Computers, Materials & Continua
DOI:10.32604/cmc.2021.016887
Article

Tech Science Press

A Link Analysis Algorithm for Identification of Key Hidden Services

Abdullah Alharbi[1], Mohd Faizan[2], Wael Alosaimi[1], Hashem Alyami[3], Mohd Nadeem[2] Suhel Ahmad Khan[4], Alka Agrawal[2] and Raees Ahmad Khan[2,*]

[1]Department of Information Technology, College of Computers and Information Technology, Taif University,
P. O. Box 11099, Taif, 21944, Saudi Arabia
[2]Department of Information Technology, Babasaheb Bhimrao Ambedkar University, Lucknow, 226025, India
[3]Department of Computer Science, College of Computers and Information Technology, Taif University,
P. O. Box 11099, Taif, 21944, Saudi Arabia
[4]Department of Computer Science, Indira Gandhi National Tribal University, Amarkantak, 84886, India
[*]Corresponding Author: Raees Ahmad Khan. Email: khanraees@yahoo.com

Received: 12 January 2021; Accepted: 14 February 2021

Abstract: The Tor dark web network has been reported to provide a breeding ground for criminals and fraudsters who are exploiting the vulnerabilities in the network to carry out illicit and unethical activities. The network has unfortunately become a means to perpetuate crimes like illegal drugs and firearm trafficking, violence and terrorist activities among others. The government and law enforcement agencies are working relentlessly to control the misuse of Tor network. This is a study in the similar league, with an attempt to suggest a link-based ranking technique to rank and identify the influential hidden services in the Tor dark web. The proposed method considers the extent of connectivity to the surface web services and values of the centrality metrics of a hidden service in the web graph for ranking. The modified PageRank algorithm is used to obtain the overall rankings of the hidden services in the dataset. Several graph metrics were used to evaluate the effectiveness of the proposed technique with other commonly known ranking procedures in literature. The proposed ranking technique is shown to produce good results in identifying the influential domains in the tor network.

Keywords: Dark web; Tor hidden services; ranking; link analysis

1 Introduction

The rapid growth of Internet technology and its outreach to the general public has made it easier to conduct a range of services. Along with this, the ever-growing technology has also been used to perform illegal and criminal activities. The dark web is a prime example of the abuse of Internet technology where unethical and illegal acts are predominantly present. The sale of forged documents, counterfeits, illegal drugs and firearm trafficking are some of the grey activities that are carried out on the dark web [1]. The *Onion Router (Tor), Freenet and I2P* are the methods to access the dark web with the Tor being the most commonly used method [2]. These methods ensure the anonymity of the users while surfing the dark web. The websites or the web services on the Tor dark web network are commonly referred to as the *hidden services*.

The profound level of privacy and anonymity provided by the Tor dark web has increased the illegal activities on these anonymous platforms. The law enforcement agencies have been continuously monitoring the hidden services to track the criminals. Among all the hidden services in the Tor network, some of the influential or key hidden services that exclusively offer illegal products and services usually have a significant user base. Such hidden services need immediate and dedicated efforts by the law enforcement agencies to track them down. The identification of such hidden services could prove beneficial and to the initiatives being undertaken by the law enforcement agencies in monitoring the illegal activities over the Tor dark web network.

Most of the studies on the Tor dark web network explore the attributes and working of the hidden services by scraping the data using a web crawler followed by its analysis to get insight into the hidden services and possible conclusions. However, the data collected could also be used to identify the prominent hidden services among others by using the graph theory and link analysis algorithms. The influential domain in a Tor network is the one on which the users end after surfing the network moving from one hidden service to the other. In this paper, we present a hyperlink based algorithm for ranking the Tor hidden services to identify the influential services in the Tor network.

The government and law enforcement agencies could focus their efforts on the top-ranked hidden services for further surveillance and monitoring. Though the top-ranked websites identified by the proposed ranking algorithm represent the key hidden services, the other hidden services should not be left off, instead, the concerned authorities should devise strategies to appropriately divide their efforts in targeting the illicit content on the Tor dark web. The proposed ranking algorithm is shown to be performing better than the other link analysis algorithm for the dark web. The performance of the proposed ranking algorithm is evaluated on the metrics like the graph density, clustering coefficient, average shortest path length, giant component and diameter of the equivalent graph representing the Tor network. The dataset for the experiment was collected with the help of the *customized Python crawler*.

The rest of the paper is divided as follows: Section 2 reviews the existing work on the link analysis algorithms. Section 3 describes the proposed ranking methodology. Section 4 provides the experimental setup followed by results and discussion in Section 5. Section 6 provides the conclusion of the work.

2 Materials and Methods

2.1 Pertinent Related Work

The Tor dark web has experienced a considerable increase in the number of hidden services after the rise of the Silk Road marketplace that has reported the business of more than USD 150 million [3]. Many of the studies on the Tor network have focused on the characteristics and working of the hidden services offering illicit products [4–7]. One of the preliminary studies on the hidden services was conducted on the Silk Road marketplace [8]. The author of the study meticulously collected the data over a period followed by its analysis. The data showed that drugs and psychoactive substances were the most popular choices among the users. Other studies have tried to find the operational style of the product vendors, the geographical distribution of customers and vendors, business environment on the Tor network, quality of product and services and advertising strategies [9–11]. Additionally, the dark web forums have also been studied to get customers' feedback and reviews about the products and services based on their shopping experience [12].

Al-Nabki et al. [13] leveraged the data collected by their customized crawler from the Tor domains to propose a novel hyperlink analysis algorithm called ToRank for ranking the Tor hidden services. The authors presented the dataset of Tor domains called DUTA-10K and implemented the ToRank algorithm on the dataset to evaluate the effectiveness of the proposed ranking algorithm. The performance of the ToRank algorithm was compared with the benchmark link analysis algorithm: PageRank, HITS, and Katz. The ToRank outperformed the benchmark algorithm with a significant margin on several graph robustness metrics like the graph density, clustering coefficient, etc. The rankings of the hidden services generated by the ToRank algorithm could be utilized to detect the influential domains.

In another study, a content-based approach was proposed for ranking the Tor hidden services [14]. However, the proposed approach specifically targeted illegal drug trafficking on the dark web cryptomarkets. The commonly available drugs on the Tor hidden services were assigned a harm score based on several criteria. A metric was proposed to calculate the overall harm score of a hidden service using the individual score of the drug substance present on that hidden service. The overall harm score reflects the amount of threat posed by the hidden service. The hidden services were then ranked based on their harm score to detect the most dangerous service among others. The efficacy of the proposed ranking approach was evaluated on the ground truth data by using the well-known ranking metrics. The content-based approach could easily detect even the isolated nodes (*hidden services with no incoming or outgoing hyperlinks)* that would otherwise have been skipped by the link-based approach.

While only limited work has been performed to rank the Tor content for identifying key services, plenty of work has been done on the same ranking task for the surface web content. Both the link-based and content-based approaches were adopted including the hybrid approach for ranking the surface web services [15,16]. Many studies in this context have proposed approaches for detecting influential users on social network platforms. A study has analyzed the distribution of followers of the users on the micro-blogging platform Twitter to identify the most influential user. The influence of the user was estimated by constructing the graph representing the follower's network and then applying the PageRank algorithm [15]. In another study, an algorithm was proposed for identifying the influence of nodes in the micro-blogging network. The algorithm considers the out-degree of the nodes and their neighbors to calculate the influence of the node. The identification of the influential nodes in the network could help in appropriately putting the advertisement in the network [16]. A model was proposed to estimate the variation in the user's influence by using the *continuous-time Markov process*. The nodes in the model represent the user, while the edges join the users having an interest in common topics [17].

In yet another study, several graph metrics like *degree centrality, betweenness* and *closeness centrality* were employed to track money laundering crimes [18]. A modified PageRank algorithm called the User Rank algorithm was proposed to detect influential users in a network through the analysis of the message content and user's attentiveness [19]. The combination of content and link-based approaches could be employed to obtain a hybrid approach for identifying influential nodes in a network. A hybrid approach was proposed to identify radical users in the dark web discussion forums [20]. The authors developed a metric to calculate the radical nature of a user in the forum. The radical score was incorporated into a modified PageRank algorithm to obtain a ranked list of radical users.

The ToRank algorithm only considers the hyperlinks between the hidden services to rank them and to detect the influential domains. However, it has been found that most of the Tor hidden services also have connectivity to the surface web [21]. The rise in the number of hyperlinks to the surface websites increases the chance of data theft and information leakage through the malicious hidden services [21]. Therefore, the out-going links to the surface websites may indicate the influential nature of the hidden services in the Tor network. Therefore, identifying such hidden services could lead to a disruption in the network by law enforcement agencies. Hence, in our approach of ranking, we have also considered the significance of the out-going links with the surface web.

2.2 Methodology

The proposed ranking approach requires the generation of a web graph corresponding to the Tor hidden services in the dataset. The ranking algorithm is then applied on the graph to obtain a list of the hidden services in a ranking order.

2.2.1 Construction of Web Graph

The web graph is created by representing each of the hidden services in the dataset as the nodes or vertices of the dataset. The hyperlinks between the two hidden services are represented by a directed edge between their corresponding nodes in the graph. All the self-loops and parallel edges were removed from the graph. The out-going hyperlinks to the surface websites were recorded for each of the nodes. In case of several hyperlinks from a node to the different web pages of a surface website, only one hyperlink was considered for that node. A hyperlink is discarded if it points to a hidden service that does not exist in the dataset. The web graph thus obtained is used by the algorithm for ranking.

2.2.2 Ranking Procedure

The proposed ranking approach consists of three different components to measure the influence of nodes in the graph. The definition of the various symbols is given in Tab. 1.

Table 1: Definition of the symbols used

Symbol	Definition
$V = \{v_1, v_2, v_3, \ldots\ldots v_n\}$	Set of nodes in the Tor web graph
$E = \{e_1, e_2, e_3, \ldots\ldots e_m\}$	Set of edges in the Tor web graph
$surf\,(v_i)$	Number of surface web hyperlinks of a node v_i
$deg\,(v_i)$	Degree centrality of a node v_i
$btw\,(v_i)$	Betweenness centrality of a node v_i
$cls\,(v_i)$	Closeness centrality of a node v_i
$r\,(v_i)$	The overall influence of the node v_i

2.2.3 Influence of Surface Web Hyperlinks

In this paper, we have considered the significance of the out-going surface web links to the importance of a hidden service in the network. The hyperlinks to surface websites may indicate that the hidden service is willing to advertise its product and services over the open web where a relatively large number of users are present. The influence of the surface web hyperlinks of a node v_i is τ(v_i) which is defined as the ratio of the number of surface web hyperlinks to the degree centrality of v_i. τ(v_i) is given by Eq. (1), note that one is added in the denominator to avoid the indeterminate form. In the case of isolated nodes, the τ(v_i) will be the total number of out-going hyperlinks to the surface web.

$$\tau\left(v_i\right) = \frac{surf\left(v_i\right)}{deg\left(v_i\right) + 1} \tag{1}$$

2.2.4 *Influence of Node Connectivity*

The influence of the node connectivity is defined by the significance of the location of the node in the Tor web graph. A user arriving at a node with a good location in the network would have a better probability of moving to other regions of the network through the neighboring nodes, a neighbor of the neighboring nodes and so on until the user reaches the desired node. Therefore, such nodes hold an influential position in the Tor ecosystem and identification of the same may be fruitful in terms of law enforcement perspective.

The centrality metrics used in the graph theory indicates the relative significance of a node in the graph. The different centrality measures reflect the importance of the node's location in the graph. The centrality value can indicate the ability of the node to bypass a large number of users because it has several paths passing through them [22]. Closeness, degree and betweenness centrality are three metrics for obtaining the centrality of the node.

In the Tor web graph, a node with a high degree centrality would facilitate the movement of a higher number of users through it than the node with a low degree centrality. If the node has a greater closeness centrality value, it would better control the movement of the user and allows the rapid dispersal of the users. The greater value of the betweenness centrality of the node would enable a quicker movement of the users to the entire web graph through a few intermediate nodes. The influence of the connectivity of the node v_i in the network is represented by μ(v_i) and is defined by Eq. (2).

$$\mu\left(v_i\right) = deg\left(v_i\right) + cls\left(v_i\right) + btw\left(v_i\right) \tag{2}$$

2.2.5 *Calculating Overall Influence*

The ranking metric presented here is used to assign an influence score to a node based on the influence value of the node's surface web links and its connectivity in the graph. The influence metric that measures the overall influence of the node differs from the centrality metrics in the sense that it considers the ability of the node in facilitating the movement of the users through the entire Tor network. Each of the nodes in the graph has a different ability to propagate the users and hence have a different influence score. The influence score of the node v_(i) is given by δ(v_i) and is defined by Eq. (3).

$$\delta\left(v_i\right) = \tau\left(v_i\right) + \mu\left(v_i\right) \tag{3}$$

In the Tor network, each of the hidden services has a different influence and significance as compared to the other services. Thus the procedure of identifying the key hidden services is equivalent to the ranking problem where the top-ranked service would be the most influential one among others. The influential nature of a hidden service is governed by other hidden services to which the former has out-going hyperlinks along with its characteristics. If a hidden service has outgoing connections with the highly influential hidden services, then it would positively contribute to its influence. The PageRank algorithm is appropriate in such scenarios for incorporating the influence of the neighboring nodes [23].

The proposed ranking algorithm is based on a modified PageRank algorithm to detect the influential hidden services from the Tor web graph. Each of the nodes in the Tor web graph is assigned an initial value reflecting the influence score and is updated iteratively according to the Eq. (4) until the convergence is achieved.

$$r\left(v_i\right) = (1-a) + a \sum_{v_j \in Q_i} \log\{r\left(v_j\right) * \delta\left(v_i\right) + 1\} \tag{4}$$

In Eq. (4), a ∈ [0, 1] is the damping factor and is set to 0.85 [24], Q_i is the set of all nodes having an incoming hyperlink from the node v_(i). The logarithm of the product of r(v_j) and δ(v_i) is used to obtain the cumulative influence of v_(i) and its immediate neighbors. Moreover, one is added to the product to avoid getting zero in the argument (when δ(v_i) is zero) for which the log function is undefined.

The final rank of the node v_i is governed by two factors: A.) the influence score δ(v_i), and B.) r(v_j) which represents the effect of connections to other nodes. If Q_i = {∅} for a node v_i, then r (v_i) will be equal to 0.15. Thus the node v_i is influential only if it provides the gateway for the users to move to other nodes in the network through it. The removal of a node with no outgoing connectivity would not cause much disruption in the Tor network, also the removal of only a single hidden service has proved to be a costly and time-consuming operation [25]. Therefore, the proposed ranking approach would identify the influential nodes whose elimination would cause much disruption in the Tor network.

2.3 Experimental Setup

2.3.1 Dataset

The dataset for the study was created using a customized Python web crawler for scraping the Tor hidden services. The web crawler was designed to connect to the hidden services using the SOCKS proxy [25]. Initially, the web crawler was provided with a small list of onion domains called *seeds* from the publicly available Tor directories [26,27]. The crawler connects with each of the seeds, upon successful connection the crawler scrapes the hidden service and searches for the new onion domains. Once all the seeds have been explored by the crawler, the newly found domains from the initial seeds were saved for subsequent operations. The above procedure is repeated on the newly discovered links. Finally, the crawler was able to found 4041 active hidden services, the content of the home page of each of the active hidden services including the HTML tags was stored in the individual files.

For our ranking procedure, the content of each of the hidden services needs to be parsed to extract only the hyperlinks and discard other textual content. A parser based on regular expressions was utilized to extract the hyperlinks to the surface web and Tor dark web. The hyperlinks to other dark web networks, internet relay chat addresses and sub-domains were eliminated.

2.3.2 Evaluation Metrics

The performance of the proposed ranking algorithm is compared with that of the PageRank [28] and ToRank [13] algorithm. In line with the earlier studies [28–31], several graph metrics that evaluate the graph structure shall be used to test the effectiveness of our proposed ranking algorithm. Graph density indicates the connectedness of the graph, a high graph density value refers to the strong connectivity of the graph. The graph density is given by Eq. (5).

$$GD = \frac{e}{n\left(n-1\right)} \tag{5}$$

The graph density curve could be used to measure the robustness of the graph [29]. The top-ranked nodes and the associated edges are eliminated from the graph one by one and the graph density is calculated at every removal. The removal process is terminated when the density becomes zero. Following the method of Al-Nabki et al. [13], the ranking algorithm that covers the smallest area under the graph density curve shall be the most effective in identifying the influential nodes in the Tor network graph. If the ranking algorithm has correctly ranked the influential nodes, the removal of the top-ranked nodes would result in a dramatic reduction in the graph density. This is because the influential nodes hold the key location in the network with good connectivity and their removal should disintegrate the graph.

The clustering coefficient [29] and the average shortest path length [30] could also be used to measure the robustness of the graph structure. A decrease in the clustering coefficient after removing the top-ranked nodes indicates a significant breakdown in the graph structure [31]. On the other hand, an increase in the average shortest path upon removal of the top-ranked nodes reflects the good ranking of the nodes [29]. The removal of top-ranked nodes one by one is a costly affair due to the high computation complexity of the clustering coefficient at every step, hence the clustering coefficient is calculated only at the removal of the top 1st, 5th, 10th and 20th percentile as in earlier study [13].

2.3.3 Parameter Settings

The Python NetworkX library was used to construct the graph from the corresponding hidden services in the dataset. The proposed ranking algorithm iteratively computes the score of each of the nodes until convergence is achieved. The algorithm is supposed to achieve convergence when the error between the score of a node in the current iteration and the preceding iteration is less than 0.0001. The parameter value of the PageRank and ToRank were set according to as in previous work [13]. All the experiments were performed on a machine running on Intel i5 CPU with 4 GB of RAM and Windows 8.1 operating system.

3 Data Interpretations and Results

The Tor web graph generated from the dataset consists of 4041 nodes representing the corresponding hidden services and 14059 edges. The influential score returned by the ranking algorithm is used to rank the nodes in descending order. The top-ranked nodes were repeatedly removed and the graph density was calculated at every iteration. The graph density curve of the PageRank, ToRank and the proposed ranking algorithm are shown in Fig. 1. The proposed ranking algorithm obtains the smallest area under the graph density curve which is 0.013. The ToRank algorithm achieves the second smallest area of 0.015 which is very close to the proposed algorithm. On the other hand, the PageRank algorithm covers a relatively large area under the curve and is the largest among the three approaches.

The performance of the proposed ranking algorithm on the other graphs' robustness is shown in Tab. 2. The values of the metrics are shown after the removal of top-ranked nodes at different percentiles and the value of metrics for the full web graph with all the nodes and edges intact is shown. The proposed algorithm produces a significant decrease in the clustering coefficient of the graph. The average shortest path also experienced an increase in its length after the removal of the top-ranked nodes. Moreover, the diameter of the graph has increased to 26 after the removal of the top 20% influential nodes. The results indicate the better performance of the proposed algorithm over the other two algorithms.

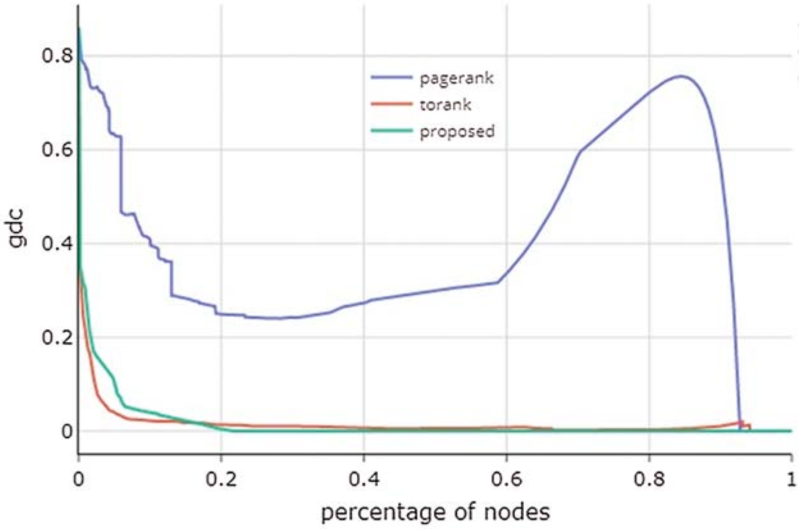

Figure 1: Graph density curves of the three algorithms

Table 2: Comparison of the proposed ranking technique with the other algorithms

Algorithms	Nodes removal (%)	Clustering coefficient	Average shortest path	Giant component	Diameter
Full graph		0.199	4.37	4761	12
PageRank	Top 1	0.288	4.42	4689	12
	Top 5	0.248	4.49	4521	12
	Top 10	0.221	4.53	3873	13
	Top 20	0.07	4.55	3325	13
ToRank	Top 1	0.041	4.85	2997	17
	Top 5	0.02	5.43	2030	20
	Top 10	0.011	6.07	752	23
	Top 20	0.006	7.12	131	25
Proposed	Top 1	0.051	4.83	3160	19
	Top 5	0.018	5.78	1606	23
	Top 10	0.007	6.24	537	25
	Top 20	0.0	7.69	10	26

The proposed ranking algorithm is based on an iterative calculation of the influence score depending on the connectivity of the node with its immediate neighbors. Initially, a small influence score is assigned to each of the nodes, once the algorithm achieves convergence, the node that has connections to the other influential nodes in the network will have a much better score than the other. However, being a link analysis algorithm, it cannot evaluate the influence of isolated nodes with zero outgoing and incoming hyperlinks.

4 Conclusions

An iterative algorithm is proposed in this work to calculate the influence of a hidden service in the Tor network. The influence of a hidden service depends upon its location and connectivity in the Tor network as well as its connectivity to the surface web. Moreover, the influential nature of the hidden service is also determined by its connection to the other services in the network. All these factors are incorporated into an algorithm to determine the overall influence of a hidden service. The proposed algorithm is implemented on a dataset of Tor hidden services and is compared with other link analysis algorithm from the existing literature. The proposed algorithm achieves better performance than the other algorithms measured in terms of various graphs' robustness metrics.

Acknowledgement: This research was supported by Taif University Researchers Supporting Project Number (TURSP-2020/231), Taif University, Taif, Saudi Arabia.

Funding Statement: Funding for this study was granted by Taif University Research Supporting Projects at Taif University, Kingdom of Saudi Arabia under the Grant No. TURSP-2020/231.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. Faizan and R. A. Khan, "Exploring and analyzing the dark web: A new alchemy," *First Monday*, vol. 24, no. 5, pp. 1–20, 2019.

[2] R. Graham and B. Pitman, "Freedom in the wilderness: A study of a darknet space," *Convergence*, vol. 2, no. 8, pp. 1–14, 2018.

[3] K. Kruithof, J. Aldridge, D. D. Hétu, M. Sim, E. Dujso *et al.,* "Internet facilitated drugs trade," *RAND Corporation*, vol. 2, no. 8, pp. 21–32, 2016.

[4] D. V. Gouwe, T. M. Brunt, M. V. Laar and P. V. D. Pol, "Purity, adulteration and price of drugs bought on-line versus off-line in the Netherlands," *Addiction*, vol. 112, no. 4, pp. 640–648, 2017.

[5] D. Rhumorbarbe, L. Staehli, J. Broséus, Q. Rossy and P. Esseiva, "Buying drugs on a darknet market: A better deal? Studying the online illicit drug market through the analysis of digital, physical and chemical data," *Expert Systems with Applications*, vol. 267, no. 8, pp. 173–182, 2016.

[6] A. Bancroft and P. S. Reid, "Concepts of illicit drug quality among darknet market users: purity, embodied experience, craft and chemical knowledge," *International Journal of Drug Policy*, vol. 35, no. 8, pp. 42–49, 2016.

[7] J. Martin, J. Cunliffe, D. D. Hétu and J. Aldridge, "The international darknet drugs trade-a regional analysis of cryptomarkets," *Australasian Policing*, vol. 10, no. 3, pp. 25–29, 2018.

[8] N. Christin, "Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace," in *Proc. of the 22nd Int. Conf. on World Wide Web*, Rio de Janeiro, Brazil, pp. 213–224, 2013.

[9] I. Ladegaard, "Instantly hooked? freebies and samples of opioids, cannabis, MDMA, and other drugs in an illicit e-commerce market," *Journal of Drug Issues*, vol. 48, no. 2, pp. 226–245, 2018.

[10] M. P. Clouston, D. D. Hetu and C. Morselli, "Assessing market competition and vendors size and scope on alpha bay," *International Journal of Drug Policy*, vol. 54, no. 8, pp. 87–98, 2018.

[11] E. Wadsworth, C. Drummond and P. Deluca, "The dynamic environment of cryptomarkets: The lifespan of new psychoactive substances (NPS) and vendors selling NPS," *Expert Systems with Applications*, vol. 8, no. 3, pp. 46–47, 2018.

[12] A. Bancroft and P. S. Reid, "Concepts of illicit drug quality among darknet market users: Purity, embodied experience, craft and chemical knowledge," *International Journal of Drug Policy*, vol. 35, no. 8, pp. 42–49, 2016.

[13] M. W. Al-Nabki, E. Fidalgo, E. Alegre and L. F. Robles, "ToRank: Identifying the most influential suspicious domains in the Tor network," *Expert Systems with Applications*, Article in Press, pp. 14–27, 2019.

[14] M. Faizan, R. A. Khan and A. Agrawal, “Ranking potentially harmful tor hidden services: Illicit drugs perspective,” *Applied Computing and Informatics*, Article in Press, pp. 1–9, 2020.

[15] H. Kwak, C. Lee, H. Park and S. Moon, “What is Twitter, A social network or a news media?,” in *Proc. 19th Int. World Wide Web Conf. Committee*, Raleigh, NC, USA, pp. 591–600, 2010.

[16] F. Hao, M. Chen, C. Zhu and M. Guizani, “Discovering influential users in micro-blog marketing with influence maximization mechanism,” in *Proc. IEEE Global Communication Conf.*, Anaheim, CA, USA, pp. 470–474, 2012.

[17] J. Li, P. Wei, T. Li, T. Sun, Q. Li *et al.,* “Social network user influence sense-making and dynamics prediction,” *Expert Systems with Applications*, vol. 41, no. 11, pp. 5115–5124, 2014.

[18] F. Bodendorf and C. Kaiser, “Detecting opinion leaders and trends in online social networks,” in *Proc. 2nd ACM Workshop Social Web Search Mining*, Hong Kong, China, pp. 65–68, 2009.

[19] C. C. Yang, X. Tang and B. M. Thuraisingham, “An analysis of user influence ranking algorithms on dark web forums,” in *Proc. 16th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, Washington D.C.,USA, pp. 1–7, 2010.

[20] T. Anwar and M. Abulaish, “Ranking radically influential web forum users,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 6, pp. 1289–1298, 2015.

[21] A. Java, P. Kolari, T. Finin and T. Oates, “Modeling the spread of influence on the blogosphere,” in *Proc. of the WWW Workshop*, CA, USA, pp. 14–24, 2006.

[22] J. Zhang, M. S. Ackerman and L. Adamic, “Expertise networks in online communities: Structure and algorithms,” in *Proc. 16th Int. World Wide Web Conf.*, Banff Alberta, Canada, pp. 221–230, 2007.

[23] S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine,” *Computer Networks and ISDN Systems*, vol. 30, no. 1–7, pp. 107–117, 1998.

[24] D. D. Hétu and L. Giommoni, “Do police crackdowns disrupt drug cryptomarkets? A longitudinal analysis of the effects of Operation Onymous,” *Crime Law and Social Change*, vol. 67, no. 1, pp. 55–75, 2017.

[25] A. F. Colladon and P. A. Gloor, “Measuring the impact of spammers on e-mail and twitter networks,” *International Journal of Information Management*, vol. 67, no. 1, pp. 14–26, 2018.

[26] A. F. Colladon and E. Remondi, “Using social network analysis to prevent money laundering,” *Expert Systems with Applications*, vol. 67, no. 5, pp. 49–58, 2017.

[27] Y. Wang, N. Nelissen, K. Adamczuk, A. S. D. Weer, M. Vandenbulcke *et al.,* “Reproducibility and robustness of graph measures of the associative semantic network,” *PLoS One*, vol. 9, no. 12, pp. 1–22, 2014.

[28] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small world’ networks,” *Nature*, vol. 393, no. 440, pp. 14–26, 1998.

[29] G. Mao and N. Zhang, “Fast approximation of average shortest path length of directed BA networks,” *Physica A: Statistical Mechanics and Its Applications*, vol. 466, no. 8, pp. 243–248, 2017.

[30] V. Chang, “A cybernetics social cloud,” *Journal of Systems and Software*, vol. 124, no. 4, pp. 195–211, 2017.

[31] S. Iyer, T. Killingback, B. Sundaram and Z. Wang, “Attack robustness and centrality of complex networks,” *PLoS One*, vol. 8, no. 4, pp. 1–18, 2013.