Extracting Social Structure from DarkWeb Forums

Elizabeth Phillips, Jason R.C. Nurse, Michael Goldsmith, Sadie Creese
Cyber Security Centre,
Department of Computer Science,
University of Oxford, UK
Email: $\{firstname.lastname\}@cybersecurity.ox.ac.uk$

***Abstract*—This paper explores various Social Network Analysis (SNA) techniques in order to identify a range of potentially *'important'* members of Islamic Networks within Dark Web Forums. For this experiment, we conducted our investigation on five forums collected in previous work as part of the Dark Web Forum portal and built upon the tool support created in our previous research in order to visualise and analyse the network. Whilst existing work attempts to identify these structures through state-of-the-art Computational Linguistic techniques, our work relies on the communication metadata alone. Our analysis involved first calculating a range of SNA metrics to better understand the group members, and then apply unsupervised learning in order to create clusters that would help classify the Dark Web Forums users into hierarchical clusters. In order to create our social networks, we investigated the effect of repeated author resolution and various weighting schemes on the ranking of forum members by creating four social networks per forum and evaluating the correlation of the top n users (for $n = 10, 20, 30, 40, 50$ and 100). Our results identified that varying the weighting schemes created more consistent ranking schemes than varying the repeated author resolution.**

Keywords–Social Network Analysis; Dark Web Forum; Jihadist forums; Social Structure

I. INTRODUCTION

With the dawn of the digital revolution, the Internet has transformed the way in which people communicate with each other. One area in which this has been seen is within terrorist networks[1], [2]. The near-instantaneous responses from users now achievable with these developing technologies and the increased audience has meant that previous face-to-face interactions and radical discussions have migrated to online mediums [3], [4]. This migration has led to an increase in the popularity of Dark Web Forums as a means of sharing text based content as well as links to other sites, videos and rich Web 2.0 features [5].

Even before the increased adoption of dark web forums, researchers have used social network analysis to identify key individuals within these groups. Numerous researchers have investigated various techniques in order to retrospectively understand the inner-workings of the 19 terrorists involved in the September 11th attacks in 2001[6]. Early research involved sourcing news articles related to the known suspects after the event and analysing the structure of the group after the event and with known targets [7]. Others have focused on researching how different terrorists organisations work together [8] by performing social network analysis [9].

The widespread adoption of the Dark Web Forums have enabled researchers to analyse the rich source of data [10] and has enabled researchers to establish a greater insight into the inner workings of some of these groups. Whilst existing research has been undertaken into creating a social network topology for Dark web Forums [11], these existing research focus on small communities and networks and rely on sophisticated sentiment analysis techniques, which are difficult to scale when evaluating millions of messages[12].

Since the revelations of metadata collection exposed by Edward Snowden in 2013 [13] , the importance of metadata from emails has gained awareness. In light of these revelations, many individuals have been unsure of their own risk exposure using these metadata techniques alone. Existing work has been effective at establishing hierarchy from the metadata of email communications alone [14], [15]. Such metadata include the post's author and timestamp. In this paper, we set out to investigate whether similar techniques could be applied within Dark Web Forums and assess to what extent we can identify the social structure of the networks.

Our paper is focused on the research question "Using communication metadata alone, can we make reasonable inferences about the structure of terrorist groups". In particular, we set out to trial various Social Network Analysis (SNA) techniques in order to identify a range of potentially *'important'* members of the forums.

A. Social Network Analysis (SNA)

For decades, complex interactivity between entities has been modelled as networks. These include the internet [16], food webs and biochemical networks [17]. For each network, entities (such as computers, routers, animals, etc.) are considered as nodes or vertices that are connected together by links or edges (such as communications between computers, the flow of energy within a food network, etc.)

Link Analysis (LA) is the analysis of information flow within the networks above and has been a topic of study for several decades [18], [19]. A *Social Network* (SN) is defined as the representation of communication networks where the nodes are people and the edges correspond to the relationships between and *Social Network Analysis* (SNA) is defined as the application of Link Analysis to a social network. We can perform SNA on our newly created social network, where the flow of information corresponds to the flow of information on the forums which allows us to perform SNA on our network to determine hidden network structures.

II. METHODOLOGY

In order to assess the extent of the network discovery of our forums, we begin by collecting the forum posts. This in turn

 ISBN: 978-1-61208-443-5

allowed us to extract the metadata from each forum from which to build our network. Once we had extracted the metadata mentioned above, the next step was to create a social network representation of each forum. We then set out to experiment ways in which we create a social network representation of the communication data [see section II-A]. Figure 1 shows an overview of our process.

Figure 1. An outline of our approach to extract social structure from online communications

From this social network, we calculated the top-ranked SNA metrics from previous work [14]. After calculating the metrics for each node within the network, we then used unsupervised machine learning to identify clusters of interest within our network. Unsupervised machine learning was selected due to the unavailability of ground truth of the seniority of the forum members (other than identifying the top posters). Given the nature of our metrics chosen, the clusters will be grouped based on the similarity of their metrics, which in turn we hypothesise reflects the importance of the individual within the network.

A. Creating the Social Network

Once the metadata for each forum was collected, we then set out to experiment the ways in which to convert the metadata into a Social Network. We convert the metadata into a social network using algorithm 1.

Algorithm 1 Algorithm to convert a forum into a social network where w(a, s, m) is the weighted score of the message m from sender a to sender s and is determined based on our design decisions below and SNW(s_1, s_2) is the overall weight of the directed edge between s_1 and s_2 in our Social Network for a given forum f.

```
for each thread t in forum f do
    for each message m in t do
        a = author(m)
        for each previous sender s in thread t do
            SNW(a, s) = SNW(a, s) + w(a, s, m)
        end for
    end for
end for
```

For this evaluation we experimented with two distinct methods of handling when an author posted multiple times in the same thread (Repeated Author Resolution) and two distinct ways of increasing the weight between two members based on the Time difference between responses.

In order to illustrate the application of our methods, Figure 2 shows an example of a forum with two members and four posts.

1) **Repeated Author Resolution:** In Figure 2, our forum posts shows two posts for $sender1$ and two posts for $sender2$. Our experiment tested 2 ways of handling repeated authors within a thread.

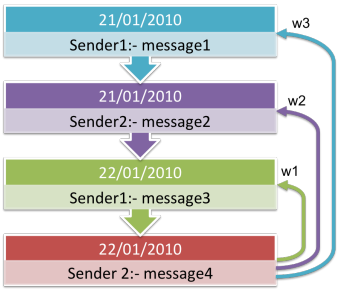

Figure 2. An example forum post with two members and four posts

Unique Senders: For this method, we only consider those senders that are unique within each thread. In this case, if we see an author reply to a thread they have previously commented on, we only add the weight for the most recent comment on the thread. In our example in Figure 2 when updating the scores after message4, only w_1 and w_2 would be added to SNA$(sender2, sender1)$ and SNA$(sender1, sender1)$ respectively and w_3 would not be calculated or added.

All Senders: For this method, we consider every sender within each thread and include duplicate senders. In this case, if we see an author reply to a thread they have previously commented on, we add the weight for each comment on the thread. In our example in figure 2 when updating the scores after $message4$, SNW$(sender2, sender1)$ is increased by $w_1 + w_3$ and SNW$(sender1, sender1)$ would be increased by w_2.

2) **Weighted Date Resolution:** When a new author a contributes a message m_a to a thread t where senders $s_1, s_2, \ldots s_k$ have previously commented, let d_{s_i} be the date that sender s_i sends a message m_{s_i} in the thread t. We then need to calculate w(a, s_i, m_a) for all i. In addition to evaluating the effectiveness of including all or unique senders, we also evaluated two distinct ways to calculate the weightings based on the time difference (d_t) between the two messages. The two methods we used to calculate the weightings were *'uniform weighting'* and *'inverse proportionality weighting'*.

Uniform weighting: Our first model assumed that the strength of the connection between two messages within a thread is independent on the amount of time taken to respond to a message. In this case, $\forall i$, w$(a, s_i, m_a) = 1$. Whilst this allows for a simple view of our forum, we set out to explore whether this model may not be representative as the same weight is given to a response to a message on the same day as a message with a response delay of 4 weeks.

Inverse Proportionality weighting: In order to overcome the issue outlined above, our second model was created on the assumption that the closer a response is in time, the stronger the weight of the connection between two users. In this case $\forall i$, w$(a, s_i, m_a) = \frac{1}{1+(d_a - d_{s_i})}$. This model ensures that $\forall i, 0 <$ w$(a, s_i, m_a) < 1$.

B. SNA Metrics

For the purpose of this paper, in order to measure the graph properties of each node that reflect importance, we used the

 ISBN: 978-1-61208-443-5

TABLE I. A TABLE OUTLINING THE METRICS USED TO EVALUATE OUR SOCIAL NETWORK

Attribute Name	**Description**
Sent Messages (SM)	The number of emails sent by an employee.
Received Messages (RM)	The number of emails received by an employee.
Degree Centrality (DCS)	The number of distinct employees within the network that an employee has sent emails to.
Betweenness Centrality Score (BCS)	The betweenness centrality measure for an employee.[20]
Pagerank Score (PRS)	The PageRank score an employee. [21]
Markov Ranking (MR)	The markov ranking of an employee. [22]
HITS Authority Score (HAS)	The authority score for an employee (if several users with high hub weights send an email t the user then they will have a higher authority score). [23]
HITS Hub Score (HHS)	The hub score for an employee (if the user sends emails to users with high authority scores then they will have a higher hub score). [23]
Clique Score (CS)	The number of cliques (maximal subgraphs) an employee is in using the Bron and Kerbosch algorithm.[24]
Weighted Clique Score (WCS)	The weighted clique score for each user, weighted by the number of users within each clique.
Average Distance Score (ADS)	The average distance between the user and all other users in the graph.
Clustering Coefficient (CC)	The extent to which vertices in a graph tend to cluster together. [25]

SNA metrics outlined in [14] and are shown in Table I. These metrics can be split into five main categories that can be used to identify relevant properties of our network. These categories are highlighted in Figure 3. By selecting a cross-section of metrics that cover all five main categories of metrics, our machine learning model is able to capture as much information from our network as possible.

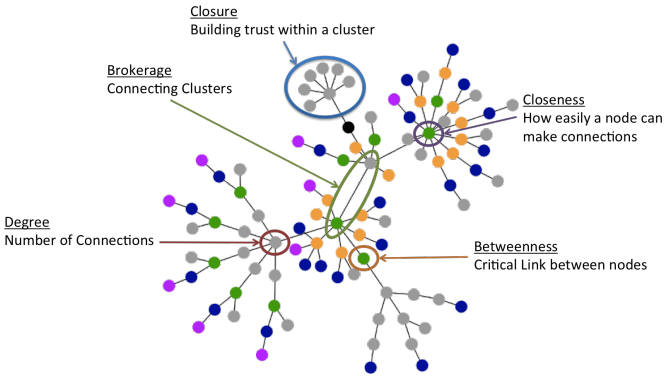

Figure 3. An outline of the five main categories of SNA metrics, namely brokerage, degree, closeness, closure and betweenness.

III. EXPERIMENTAL SETUP

A. Datasets

For this experiment we evaluated 10 of the forums collected as part of the Dark Web Forum Portal dataset[26]. This dataset was collected by crawling a variety of radical websites [27] and allows us to compare the results and its applicability over multiple forums. Table II shows a breakdown of each forum used as part of the experiment.

In order to ensure that the connections in our graph reflect meaningful connections within the network, members whose edge weights are less than 10 are pruned from our graph. This allows the graph to reflect the strong connections whilst removing those connections that do not play a central role within the network. As such, given the different weighting approaches for each scheme, after pruning edges with edge weight less than 10, each social network contained varying numbers of vertices with 1 or more edges.

For each of our seven terrorist network datasets, we created four social networks using the combination of metrics below. Once the network was created, we then calculated our SNA metrics and performed unsupervised machine learning.

- Uniform weighting and unique senders
- Uniform weighting and all senders
- Inverse-Proportionality weighting and unique senders
- Inverse-Proportionality and all senders

In order to evaluate the appropriateness of each of our methods, we performed Unsupervised Learning using Expectation Maximisation (EM) [28] in order to create clusters in order to classify forum members. EM was chosen as it is able to handle unobserved data and missing datapoints, which can reflect the fact that we may only have a partial view of the network by observing the online forums alone.

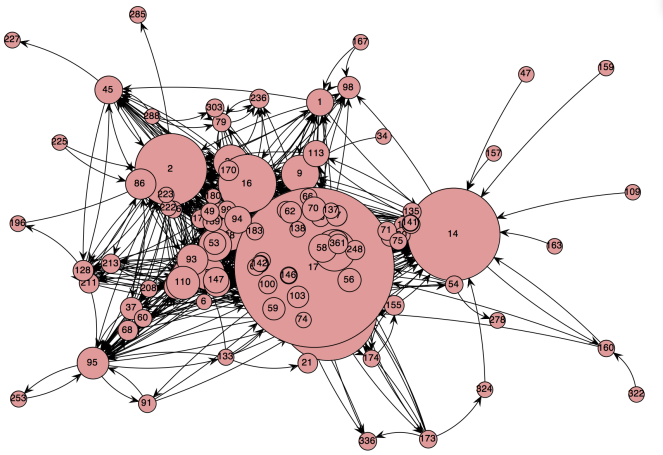

Figure 4. An example screenshot of our toolsuppport where each node is sized by the node's HITS hub score.

B. Tool Support

To allow us to visualise the results of our social network analysis and easily switch between the different weighting schemes and repeated author resolution, we built upon our existing tool [14] by adding additional features to allow users to provide their own weighting scheme to create their own Social Network Representation of the underlying forum data.

Figure 4 shows an example screenshot of our tool support displaying the social Network of the Ansar1 Network using repeated authors and inverse proportionality weighting.

 ISBN: 978-1-61208-443-5

TABLE II. AN OVERVIEW OF THE ENGLISH LANGUAGE FORUMS USED AS PART OF OUR EXPERIMENTS

Forum	#Messages:	# Threads:	# Members:	Start Date:	End Date:	Forum URL:
Ansar AlJihad Network (Ansar1)	29492	11244	382	12/08/08	01/20/2010	http://wwww.ansar1.info/
Gawaher (Gawaher)	372499	53235	9269	10/24/2004	06/07/12	http://www.gawaher.com
Islamic Awakening (IslamicAwakening)	201287	32879	3964	04/28/2004	05/22/2012	http://forums.islamicawakening.com
Islamic Network (IslamicNetwork)	91874	13995	2082	06/09/04	11/10/10	http://talk.islamicnetwork.com
Islamic Web-Community (Myiwc)	25016	6310	756	11/05/00	02/19/2010	http://www.myiwc.com/forums/index.php
Turn To Islam (TurnToIslam)	335338	41654	10858	06/02/06	05/20/2012	http://www.turnintonislam.com/forum/
Ummah	1491957	91527	21013	04/01/02	05/18/2012	http://www.ummah.com/forum/

IV. RESULTS AND ANALYSIS

In this section we outline some of the main highlights of our research findings. For our Ansar1 Dataset, we identified five clusters. $Cluster0$ contained 26 members, $cluster1$ contained 37 members, $cluster2$ contained 1 member, $cluster3$ contained 5 members and $cluster4$ contained 42 members. Table III provides further detail on the distribution of each cluster identified by the EM model.

TABLE III. OUTLINE OF OUR CLUSTERING ALGORITHMS FOR OUR ANSAR1 DATASET USING A UNIFORMED WEIGHTING SCHEME AND INCLUDING ALL SENDERS.

Attribute	Cluster				
	0	1	2	3	4
degree					
mean	0.0953	0.0095	0.8559	0.3568	0.0293
std.dev	0.0419	0.0021	0.1104	0.0944	0.0124
betweenness					
mean	30.5428	0.0535	5549.8267	551.7925	0.0002
std.dev	42.8349	0.3254	541.1433	337.2082	0.013
PageRank					
mean	0.0105	0.0031	0.1488	0.0464	0.0053
std.dev	0.005	0.0007	0.017	0.0223	0.002
Markov					
mean	0.0113	0.002	0.1513	0.0538	0.0048
std.dev	0.0064	0.0009	0.0185	0.0258	0.0027
HITS_authority					
mean	0.0977	0.0168	0.4142	0.2655	0.0477
std.dev	0.0451	0.0118	0.0715	0.0601	0.0273
HITS_hub					
mean	0.1085	0.0149	0.4118	0.2502	0.0445
std.dev	0.0423	0.0126	0.0712	0.0506	0.0291
weighted_cliques					
mean	538.4462	2.107	5108	3875.3263	17.6769
std.dev	666.9511	0.458	997.1299	669.2559	28.3311
cliques					
mean	5.968	1.0532	110	42.0026	1
std.dev	4.5005	0.2258	13.7192	14.3096	13.7192
avgDistance					
mean	1.5016	1.4613	1.8868	1.5895	1.474
std.dev	0.027	0.0414	0.0586	0.0342	0.0306
clusteringCoeff					
mean	0.6634	0	0.0589	0.257	1
std.dev	0.1631	0.0029	0.4387	0.0935	0.4387

Our evaluator identified 5 main clusters ranging in size from 1 to 42 users. From our results, we can immediately identify that cluster 2 (with 1 user) has a significantly greater score on almost all metrics compared to the other clusters followed by cluster 3. This leads us to believe that user 17 (the sole user in cluster 2) has significantly higher influence within the network, compared followed by those members in cluster 3 and as such, these users are potential candidates for being "important" users within the Ansar1 network. The significantly lower standard deviation for clusters 1 and 4 indicate that these users are likely to be less important in the network with their low scores in the ranking metrics (e.g. PageRank, HITS Hub, etc.).

We also evaluated the difference in ranking of each created social network and Table IV shows the top ten ranked users

TABLE IV. THE TOP TEN USERS USING ALL FOUR SOCIAL NETWORK METRICS SCHEME USING OUR ANSAR1 DATASET RANKED USING THE PAGERANK ALGORITHM

Inverse-1Date	Inverse-AllDates	Uniform-AllDates	Uniform-1Date
17	17	17	17
14	14	0	0
0	0	14	14
16	2	2	2
2	16	16	16
11	39	39	39
9	11	11	11
39	9	12	9
33	33	33	33

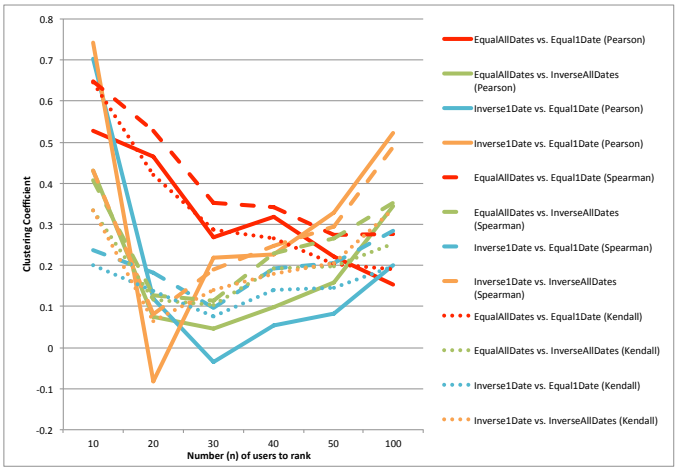

Figure 5. An outline of the correlation coefficients for each of our four networks using Spearman's, Kendall's and Pearson's correlation coefficients.

(using the PageRank algorithm) for each of the four networks using the Ansar1 dataset. The table shows that all four networks identified user 17 as the most influential user and agreed on the top three users (but had different orderings of users 0 and 14). Similarly, they all agree on the top 10 users (with the exception of the UniformAllDates network, which substitutes user 9 for user 12.

In order to evaluate the effect that weighting and repeated author resolution has on our model, we evaluated the consistency of our rankings with schemes by comparing the top n ranked users (for $n = 10, 20, 30, 40, 50$ and 100) using Spearman's Coefficient Ranking [29], Kendall's Tau Measure [30] and Pearson's Correlation Coefficient [31]. Table V and Figure 5 shows the ranking scores of each comparison with only one degree of freedom (either identical weighting and different repeated author resolution or vice versa). Our results showed that on average, varying the repeated author resolution caused less variation in the ranking of the metrics in the final social network when applied for $n < 50$. However, the difference between rankings based on their repeated author

 ISBN: 978-1-61208-443-5

TABLE V. COMPARISON OF CORRELATION BETWEEN RANKING ALGORITHMS USING OUR ANSAR1 DATASET

Attribute	Number of users (n) to compare ranking using PageRank																	
	10			20			30			40			50			100		
	P	S	K	P	S	K	P	S	K	P	S	K	P	S	K	P	S	K
UniformAllDates vs.Uniform1Date	0.529	0.648	0.644	0.466	0.528	0.421	0.270	0.354	0.287	0.318	0.343	0.267	0.221	0.274	0.203	0.152	0.276	0.191
UniformAllDates vs. InverseAllDates	0.431	0.406	0.333	0.074	0.128	0.116	0.045	0.113	0.103	0.099	0.231	0.192	0.157	0.265	0.198	0.346	0.352	0.253
Inverse1Date vs. Uniform1Date	0.704	0.236	0.200	0.119	0.182	0.137	-0.035	0.097	0.076	0.054	0.192	0.141	0.084	0.205	0.146	0.202	0.285	0.197
Inverse1Date vs. InverseAllDates	0.741	0.430	0.333	-0.083	0.081	0.063	0.218	0.191	0.140	0.227	0.247	0.179	0.328	0.294	0.207	0.521	0.489	0.338

resolution diminishes as n increases. This leads us to believe that performing two social networks, one using repeated authors and one using unique authors only may provide subtly different views of the graph, which in turn will allow us to gain more insight from the social network.

V. CONCLUSIONS AND FUTURE WORK

In our paper, we set out to investigate the effectiveness of using metadata alone to identify influential and important users on Dark Web Forums. We set out to investigate the effect of repeated author resolution and various weighting schemes on our rankings by creating four social networks per forum and evaluating the consistency of the top n users (for $n = 10, 20, 30, 40, 50, 100$). We also performed unsupervised machine learning for each of network in order to identify clusters by user's importance. Our results showed us that difference in rankings from different weighting schemes were more consistent on average than those using different repeated author resolution techniques.

In order to allow us to further evaluate the validity of our work, we hope to establish an authoritative ground truth in order to assess the relative performance of our work. Another direction we hope to take the research is to perform dynamic analysis on our network and assess how the network changes over time. This technique could then be applied to identify potential insider threats within the Dark Web Forums by observing abnormal dynamic behaviour. In order to allow us to further evaluate the validity of our work, we hope to establish an authoritative ground truth in order to assess the relative performance of our work. Another direction we hope to take the research is to perform dynamic analysis on our network and assess how the network changes over time. This technique could then be applied to identify potential insider threats within the Dark Web Forums by observing abnormal dynamic behaviour.

ACKNOWLEDGMENT

This work has been made possible through the EPSRC Scholarship as part of the Oxford University Centre for Doctoral Training in Cyber Security.

REFERENCES

[1] K. Crilley, "Information warfare: new battle fields terrorists, propaganda and the internet," in Aslib Proceedings, vol. 53, no. 7. MCB UP Ltd, 2001, pp. 250–264.

[2] Y. Zhou, E. Reid, J. Qin, H. Chen, and G. Lai, "Us domestic extremist groups on the web: link and content analysis," Intelligent Systems, IEEE, vol. 20, no. 5, 2005, pp. 44–51.

[3] P. B. Gerstenfeld, D. R. Grant, and C.-P. Chiang, "Hate online: A content analysis of extremist internet sites," Analyses of social issues and public policy, vol. 3, no. 1, 2003, pp. 29–44.

[4] S. Mehrotra, Intelligence and Security Informatics: IEEE International Conference on Intelligence and Security Informatics, ISI 2006, San Diego, CA, USA, May 23-24, 2006. Springer Science & Business Media, 2006, vol. 3975.

[5] H. Chen, S. Thoms, and T. Fu, "Cyber extremism in web 2.0: An exploratory study of international jihadist groups," in Intelligence and Security Informatics, 2008. ISI 2008. IEEE International Conference on. IEEE, 2008, pp. 98–103.

[6] V. E. Krebs, "Mapping networks of terrorist cells," Connections, vol. 24, no. 3, 2002, pp. 43–52.

[7] V. Krebs, "Uncloaking terrorist networks," First Monday, vol. 7, no. 4, 2002.

[8] A. Basu, "Social network analysis of terrorist organizations in india," in North American Association for Computational Social and Organizational Science (NAACSOS) Conference, 2005, pp. 26–28.

[9] C. C. Yang, N. Liu, and M. Sageman, "Analyzing the terrorist social networks with visualization tools," in Intelligence and security informatics. Springer, 2006, pp. 331–342.

[10] T. Stevens, "Regulating the ?dark web?: How a two-fold approach can tackle peer-to-peer radicalisation," The RUSI Journal, vol. 154, no. 2, 2009, pp. 28–33.

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

[12] G. L'Huillier, S. A. Ríos, H. Alvarez, and F. Aguilera, "Topic-based social network analysis for virtual communities of interests in the dark web," in ACM SIGKDD Workshop on Intelligence and Security Informatics, ser. ISI-KDD '10. New York, NY, USA: ACM, 2010, pp. 9:1–9:9. [Online]. Available: http://doi.acm.org/10.1145/1938606.1938615 [Accessed: 2015-10-19]

[13] T. Guardian, "Edward snowden | world news | the guardian," 06 2015. [Online]. Available: http://www.theguardian.com/world/edward-snowden [Accessed: 2015-10-10]

[14] E. Phillips, J. R. C. Nurse, M. Goldsmith, and S. Creese, "Applying social network analysis to security," in International Conference on Cyber Security for Sustainable Society, 2015, pp. 11–27.

[15] M. Hay, G. Miklau, D. Jensen, D. Towsley, and P. Weis, "Resisting structural re-identification in anonymized social networks," Proceedings of the VLDB Endowment, vol. 1, no. 1, 2008, pp. 102–114.

[16] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," ACM SIGCOMM Computer Communication Review, vol. 29, no. 4, 1999, pp. 251–262.

[17] A. Golightly and D. J. Wilkinson, "Bayesian parameter inference for stochastic biochemical network models using particle markov chain monte carlo," Interface Focus, 2011, p. rsfs20110047.

[18] L. Getoor and C. P. Diehl, "Link mining: A survey," SIGKDD Explor. Newsl., vol. 7, no. 2, Dec. 2005, pp. 3–12. [Online]. Available: http://doi.acm.org/10.1145/1117454.1117456 [Accessed: 2015-10-18]

[19] S. Wasserman, Social Network Analysis: Methods and Applications. Cambridge University Press, Nov. 1994.

[20] L. C. Freeman, "Centrality in social networks conceptual clarification," Social networks, vol. 1, no. 3, 1979, pp. 215–239.

[21] I. Rogers, The Google Pagerank algorithm and how it works, 2002. [Online]. Available: http://mira.sai.msu.ru/~megera/docs/IR/search/pagerank/pagerank_explained.pdf [Accessed: 2015-09-15]

[22] D. Koschtzki, K. A. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, and O. Zlotowski, "Centrality Indices," in Network Analysis, ser. Lecture Notes in Computer Science, U. Brandes and T. Erlebach, Eds. Springer Berlin Heidelberg, 2005, no. 3418, pp. 16–61. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-540-31955-9_3 [Accessed: 2015-05-19]

[23] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," J. ACM, vol. 46, no. 5, Sep. 1999, pp. 604–632. [Online]. Available: http://doi.acm.org/10.1145/324133.324140 [Accessed: 2015-10-15]

[24] C. Bron and J. Kerbosch, "Algorithm 457: finding all cliques of an

 ISBN: 978-1-61208-443-5

undirected graph," Communications of the ACM, vol. 16, no. 9, 1973, pp. 575 – 577. [Online]. Available: http://dl.acm.org/citation.cfm?id=362367 [Accessed: 2015-06-15]

[25] S. N. Soffer and A. Vazquez, "Network clustering coefficient without degree-correlation biases," Physical Review Series E-, vol. 71, no. 5, 2005, p. 057101.

[26] "Dark web forum portal." [Online]. Available: http://cri-portal.dyndns.org/portal/Home.action [Accessed: 2015-10-10]

[27] H. Chen, "Dark web forum portal," in Dark Web. Springer, 2012, pp. 257–270.

[28] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," Journal of the royal statistical society. Series B (methodological), 1977, pp. 1–38.

[29] S. B. Lyerly, "The average spearman rank correlation coefficient," Psychometrika, vol. 17, no. 4, 1952, pp. 421–428.

[30] G. S. Shieh, "A weighted kendall's tau statistic," Statistics & Probability Letters, vol. 39, no. 1, 1998, pp. 17–24.

[31] P. Sedgwick et al., "Pearsons correlation coefficient," BMJ, vol. 345, 2012.

 ISBN: 978-1-61208-443-5