.030
3	qyh35wx75lkf6z55.onion	0.024
4	2pneiouz2aj27kjs.onion	0.024
5	sk3w2x7g2gksovk6.onion	0.024
6	z5taguvepvuevyp3.onion	0.022
7	coud657j6pcfkp6z.onion	0.017
8	lolipornngyulcxe3goplk3cjao65lpxfhgkph6fdompotzjxlkplmid.onion	0.017
9	lolipornp7zsenhy3k3khowbdhyn4afshaj6axw7lakg5xgguezcinqd.onion	0.016
10	loliporncty s4f7k6rmr4rqutnsbjvmbrpwyb6nzhcvxa6tmvuwh34yd.onion	0.016

Table 21: Top 10 nodes by in-degree centrality in snapshots A-F. (n=172,740)

rank	node	in-degree centrality
1	oa7afrpjhd6ffocc.onion	0.014
2	m6kshkn2vcovss35.onion	0.012
3	qyh35wx75lkf6z55.onion	0.010
4	2pneiouz2aj27kjs.onion	0.010
5	sk3w2x7g2gksovk6.onion	0.010
6	z5taguvepvuevyp3.onion	0.008
7	ahhq2dssfo7e77ui.onion	0.007
8	blockchainbdgpzk.onion	0.006
9	cpmovieswo2ww7346l6qydxeg7vbeoey3hk23uh4ivdvdb5w7lpwahid.onion	0.006
10	lolipornngyulcxe3goplk3cjao65lpxfhgkph6fdompotzjxlkplmid.onion	0.006

Table 22: Top 10 nodes by PageRank in snapshot A. (n=48,360)

rank	node	PageRank
1	zlal32teyptf4tvi.onion	0.045
2	56wr4dvq3abd2ivkf5z36nortvu7dgona55zqsihfaqo2aeg5er4moid.onion	0.041
3	msydqstlz2kzerdg.onion	0.027
4	msydqstjd6vtexpg.onion	0.026
5	msydqstlzbk3tr5q.onion	0.026
6	directoryvi6plzm.onion	0.024
7	4doqhu4gw5xoddmn.onion	0.023
8	wrrkz262g55scqoj.onion	0.022
9	zdxgqrvvvpwnuj2n.onion	0.010
10	3wfzef3ql23dngkh.onion	0.009

Table 23: Top 10 nodes by PageRank in snapshot B. (n=53,300)

rank	node	PageRank
1	zlal32teyptf4tvi.onion	0.072
2	56wr4dvq3abd2ivkf5z36nortvu7dgona55zqsihfaqo2aeg5er4moid.onion	0.068
3	msydqstlz2kzerdg.onion	0.052
4	pejjyyh7rhv5ctyu.onion	0.046
5	2xyqdwad2laqcd3v.onion	0.014
6	acjhxk5yqwnw2jdu.onion	0.014
7	dhosting4xxoydyaivckq7tsmtgi4wfs3flpeyitekkmqwu4v4r46syd.onion	0.013
8	darkweb2zz7etehx.onion	0.010
9	cratedvnn5z57xhl.onion	0.010
10	msyd5324jqjq4jw6.onion	0.009

Table 24: Top 10 nodes by PageRank in snapshot C. (n=70,684)

rank	node	PageRank
1	visiwnqyii4r5f5l.onion	0.091
2	pejjyyh7rhv5ctyu.onion	0.067
3	dhosting4xxoydyaivckq7tsmtgi4wfs3flpeyitekkmqwu4v4r46syd.onion	0.017
4	msydqstlz2kzerdg.onion	0.017
5	msyd5324jqjq4jw6.onion	0.014
6	3bbad7fauom4d6sgppalyqddsqbf5u5p56b5k5uk2zxsy3d6ey2jobad.onion	0.014
7	2xyqdwad2laqcd3v.onion	0.013
8	jld3zkuo4b5mbios.onion	0.012
9	acjhxk5yqwnw2jdu.onion	0.012
10	3bbaaaccczcbdddz.onion	0.011

Table 25: Top 10 nodes by PageRank in snapshot D. (n=103,674)

rank	node	PageRank
1	visicgxrfb443cqh.onion	0.060
2	pejjyyh7rhv5ctyu.onion	0.048
3	msydqstlz2kzerdg.onion	0.029
4	msyd6emf7clejhld.onion	0.027
5	42yn43ahvsm7tonn.onion	0.015
6	kj7fx3wribkd4ara.onion	0.011
7	5uqoveemnt3nlb3o.onion	0.011
8	dhosting4xxoydyaivckq7tsmtgi4wfs3flpeyitekkmqwu4v4r46syd.onion	0.009
9	3bbaaaccczcbdddz.onion	0.007
10	jo7np7o7hhwiksse.onion	0.007

Table 26: Top 10 nodes by PageRank in snapshot E. (n=54,509)

rank	node	PageRank
1	msydqstlz2kzerdg.onion	0.104
2	msyd6emf7clejhld.onion	0.054
3	3bbaaaccczcbdddz.onion	0.015
4	jo7np7o7hhwiksse.onion	0.011
5	visicgxrfb443cqh.onion	0.007
6	oa7afrpjhd6ffocc.onion	0.003
7	runionwe25lxr3is.onion	0.002
8	il2tacbrarmadm4q.onion	0.002
9	m6kshkn2vcovss35.onion	0.002
10	2f7f4rlcu62w4zph.onion	0.002

Table 27: Top 10 nodes by PageRank in snapshots A-F. (n=172,740)

rank	node	PageRank
1	msydqstlz2kzerdg.onion	0.042
2	msyd6emf7clejhld.onion	0.022
3	zlal32teyptf4tvi.onion	0.014
4	56wr4dvq3abd2ivkf5z36nortvu7dgona55zqsihfaqo2aeg5er4moid.onion	0.014
5	visiwnqyii4r5f5l.onion	0.013
6	visicgxrfb443cqh.onion	0.013
7	dhosting4xxoydyaivckq7tsmtgi4wfs3flpeyitekkmqwu4v4r46syd.onion	0.012
8	3bbaaaccczcbdddz.onion	0.011
9	3bbad7fauom4d6sgppalyqddsqbf5u5p56b5k5uk2zxsy3d6ey2jobad.onion	0.009
10	42yn43ahvsm7tonn.onion	0.009

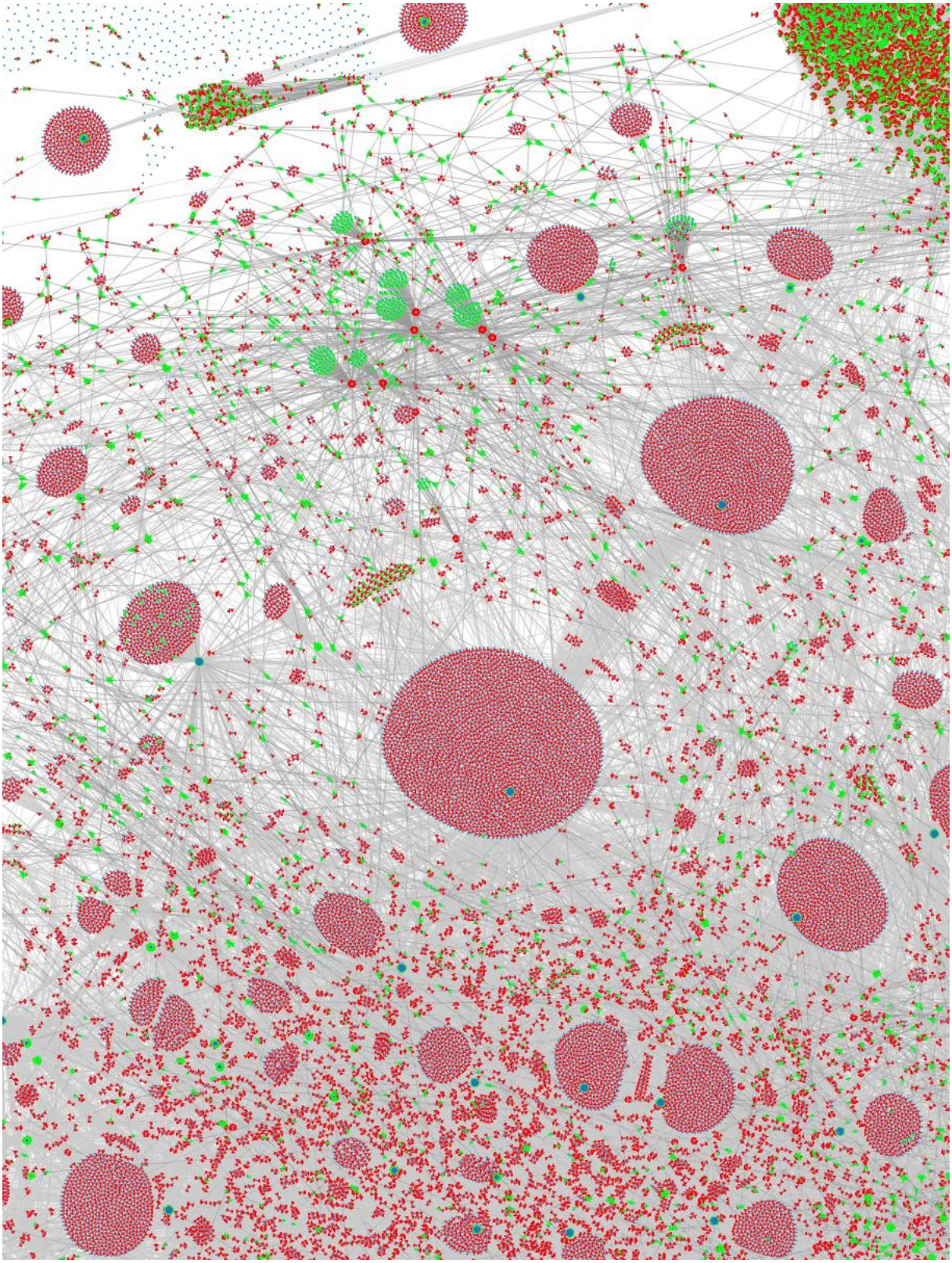

Figure 20: Enlarged view of the dark web hyperlink network graph from June 1, 2018, to January 30, 2021.