# Employing social network analysis to dark web communities

Sotirios Nikoletos* and Paraskevi Raftopoulou†
Dept. of Informatics & Telecommunications, University of the Peloponnese, GR22131 Tripolis, Greece
email: *sothrhsn@hotmail.com, †praftop@uop.gr

***Abstract*—Deep web refers to sites that cannot be found by search engines and makes up the 96% of the digital world. The dark web is the part of the deep web that can only be accessed through specialised tools and anonymity networks. To avoid monitoring and control, communities that seek for anonymization are moving to the dark web. In this work, we scrape five dark web forums and construct five graphs to model user connections. These networks are then studied and compared using data mining techniques and social network analysis tools; for each community we identify the key actors, we study the social connections and interactions, we observe the small world effect, and we highlight the type of discussions among the users. Our results indicate that only a small subset of users are influential, while the rapid dissemination of information and resources between users may affect behaviours and formulate ideas for future members.**

***Index Terms*—dark web, social network analysis, data mining, key nodes, social interactions**

## I. Introduction

Regularly, whenever we perform an Internet search we have access to the visible (or else surface) web, which includes all the websites that can be found by the classic search engines, such as Google, Bing or Yahoo, and it constitutes only the 4% of the total Internet. However, the Internet is a strange and mysterious place; beneath its surface, there exists a huge tangle of invisible web pages, much larger than we know or can imagine. This is the *deep web*; it refers to sites that cannot be found by search engines and makes up the 96% of the digital world. The *dark web* is the part of the deep web that can only be accessed through specialised software/tools and authorization to access. One of the popular tools used to access the dark web is The Onion Router (Tor), that is a free and open-source software for enabling anonymous communication.

With the increasing monitoring and control by the platforms, communities that seek for anonymization are moving to the dark web, which hosts content ranging from complaints and privacy to cybercrime, sexual harassment and drugs [1]. Previous research on dark web has focused on identifying the topics of discussion [2], uncovering the geographical origin of the users [3], understanding the profiles of criminals [4], detecting information in the areas of cyber security and fraud [5], [6], identifying terrorist content and extremist groups [7]–[9], applying machine learning for proactive cyber-threat intelligence [10]–[13], etc. However, dark web presents a rich ecosystem of communities, since at the core of these underground forums are members who interact with each other. Although there are research works that explore digital underground economies and the key actors [14]–[19], the structures, functions, and interactions of the dark web communities have not yet been discovered at all their extend, mainly due to the difficulties associated with the data collection. Social network analysis may though provide valuable information on how these communities operate. We thus, turn our attention to the underground forums, highlight the communities built inside these networks, identify the key actors, and examine in what way members connect and interact to each other.

In this work, we scraped *five dark web forums* of different languages -related to cybercrime, use and trafficking of drugs, personal experiences, politics, etc.- to obtain information about their organisational structures that form the basis of information and resource flows within these communities. It is important to note that these networks are constantly evolving making it difficult to monitor the involved actors and the corresponding connections. Thus, our study focuses on a static snapshot of the network that concerns the period between *September* and *December 2021*. We then, constructed five directed graphs to model the connections among the members of the forums. These networks are studied and compared using data analysis techniques and social network analysis tools. More specifically, we

- identify the key actors in each dark web (sub-)forum,
- study the social connections and interactions within each community,
- observe the small world effect in these communities, and
- highlight the type of discussions among the users of the forums.

The rest of the paper is organised as follows. In Section II, we present state-of-the-art research related to dark web analysis using social network principles, while Section III presents the metrics used to understand the social structures and the users' characteristics. In Section IV, we describe the process of retrieving the data of five forums from the dark web and present the characteristics of each forum. In Section V, we present and discuss the results of our research. Finally, Section VI concludes this paper giving also future directions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme Foresight under grant agreement no. 833673. The work reflects only the authors' view and the Agency is not responsible for any use that may be made of the information it contains.

## II. Related Work

In the Internet we know, that holds 4% of the total Internet, we resort to Social Network Analysis (SNA) techniques, in order to gather valuable information about social network structure (experts/influencers, user groups, followers, etc.) based on the relationships between social network members, which are formalised as a graph representation [20]. The huge tangle of invisible web pages in the deep web along with the constant moving of users from the surface to the dark web have recently drawn the evolving interest of the research community. Can we apply the same or similar techniques used in the surface web in order to explore digital underground economies? The research in [21] suggests that SNA is capable of revealing significant insights into the dynamics of dark networks, particularly in the identification of critical nodes. In what follows, we present the state-of-the-art research related to dark web analysis using social network principles.

A large part of the research community is concerned with examining the structure of the dark web network. Towards this direction, the work in [15] forms a network based on Tor hyperlinks and explores traditional social organisation principles to draw comparisons between these virtual communities and real-life crime-prone neighborhoods. Similarly, [16] collects and analyses the dark web hyperlink graph; the presented results indicate that the graph under investigation is highly dissimilar to the well-studied world wide web hyperlink graph (e.g., $> 87\%$ of dark web sites never link to another site) and eventually, suggest to view the dark web as a set of largely isolated dark silos. The study in [19] investigates the structure of the dark web and finds that its topology is characterised by a non-homogeneous distribution of connections and low-connected nodes, revealing that this structure makes the dark web much more resilient than the Internet to random failures, targeted attacks, and cascade failures.

Another part of the research community, which is more relevant to our study, examines the social ties of users inside dark web networks. For example, the work in [14] uses network analysis techniques to identify members with central roles inside Islamic virtual communities. In [18], theoretical network approaches are used to analyse the data of two types of interaction networks and establish several characteristic behavioural patterns, while [17] analyses dark web forums and social networks with topics of interest such as drugs, guns, hacking, etc. aiming at observing the structure of each network, highlighting structural patterns, and identifying nodes of importance. Our study extends the work in [17] by (i) investigating a more recent time period, (ii) analysing also the sub-forums (as resulted by different discussion topics), and (iii) contrasting dark web and surface web social networks; we endeavore to investigate the reciprocal action or influence of the underlying communities and to identify members' roles based on each node's interconnections and posting activity.

## III. Theoretical Background

SNA is the application of graph theory on social networks, focusing on social structures or pattens of relations among people or groups. As social communities emmerge in the Internet, SNA has been used as a critical tool to study Web hyperlinks and identify the network structures in terms of nodes (or users or actors) and the links (or relationships or ties or edges) that connect them [14]–[19]. SNA provides a set of powerful metrics for perceiving of the social structures and the individuals and groups within them [22].

**Degree centrality** measures the number of incoming or outgoing (or both) links held by each node. Although it is the simplest measure of *node connectivity*, it can determine the nodes of importance in a social network. It ranges from 0 (if a node has no connections) to $total\ number\ of\ nodes - 1$ (if the node connects to all other nodes in the network).

**Closeness centrality** scores each node based on its *closeness* to all other nodes in the network; it is used as a way of detecting the nodes who are best placed to spread information and influence the entire network most quickly. It is calculated as the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the graph:

$$C_x = 1/\sum d(x, y),$$

where $d(x, y)$ is the distance between nodes $x$ and $y$. Closeness centrality ranges from 0 to 1. Values close to 0 indicate that a given node is far from other nodes in the network (i.e., many links must be traversed to get from that node to other nodes in the network), whereas a value close to 1 indicates that a given node is close to other nodes in the network (i.e., few links must be traversed).

**Eigenvector centrality** is a measure that identifies nodes with *influence* over the whole network, not just those directly connected to it. Eigenvector centrality measures a node's influence based on the number of links it has to other nodes in the network, taking also into account how *well-connected* these nodes are; links originating from high-scoring nodes contribute more to the score of this node than connections from low-scoring nodes. It ranges between 0 and 1; a high eigenvector score means that a node is connected to many nodes who themselves have high scores.

**Betweenness centrality** measures the number of times a node lies on the shortest path between other nodes; this is a way of detecting the nodes that *influence the flow of information* in a network and it is often used to find nodes that serve as a bridge between different sub-parts of a network. The betweenness centrality of a node $x$ is given by the expression:

$$B_x = \sum_{x \neq y \neq z} (\sigma_{yz}(x)/\sigma_{yz}),$$

where $\sigma_{yz}$ is the total number of shortest paths from node $y$ to node $z$ and $\sigma_{yz}(x)$ is the number of those paths that pass through $x$. The betweenness centrality scales with the number of pairs of nodes; scores close to 0 mean that a node is isolated.

## IV. Datasets

In what follows, we describe the process of retrieving the data of five forums from the dark web and present the characteristics of each forum.

TABLE I
OVERVIEW OF THE SCRAPED FORUMS

| *Forum* | *# of Nodes* | *# of Edges* | *Density* |
|---|---|---|---|
| *Forum A* | 14,845 | 37,427 | 0.00033 |
| *Forum B* | 1,188 | 3,637 | 0.0051 |
| *Forum C* | 1,493 | 3,712 | 0.0033 |
| *Forum D* | 1,001 | 767 | 0.0015 |
| *Forum E* | 747 | 621 | 0.0022 |

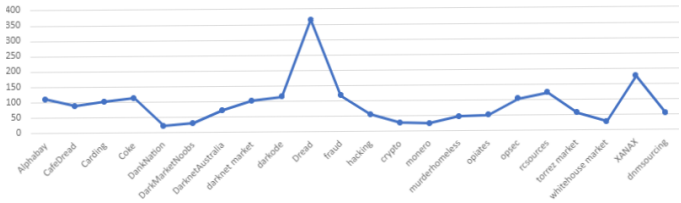


Fig. 1. Forum A: Maximum degree centrality per subforum

### A. Data crawling and storing

For extracting and retrieving the data from dark web forums, we used a web scraping module based upon (1) the open-source python framework Scrapy[1] and (2) the python library Beautiful Soup.[2] By utilising this module and TOR proxies, we performed in-depth crawls on specific websites on the dark web. To do so, the crawler was provided with a number of onion links that corresponded to the forums and monitored the underground discussions. To overcome user authentication mechanisms that are often in place in dark web forums, the dark web crawler required an initial manual login. After a successful user authentication, the session cookies were stored and were utilised (via HTTP requests) in subsequent visits of the crawler to simulate user login. After each crawl completed, the crawled HTML pages were parsed and the extracted content was stored in json format in the noSQL Neo4j[3] graph database. The data has the following structure: *Author Name, Title of Post, Body of Post, Commentator Name, Comment Body*. Based on this structure, we constructed graphs to model relationships between authors and commentators.

### B. Forums

We scraped *five* forums of different languages (four in english and one in spanish) in the dark web; the scraped data concerns the period from *September* since *December 2021*. Each of the scraped forums has subforums based on the various topics of interest, ranging from cybercrime and trafficking of drugs to personal experiences and politics. The forums' names are anonymised to prevent indirect publicity, by assigning a distinct label to each. Table I gives an overview of the quantitative and qualitative characteristics of the scraped forums. As seen in Table I, Forum A is the largest one, having one order of magnidute more nodes and connections compared to the rest of the forums; Forums D and E are the smallest ones mainly in terms of node connections, having less that the 20% of the edges appearing in the immediately larger forum; Forums B and C are quite similar in terms of number of nodes and edges. In more detail:

**Forum A** is the largest scraped and one of the most famous underground forums in the deep web. Our dataset contains 26 subforums, which feature topics on trafficking and usage of drugs, cybercrime, carding, fraud, and cryptocurrency.

[1]https://scrapy.org
[2]https://beautiful-soup-4.readthedocs.io/en/latest/
[3]https://neo4j.com

**Forum B** is the Spanish edition of a former underground forum based on free speech and general discussion. Our dataset contains 12 subforums featuring drugs, discussions about programming, personal issues, health, and politics.
**Forum C** is solely a hacking and programming forum; new users must take up succesfully some hacking challenges thrown by the administrators in order to have access to free tutorials. Our dataset contains six subforums under this one.
**Forum D** is a marketplace containing ten subforums, where users, while performing their buying online, can also have discussions about drug reviews, drug vendors, and anonymity/security issues.
**Forum E** is also a marketplace of drugs; it contains 11 subforums and offers the possibility to its members to discuss on drugs and security.

## V. RESULTS

In what follows, we apply the SNA metrics (Section III) to identify the social structures, as well as whether there are nodes of importance in each of the above forums. Our findings confirm and extend the results in [17] by also analysing the subforums resulted by the different discussion topics.

### A. Forum A

Starting with the largest forum, we calculated the degree centrality to highlight influential nodes. The maximum values of degree centrality for all subforums of Forum A are presented in Figure 1. There, we identify in one of the subforums (labeled *Dread*) a highly-connected node (with degree centrality equal to 350), while in the rest subforums the maximum degree centrality varies from 15 to 100. Knowing also that the average values of degree centrality for all subforums are close to 2.5-3, we can infer that only a small group of users are influential and famous, while the largest amount of members are casual users with little to no connections.

Next, to detect the nodes that are best placed to spread information and influence the entire network most quickly, we calculated the closeness centrality for Forum A. As shown in Figure 2, the average values of closeness centrality for each subforum of Forum A are all below 0.3. Note also that the maximum value calculated in Forum A for closeness centrality is 1. These scores indicate that, in each subforum, the largest amount of central actors were not that close to each other, resulting to a sparse network with low density.

The eigenvector centrality of Forum A appears quite the same with its degree centrality, although it also takes into

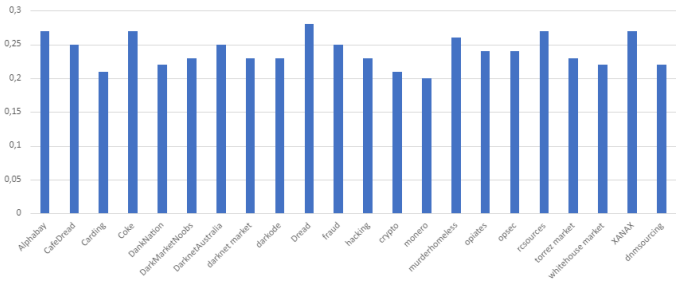


Fig. 2. Forum A: Average closeness centrality per subforum

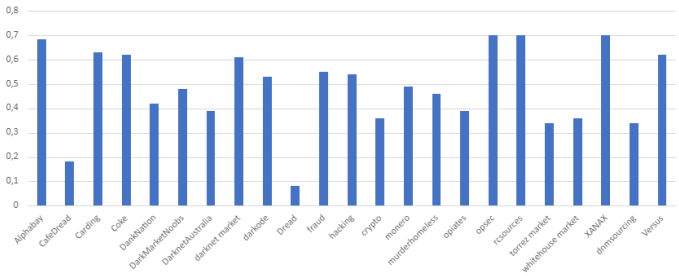


Fig. 3. Forum A: Maximum eigenvector centrality per subforum

account the connections of nodes' neighbours. In Figure 3, we demonstrate the maximum values of the eigenvector centrality calculated for each subforum of Forum A; in most subforums the eigenvector centrality reached up to 0.7, which means that most of the nodes were connected to high-scoring nodes.

Finally, calculating the betweenness centrality for Forum A, we get the results presented in Figure 4. As it is shown there, the highest scores were detected in subforums relative to drugs and dark markets, which means that in those subforums network structures are more dense and users tend to propagate information faster.

We also used a metric based on Louvain method (i.e., an algorithm to detect communities in large networks) that maximises a *modularity* score for each community, where the modularity quantifies the quality of an assignment of nodes to communities [23]. By using it, we can evaluate how densely connected the nodes within a community are, compared to how connected they would be in a random network. Modularity is measured in a scale between $-0.5$ (non-modular clustering) and 1 (fully modular clustering). Using this method, the number of communities (left-axis) and the modularity score (right-axis) calculated for each subforum of Forum A are presented in Figure 5. The results indicate that those subforums that presented the highest scores of betweenness centrality have also high modularity scores; most communities were detected in subforums labeled Dread, Fraud, Dark Market.

### B. Forum B

Concerning Forum B (that is the Spanish forum), the maximum centrality scores are presented in Table II and the average values are shown in Table III. The scores obtained for maximum degree centrality for all subforums indicate that

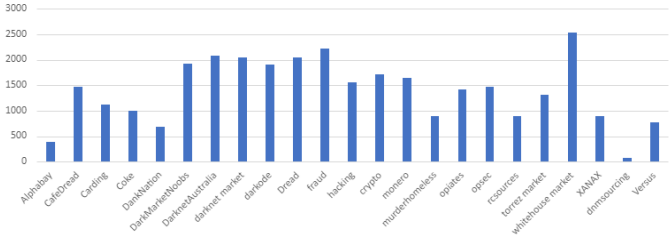


Fig. 4. Forum A: Average betweenness centrality per subforum

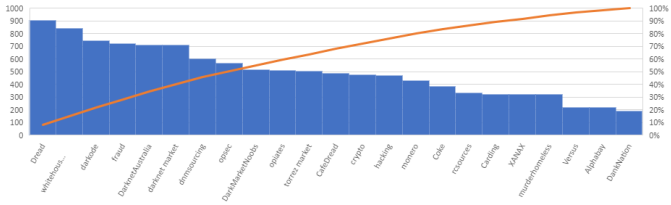


Fig. 5. Forum A: Number of communities and modularity score per subforum

users are not highly connected. However, the high scores achieved for the maximun closeness centrality mean that few links must be traversed to get from some key node to other nodes in the network, highlighting that there are nodes that can cause a rapid flow of information. Similarly, the average centrality scores indicate that the network is sparse and only a handful of users hold a central role.

### C. Forum C

The results for Forum C, that is a hacking and programming forum, are presented in Tables IV and V. While average centralities indicate sparse networks mostly populated by casual users, the maximum centralities scores point out the existence of highly-connected users in all subforums, who may spread information and influence the entire network quickly.

### D. Forum D

For Forum D, the results are displayed in Table VI for maximum and Table VII for average centrality scores. This forum is quite different from the previous ones since its main focus is buying and selling rather that discussing, resulting in low connectivity among its users as it is made obvious by the small number of edges (Table I). The average degree centralities indicate that most of the users have barely one connection to other users in the network, while also users holding a more central role appear low connected. These observations suit also our initial intuition, since a marketplace is focused towards transactions and not in social connections among its users. Although notice, that average closeness centrality scores are higher when compared to the previous forums; this is explained by the small path lengths between the nodes of this network. Also, notice that one of the subforums, labeled Introduction, has high values of betweenness centrality; most probably in this subforum new users interact with the administrators (or older users) to get answers to questions,

TABLE II
FORUM B: MAXIMUM CENTRALITIES PER SUBFORUM

| Forum B | | | | |
|---|---|---|---|---|
| Subforums | *Degree* | *Closeness* | *Eigenvector* | *Betweenness* |
| *Crypto* | 11 | 0.63 | 0.52 | 65 |
| *Darknet* | 21 | 0.6 | 0.445 | 792 |
| *Erotic Discussion* | 20 | 0.57 | 0.35 | 402 |
| *Hacking* | 15 | 0.6 | 0.37 | 102 |
| *Health* | 42 | 0.64 | 0.38 | 536 |
| *Knowledge Information* | 42 | 1 | 0.45 | 2709 |
| *Money* | 17 | 0.59 | 0.44 | 375 |
| *Other* | 43 | 0.67 | 0.4 | 3704 |
| *Programming* | 11 | 0.588 | 0.49 | 57 |
| *SadTimes* | 20 | 0.59 | 0.45 | 398 |
| *Technology* | 24 | 0.58 | 0.48 | 509 |
| *WTF* | 29 | 0.59 | 0.43 | 690 |

TABLE III
FORUM B: AVERAGE CENTRALITIES PER SUBFORUM

| Forum B | | | | |
|---|---|---|---|---|
| Subforums | *Degree* | *Closeness* | *Eigenvector* | *Betweenness* |
| *Crypto* | 2.16 | 0.38 | 0.043 | 1.89 |
| *Darknet* | 2.7 | 0.35 | 0.04 | 24.5 |
| *Erotic Discussion* | 2.3 | 0.37 | 0.078 | 14.81 |
| *Hacking* | 2.21 | 0.35 | 0.04 | 3.85 |
| *Health* | 2.35 | 0.38 | 0.075 | 12.25 |
| *Knowledge Information* | 3.24 | 0.39 | 0.036 | 77 |
| *Money* | 2.51 | 0.35 | 0.049 | 10.20 |
| *Other* | 5.26 | 0.40 | 0.035 | 127 |
| *Programming* | 1.9 | 0.36 | 0.068 | 2.15 |
| *SadTimes* | 2.87 | 0.38 | 0.073 | 24.98 |
| *Technology* | 2.35 | 0.37 | 0.06 | 11.11 |
| *WTF* | 3.14 | 0.37 | 0.04 | 30.6 |

understand the functionality and structure of the marketplace, or gain access to transactions.

### *E. Forum E*

Forum E is also a marketplace, presenting similar characteristics with Forum D (see Table I). As also in the previous marketplace, in Forum E we encountered low values of betweenness centrality, which imply poor social ties between users. Detailed results are ommitted due to space constraints.

### *F. Discussion*

Through SNA metrics, we are given the opportunity to have an overview of user roles within five dark web forums of different characteristics. On the one hand, we have the fairly large Forum A, which, due to its size, has very low density. In this forum, we found out that there is one or more central nodes per subforum (Figures 1 and 3). Depending on the type of subforum, whether it is a drug market or a hacking subforum, these key users can either be administrators, publishers, professionals, or users, for example responsible for integrating new members in the community and building more communities in the network. The large variation in degree centralities (Figure 1) per subforum indicate that there is the possibility of disseminating information and resources between dissimilar users, for example new users versus older and more experienced ones; this condition may affect behaviours and formulate ideas for future users. On the other hand, Forums D and E are more concentrated around cryptocurrencies, with less emphasis on interacting with new users (Tables VI and VII). Additionally, Forum C, because of its focus on cybercrime and hacking, enables new users through trials and eventually provides them with access to important manuals and learning tools; through this process there is enough interaction among users and rapid dissemination of information in the relevant subforums (Tables IV and V). Finally, in Spanish Forum B, which was easily accessible and proponent of free speech, we found that it has the highest density (Tables II and III). This is explained by the fact this forum has minimal rules concerning posts and does not support hierarchy of users, meaning that any user can equally have a central or more peripheral role in the network.

TABLE IV
FORUM C: MAXIMUM CENTRALITIES PER SUBFORUM

| Forum C | | | | |
|---|---|---|---|---|
| Subforums | *Degree* | *Closeness* | *Eigenvector* | *Betweenness* |
| *Hacking* | 54 | 1 | 0.39 | 99894 |
| *Programming* | 13 | 1 | 0.54 | 5050 |
| *Discussions on darknet* | 21 | 1 | 0.38 | 17494 |
| *Newbies* | 29 | 1 | 0.404 | 47047 |
| *Sell* | 21 | 1 | 0.35 | 3857 |
| *Support* | 8 | 1 | 0.508 | 2113 |

TABLE V
FORUM C: AVERAGE CENTRALITIES PER SUBFORUM

| Forum C | | | | |
|---|---|---|---|---|
| Subforums | *Degree* | *Closeness* | *Eigenvector* | *Betweenness* |
| *Hacking* | 2.13 | 0.269 | 0.0062 | 840 |
| *Programming* | 1.46 | 0.228 | 0.015 | 479 |
| *Discussions on darknet* | 1.77 | 0.253 | 0.013 | 557 |
| *Newbies* | 1.74 | 0.255 | 0.011 | 774 |
| *Sell* | 1.35 | 0.27 | 0.040 | 206 |
| *Support* | 1.33 | 0.325 | 0.032 | 110 |

In an attempt to compare the patterns/trends present in surface social network communities (e.g., [24]) vs those emerging from dark web communities, we notice that (1) only a small subset of users are influential (although much fewer in dark web forums) and (2) few links must be traversed to get from some node (key node in dark web forums) to other nodes in the network (small-world effect) in both communities. However, in dark web networks, in contrast to social networks, (1) most of the users have little to no connections, (2) the network is sparse, (3) the distribution of connections is not homogeneous, and (4) less emphasis is put on node interactions.

## VI. CONCLUSION

We scraped, analysed and compared five dark web forums to obtain information about their organisational structures. Our study focused on a static snapshot of the network, concerning the period from September since December 2021. We constructed five directed graphs to model the connections between the members of the forums, studied and compared

TABLE VI
FORUM D: MAXIMUM CENTRALITIES PER SUBFORUM

| Forum D | | | | |
|---|---|---|---|---|
| Subforums | *Degree* | *Closeness* | *Eigenvector* | *Betweenness* |
| *Anonymity* | 7 | 1 | 0.48 | 14 |
| *Carding* | 11 | 1 | 0.59 | 10 |
| *General Discussion* | 8 | 1 | 0.499 | 103 |
| *Tutorials* | 10 | 1 | 0.31 | 3.6 |
| *Hacked Data* | 14 | 1 | 0.4 | 11 |
| *Hacking Tools* | 8 | 1 | 0.34 | 6 |
| *Hacking Tutorials* | 7 | 1 | 0.37 | 15 |
| *Introduction* | 9 | 1 | 0.74 | 2929 |
| *Paid hacking services* | 4 | 1 | 0.4 | 2 |
| *Programming* | 5 | 1 | 0.37 | 15 |

TABLE VII
FORUM D: AVERAGE CENTRALITIES PER SUBFORUM

| Forum D | | | | |
|---|---|---|---|---|
| Subforums | *Degree* | *Closeness* | *Eigenvector* | *Betweenness* |
| *Anonymity* | 0.91 | 0.41 | 0.067 | 1.83 |
| *Carding* | 0.86 | 0.37 | 0.022 | 0.25 |
| *General Discussion* | 0.79 | 0.31 | 0.017 | 1.68 |
| *Tutorials* | 0.64 | 0.5 | 0.126 | 0.17 |
| *Hacked Data* | 0.83 | 0.3 | 0.022 | 0.22 |
| *Hacking Tools* | 0.71 | 0.39 | 0.058 | 0.12 |
| *Hacking Tutorials* | 0.67 | 0.72 | 0.11 | 0.79 |
| *Introduction* | 0.7 | 0.28 | 0.0068 | 58 |
| *Paid hacking services* | 0.71 | 0.33 | 0.033 | 0.061 |
| *Programming* | 0.54 | 0.52 | 0.085 | 1.32 |

these networks to identify nodes of importance, understand the social connections and interactions, and highlight the differences among forums of different characteristics.

This work can be further extended by analysing an evolving network of interactions and the developping network patterns instead of a network snapshot. This way, we could better understand over time the users' behaviour, better evaluate users relationships, as well as identify how users hierarchy changes depending on their activity. Additionally, in dark web there is plenty material in German, Polish and Russian forums, which could provide us with data for evaluating and comparing social relations and interactions among different countries; for example, to understand whether the culture affects the social virtual interactions in anonymised environments. Finally, the attributes of the various node types or their special characteristics (e.g., the high percentage of nodes without outgoing links) might allow a fine-grained analysis.

## REFERENCES


[1] A. Bermudez-Villalva and G. Stringhini, "The shady economy: Understanding the difference in trading activity from underground forums in different layers of the web," in *Proceedings of the APWG Symposium on Electronic Crime Research (eCrime)*, 2021, pp. 1–10.

[2] N. Tavabi, N. Bartley, A. Abeliuk, S. Soni, E. Ferrara, and K. Lerman, "Characterizing activity on the deep and dark web," in *Proceedings of the World Wide Web Conference (WWW)*, May 2019, pp. 206—-213.

[3] M. L. Morgia, A. Mei, S. Raponi, and J. Stefa, "Time-zone geolocation of crowds in the dark web," in *Proceedings of the IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, 2018, pp. 445–455.

[4] R. Rawat, A. S. Rajawat, V. Mahor, R. N. Shaw, and A. Ghosh, "Dark web—onion hidden service discovery and crawling for profiling morphing, unstructured crime and vulnerabilities prediction," *Innovations in Electrical and Electronic Engineering. Lecture Notes in Electrical Engineering*, vol. 756, May 2021.

[5] M. Schäfer, M. Fuchs, M. Strohmeier, M. Engel, M. Liechti, and V. Lenders, "Blackwidow: Monitoring the dark web for cyber security information," in *Proceedings of the 11th International Conference on Cyber Conflict (CyCon)*, vol. 900, 2019, pp. 1–21.

[6] P. Koloveas, T. Chantzios, C. Tryfonopoulos, and S. Skiadopoulos, "A crawler architecture for harvesting the clear, social, and dark web for iot-related cyber-threat intelligence," in *Proceedings of the IEEE World Congress on Services (SERVICES)*, vol. 2642-939X, 2019, pp. 3–8.

[7] J. K. Saini and D. Bansal, "A comparative study and automated detection of illegal weapon procurement over dark web," *Cybernetics and Systems*, vol. 50, no. 5, pp. 405—-416, January 2019.

[8] J. Xu and H. Chen, "The topology of dark networks," *Communications of the ACM*, vol. 51, no. 10, pp. 58–65, 2008.

[9] H. Chen, *Dark Web: Exploring and Data Mining the Dark Side of the Web*, S.-V. N. York, Ed., 2012.

[10] S. Samtani and Y. Chai, "Linking exploits from the dark web to known vulnerabilities for proactive cyber threat intelligence: An attention-based deep learning deep structured semantic model," *MIS Quarterly*, vol. 46, no. 2, pp. 911—-946, June 2022.

[11] P. Koloveas, T. Chantzios, S. Alevizopoulou, S. Skiadopoulos, and C. Tryfonopoulos, "inTIME: A Machine Learning-Based Framework for Gathering and Leveraging Web Data to Cyber-Threat Intelligence," *Electronics*, vol. 10, no. 7, 2021.

[12] S. Alevizopoulou, P. Koloveas, C. Tryfonopoulos, and P. Raftopoulou, "Social Media Monitoring for IoT Cyber-Threats," in *Proceedings of the International Workshop on Data Science for Cyber Security (DS4CS @ IEEE CSR)*, 2021.

[13] S. Sarkar, M. Almukaynizi, J. Shakarian, and P. Shakarian, "Predicting enterprise cyber incidents using social network analysis on dark web hacker forums," in *The Cyber Defense Review SPECIAL EDITION: International Conference on Cyber Conflict (CYCON* , November 2019, pp. 87–102.

[14] E. Phillips, J. R. Nurse, M. Goldsmith, and S. Creese, "Extracting social structure from darkweb forums," in *Proceedings of the 5th International Conference on Social Media Technologies, Communication, and Informatics (OTICS)*, November 2015, pp. 97–102.

[15] B. Monk, J. Mitchell, R. Frank, and G. Davies, "Uncovering tor: An examination of the network structure," *Security and Communication Networks*, 2018.

[16] V. Griffith, Y. Xu, and C. Ratti, "Graph theoretic properties of the darkweb," *CoRR*, 2017.

[17] I. Pete, J. Hughes, Y. T. Chua, and M. Bada, "A social network analysis and comparison of six dark web forums," in *Proceedings of the IEEE European Symposium on Security and Privacy Workshops (EuroS PW)*, 2020, pp. 484–493.

[18] M. Zamani, F. Rabbani, A. Horicsányi, and A. Z. andTamas Vicsek, "Differences in structure and dynamics of networks retrieved from dark and public web forums," *Physica A: Statistical Mechanics and its Applications*, vol. 525, pp. 326—-336, July 2019.

[19] M. D. Domenico and A. Arenas, "Modeling structure and resilience of the dark network," *Physical Review E*, vol. 95, February 2017.

[20] S. A. Ríos, F. Aguilera, J. D. Nuñez-Gonzalez, and M. Graña, "Semantically enhanced network analysis for influencer identification in online social networks," *Neurocomputing*, vol. 326—327, pp. 71–81, January 2019.

[21] M. Burcher and C. Whelan, "Social network analysis as a tool for criminal intelligence: understanding its potential from the perspectives of intelligence analysts," *Trends in Organized Crime*, vol. 21, pp. 278—-294, May 2018.

[22] A. Disney, "Social network analysis 101: centrality measures explained," https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/, January 2020.

[23] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, Feb 2004.

[24] D. Linarakis, S. Vlachos, and P. Raftopoulou, "From digital footprints to facts: Mining marketing policies of the Greek community on Instagram and Youtube," in *Proceedings of the 11th International Conference on Data Science, Technology and Applications (DATA)*, 2022.