2016

Exploring Deviant Hacker Networks (DHN) On Social Media Platforms

Samer al-Kateeb
University of Arkansas at Little Rock

Kevin Conlan
University of New Haven

Nitin Agarwal
University of Arkansas at Little Rock

Ibrahim Baggili
University of New Haven, ibaggili@newhaven.edu

Frank Breitinger
University of New Haven, fbreitinger@newhaven.edu

Publisher Citation

Al-khateeb, Samer, Conlan, Kevin J., Agarwal, Nitin, Baggili, Ibrahim, and Breitinger, Frank. "Exploring Deviant Hacker Networks (DHN) On Social Media Platforms." Journal of Digital Forensics, Security and Law 11, no. 2 (2016): 7-20. http://jdfsl.org

Dr. Baggili was appointed to the University of New Haven's Elder Family Endowed Chair in 2015.

EXPLORING DEVIANT HACKER NETWORKS (DHN) ON SOCIAL MEDIA PLATFORMS

Samer Al-khateeb[1], Kevin J Conlan[2], Nitin Agarwal[1], Ibrahim Baggili[2], and Frank Breitinger[2]

[1]Department of Information Science
Center of Social Media and Online Behavioral Studies (COSMOS)
University of Arkansas at Little Rock, 2801 S University Ave, Little Rock, AR 72204
{sxalkhateeb,nxagarwal}@ualr.edu

[2]Cyber Forensics Research & Education Group (UNHcFREG)
Tagliatela College of Engineering, ECECS
University of New Haven, 300 Boston Post Rd, West Haven, CT 06516
{kconl1, ibaggili, fbreitinger}@newhaven.edu

ABSTRACT

Online Social Networks (OSNs) have grown exponentially over the past decade. The initial use of social media for *benign* purposes (e.g., to socialize with friends, browse pictures and photographs, and communicate with family members overseas) has now transitioned to include *malicious* activities (e.g., cybercrime, cyberterrorism, and cyberwarfare). These nefarious uses of OSNs poses a significant threat to society, and thus requires research attention. In this exploratory work, we study activities of one deviant groups: hacker groups on social media, which we term *Deviant Hacker Networks* (DHN). We investigated the connection between different DHNs on Twitter: how they are connected, identified the powerful nodes, which nodes sourced information, and which nodes act as "bridges" between different network components. From this, we were able to identify and articulate specific examples of DHNs communicating with each other, with the goal of committing some form of deviant act online. In our work, we also attempted to bridge the gap between the empirical study of OSNs and cyber forensics, as the growth of OSNs is now bringing these two domains together, due to OSNs continuously generating vast amounts of evidentiary data.

Keywords: deviant hacking groups, focal structure analysis, FSA, groups coordination, collective action, deviant groups.

1. INTRODUCTION

In a relatively short period of time, Online Social Networks (OSNs) have revolutionized how societies interact with the Internet, especially with regards to communication. While this new phenomenon in

online socialization has brought the world closer together, OSNs have also led to new vectors to facilitate cybercrime, cyberterrorism, and other deviant behaviors.

Online *deviant* groups have grown in parallel with OSNs. Groups with malicious actors, such as Islamic State of Iraq and the Levant (ISIL) typically disseminate their message of terror and recruit people using OSNs. Black hat hackers with the intent of coordinating cyber attacks may also be considered another group with malicious intent. The threat these deviant groups pose is real and can manifest in several forms of deviance, such as the disabling of critical infrastructure (e.g., the Ukraine power outage caused by Russian-sponsored hackers that coordinated a cyber attack in December 2015 [1], and ISIL recruiters who influence impressionable youth on Internet forums and social media outlets to join extremist groups and participate in heinous acts of terror (Al-khateeb & Agarwal, 2015a). Observable malicious behaviors in OSNs, similar to the aforementioned ones, continue to negatively impact society. This necessitates their scientific inquiry.

It would be of benefit to the Information Assurance (IA) domain, and its respective sub-domains, to conduct novel research on the phenomenon of deviant behavior in OSNs. One evident form of deviant behavior comes in the form of communications between black hat hacking groups on OSNs, which we term *Deviant Hacker Networks* (DHN). In this preliminary work we attempted to answer the following research questions: 1) Do DHNs employ social media platforms to coordinate attacks? If so, what are their levels of sophistication?; 2) Can their affiliations be tracked and monitored?; 3) Can their coordination strategies and communication networks be inferred from empirical observation, data set collection, and analysis? and; 4) Can the most influential set of nodes be identified?

It is apparent that these questions require inquiry since at the time of writing this paper, research on this topic was sparse. Our work resulted in the following contributions:

- We proposed a framework that can provide insights about DHNs from OSNs (e.g., the communications of nodes between and within their respective focal structures.
- We observed and analyzed the communication network of DHNs that use social media as a means to communicate and coordinate their attacks.
- We were able to identify key accounts (e.g., powerful sources of information through the measure of their out-degree centrality) and powerful coordinating DHNs using the Focal Structure Analysis (FSA) algorithm.

The rest of the article is organized as follows. Theoretical background of the research is discussed in Section 2. Section 3 presents the research methodology, including a description of the data set and software used to collect the data, then results and analysis, and finally our proposed framework. Section 4 concludes the study with possible future research directions.

2. RELATED WORK

2.1 Forensic Analysis of Social Applications

Most scientific work to date has focused on the acquisition of social data from digital devices as well as applications installed on

[1] U.S. helping Ukraine investigate power grid hack. *Reuters*. Jan. 12 2016. Available at: http://reut.rs/1PqNAYG.

them. Over time, OSNs have become the largest and fastest growing entities on the Internet, containing hundreds of millions of people, and now bots. OSNs, hosted on platforms like Facebook, LinkedIn, and Twitter, contain a plethora of data about its members, which is of interest to both digital forensic scientists and practitioners Huber et al. (2011). The forensic potential of these OSNs has been acknowledged by research, and there have been a number of studies on extracting this forensically relevant data from them.

In the seminal work by Al Mutawa, Baggili, and Marrington (2012), researchers conducted the primary forensic analysis of three popular social networking smartphone applications: Facebook, Twitter, and MySpace. Their results indicated that while no traces of forensically valuable artifacts were retrievable from Blackberry devices, iPhones and Android phones stored a significant amount of valuable data on user activities that may be recovered by forensic investigators. This data included usernames, passwords, chat messages, posts, location data, and pictures, just to name a few (Al Mutawa et al., 2012).

Other work also focused on the retrieval of artifacts from the network captures of social-messaging applications as opposed to data from the storage on these devices. Walnycky, Baggili, Marrington, Moore, and Breitinger (2015) forensically acquired and analyzed locally stored data and network traffic of 20 popular Android social-messaging applications. They were able to reconstruct partial or entire message contents from 16 of the 20 applications in their study. While this reflected poorly on the security of the applications, it could be seen positively with regards to evidence collection from a forensics standpoint. Their reconstruction and interception of passwords, pictures, videos, audio, and other forms of digital artifacts signifies the great potential for extracting evidence from the very OSNs that result from the use of these applications.

These examples of traditional digital forensic research initiatives are of relevance to our work, as evidence to the activities, communications, and motives of DHNs may be recovered from artifacts left behind by users of social networking applications - either from the device, or from the network traffic generated by the applications.

Nonetheless, this data would be based on the analysis of the more traditional sources of evidence found on systems and devices, such as file systems and captured network traffic. OSNs are continuously creating and storing data on multiple servers across the Internet through their respective social networks. These newer forms of data, especially the communications of hacker groups on OSNs, would likely pertain to, for example, coordination and planning. This would mean that traditional methods of forensic investigation would be insufficient, as this data would be real-time, constantly expanding, and simply not found in traditional sources of forensic evidence.

2.2 Forensic Retrieval of OSN-generated Evidentiary Data

As discussed, a growing challenge for forensic investigations is the emergence of OSNs and their multitude of online communication vectors, producing data with heightened volume, variety and velocity - big data at its essence. This data cannot be accessed using traditional digital forensic techniques related to storage and memory media analysis.

Baggili and Breitinger (2015) highlighted that social media is growing as a data source for cyber forensics, providing new types of artifacts that can be relevant to investigations. They identified key social media data sources (e.g., text posts, friends/groups, im-

ages, geolocation data, demographic information, videos, dates/times), as well as their corresponding applications to cyber forensics (author attribution, social network identification, facial/object recognition, personality profiling, location finding, cyber profiling, deception detection, event reconstruction, etc.).

Baggili and Breitinger (2015) further emphasized that in the future, practitioners should embrace the idea of using real-time intelligence to assist in investigations, and not just post-mortem data.

Key work in this realm was executed by Mulazzani, Huber, and Weippl (2012), where they addressed the challenge of traditional computer forensics in the growing number of cases that involve evidence being generated through cloud services and their use of distributed data centers (Mulazzani et al., 2012). In their work, they identified key data sources and analytical methods for automated forensic analysis on social network user data. This work is relevant to our research, as the hundreds of millions of OSN users online makes the forensic data extraction from them a requirement. Furthermore, Mulazzani et al. (2012) were able to present a method to evaluate these data sources in an automated fashion and generate visual graphs of interest. To note, their work was conducted without the need to collaborate with the social network operator (in their case, Facebook).

Based on prior work in this domain, it is clear that OSNs contain vast amounts of important, and often publicly accessible data that can service cyber forensics and related disciplines. A progression must thus be made towards developing and/or adopting methodologies to effectively collect and analyze evidentiary data extracted from OSNs, and leverage them in relevant domains outside of classical information sciences.

2.3 Intelligence-driven Analyses of OSNs

As OSNs continuously replace traditional means of digital storage, sharing, and communication (Huber et al., 2011), collecting this ever-growing volume of data is becoming a challenge. Within the past decade, data collected from OSNs has already played a major role as evidence in criminal cases, either as incriminating evidence or to confirm alibis [2] [3] [4].

Online deviant groups, like terrorist groups, criminal organizations, and in our specific research interest, DHNs, continue to utilize OSNs to promote, enhance, and facilitate their respective goals. It might be more efficient to take an intelligence-driven approach for identifying evidentiary trails. Harvesting forensically relevant data directly from targeted OSN user accounts, as we aim to do in our work, would be more efficient than traditional forensic techniques of analyzing the hardware, network traffic, file systems, and other traditional scenarios in digital forensics.

Interestingly, despite the growing importance of data that can be extracted from OSNs, there has been little academic research aimed at developing and enhancing techniques to effectively collect and analyze this data Huber et al. (2011). Our work aims to take steps towards bridging the gap between cyber forensics and social network analysis through a primary exploratory study that focuses on DHNs. Despite the lack of research in this domain, there have been seminal research efforts similar to our proposed work.

[2] Facebook status update provides alibi. *CNN*. 2009.

[3] Criminal Found via Facebook. *The New York Criminal Law Blog*. 2009.

[4] Facebook: a place to meet, gossip, share photos of stolen goods. *The Washington Post*. 2010.

A similar research thrust was conducted by Huber et al. (2011), where they utilized an automated web browser combined with an OSN third party application (in this case, Facebook) to gather *social snapshots* of data sets of user data and relational information of the targeted social network.

In our work, we used an algorithm developed by Şen, Wigand, Agarwal, Tokdemir, and Kasprzyk (2016) to discover *influential sets of individuals* in online social networks called Focal Structure Analysis (FSA) (Şen, Wigand, Agarwal, Mete, & Kasprzyk, 2014). This algorithm has been tested on many real world cases such as *the Saudi Arabian women's Oct26Driving campaign* on Twitter (Yuce, Agarwal, Wigand, Lim, & Robinson, 2014). Results showed that focal structures are *more interactive than average individuals* and *more interactive than communities* in the evolution of a mass protest, i.e., the interaction rate of focal structures are significantly higher than the average interaction rate of random sets of individuals, and the number of retweets, mentions, and replies increase proportionally with respect to the followers of the individuals within communities (Şen et al., 2016).

The 2014 Ukraine Crisis is another FSA application example. When a British journalist and a blogger known as Graham W. Phillips covered the 2014 Ukraine crisis and became a growing star on Kremlin-owned media (Seddon, 2014; Şen et al., 2016), FSA was applied on a blog-blog network and results illustrated that Graham Phillips was involved in the only focal structure of the entire network along with ITAR-TASS (the Russian News Agency) and Voice of Russia (the Russian government's international radio broadcasting service). Graham W. Phillips was actively involved in the crisis as a blogger, and maintained a single-authored blog with significant influence when compared to some of the active mainstream media blogs (Al-khateeb & Agarwal, Under review).

FSA has also been used in the case of *the Dragoon Ride Exercise* to discover the most influential set of botnets or the seeders of information used to disseminate the propaganda (bots that by working together profoundly impact propaganda dissemination) (Al-khateeb & Agarwal, Under review).

3. EXPLORATORY WORK

3.1 Data Collection

A seed data set was created by manually searching for hacker accounts on Twitter. The accounts chosen were selected based on their self-identifying statuses as hackers with deviant and malicious intentions. We started with a list of 49 DHNs known for their promotion of deviant activities and personas. To note, the researchers are aware that some self-proclaimed "hacker" accounts on Twitter are actually benevolent in nature, i.e., "white-hat hackers". These accounts were avoided. We were able to identify 62 Twitter accounts for these DHNs (some groups have more than one account, and some groups we could not find their Twitter handles). We crawled their Twitter network using NodeXl (Foundation, 2014) which uses the Twitter API to collect data for the period between 11/25/2008 7:03:30 PM UTC and 11/28/2015 9:20:06 PM UTC. This resulted in a directed graph that contained 58,120 unique twitter nodes and 76,964 unique edges. (Total edges with duplicates are 87,318). The edges were as follows:

- We obtained 2,837 "Tweet" edges: these are the edges created when a user tweets something. These are represented self-loops (from the user tweeting to him/herself).

- We obtained 6,148 "Mentions" edges: these are the edges created when a user mentions someone (i.e., include a preceeding '@' character with another Twitter handle) in the tweet or retweet of another user.

- We obtained 1,405 "Replies to" edges: these are a special form of the mention edges that occur when the user's name is at the very start of a tweet.

- We obtained 76,928 "Friends" and "Followers" edges: these are the edges created when a user follows or followed by other users.

The network also contained 6 isolates (one node alone not connected to any other nodes) and 1 big connected component (everyone is connected to some other nodes). The big component had 58,114 unique nodes and 87,318 edges. The network we collected also had a maximum geodesic distance ($Diameter$) = 9 which meant the network is quite large (it takes a maximum of 9 steps to get from one side of the network to the other) but also the network has an average geodesic distance = 3.8 which portrays that information is likely to reach all the nodes in the network quite quickly (Hanneman & Riddle, 2005).

3.2 Results & Discussion

In this section, we discuss the results we obtained from applying social network analysis on the collected data, then discuss the results we obtained form applying FSA (Şen et al., 2016).

3.2.1 Social Network Analysis of the Data

As mentioned earlier, we collected the Twitter network of the identified DHN accounts. By examining their overall network we were able to identify the most used hashtags (See Table 2 Appendix A) in the network (these are appropriate for discerning if a counter message were to be pushed to the same audience who follow these specific hashtags), or keywords/ bigrams (See Table 3 Appendix A) (which are useful for exploring what these groups are mainly talking about without having to read entire tweets), the most used URLs (to explore what messages these groups are trying to deliver e.g., is it a YouTube video? or a blog entry?, etc.), and the domains they used frequently (can be used as seed knowledge to be fed into forensic investigations to discover other related websites e.g., websites managed by the same Unique Identifier can be retrieved using digital forensic techniques, finding geolocations, IP Addresses, owner details, etc.).

Table 1. The number of nodes in each of the focal structures. The focal structure FSA_ID1 has the highest number of nodes. The FSA_ID is to distinguish the FSAs only.

Focal structure ID	Number of nodes in the FSA
FSA_ID1	18
FSA_ID2	3
FSA_ID3	5
FSA_ID4	2
FSA_ID5	2
FSA_ID6	9
FSA_ID7	3
FSA_ID8	16

From the primary analysis, we identified three key findings: 1) with regards to the *bigram* "tango,down", which has the largest entire graph count (173). The hacker group, Anonymous, often included the *hashtag* "Tango Down" to signify a successful DDoS attack[5]; 2) the *hashtag* "#OpNimr",

[5]Anonymous: CIA, Interpol websites 'tango down'. *Reuters*. 2012.

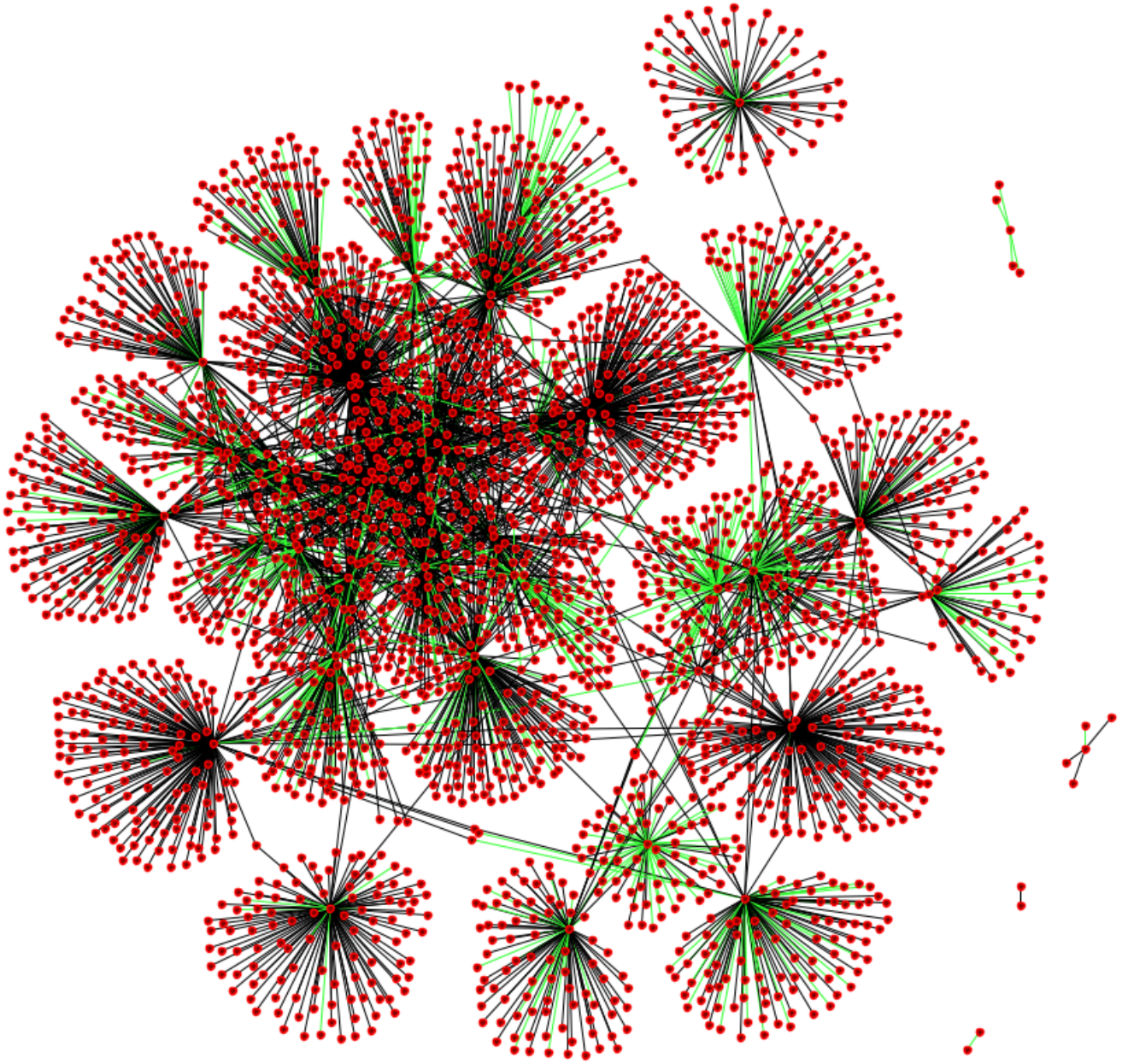

Figure 1. The Communication Network of the DHNs. Green Edges represent "Replies to" relation between the accounts. Black Edges represent "Mentions" relation between the accounts

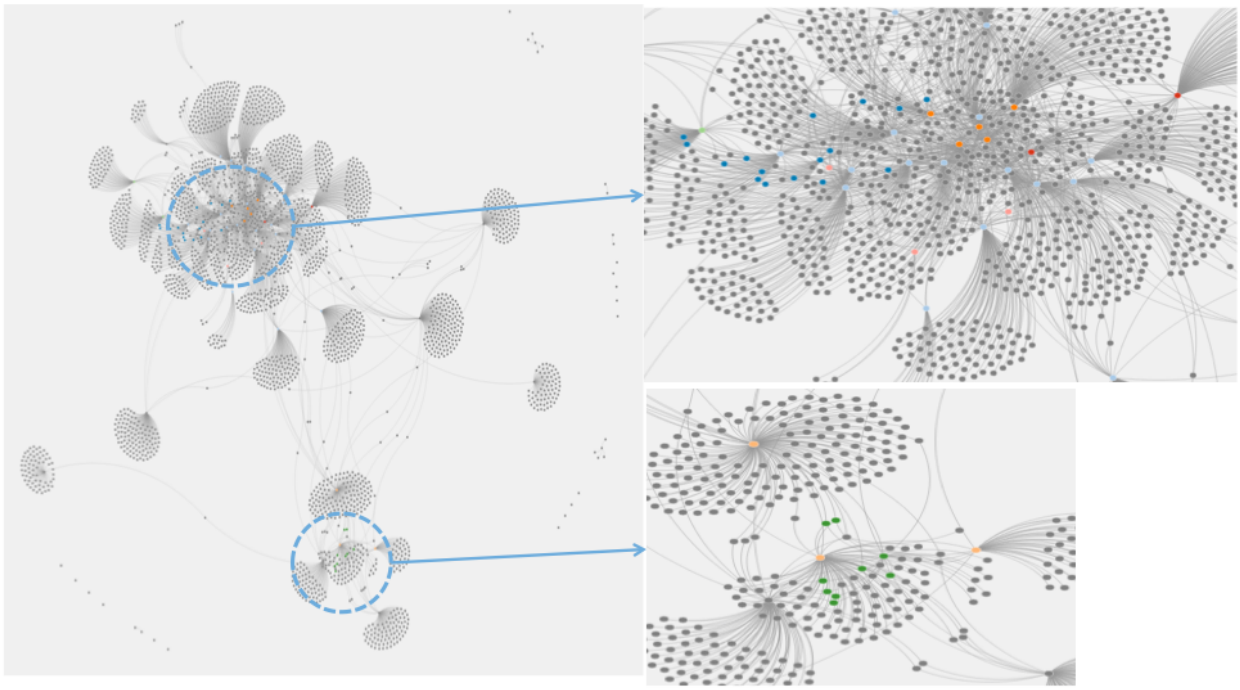

Figure 2. The communication network of the DHNs with the identified FSAs (displayed on the right). On the left side we zoom-in on the FSAs which are distinguished by different colors)

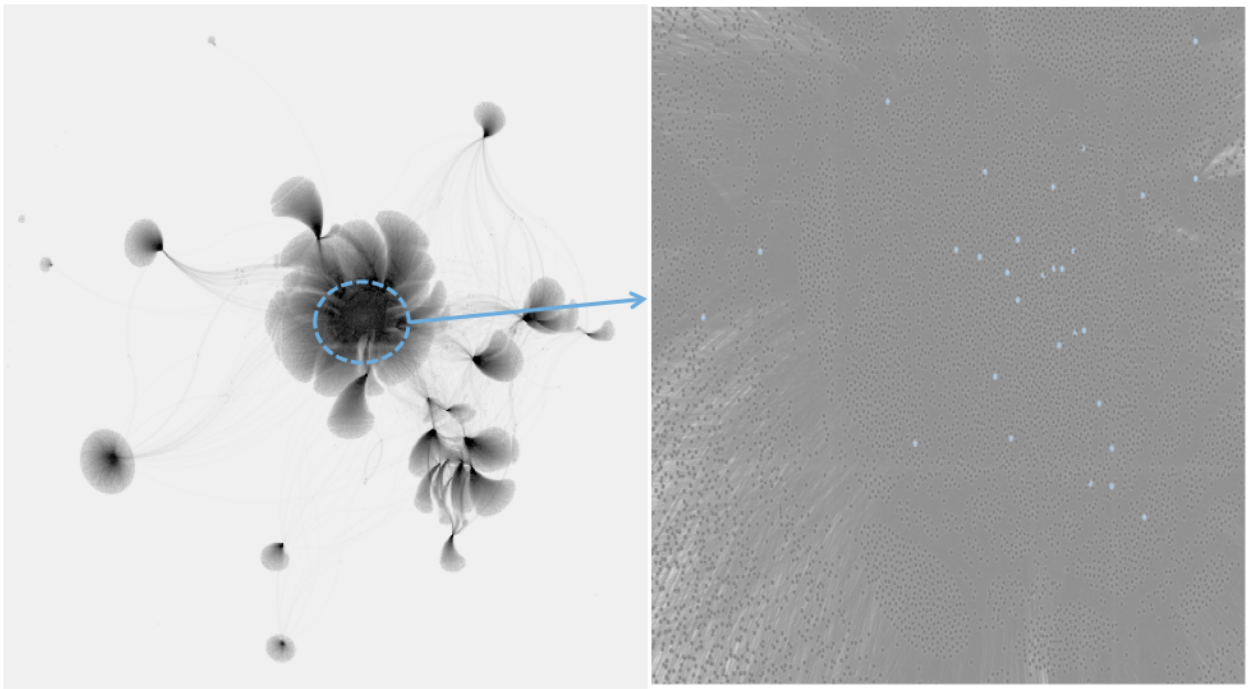

Figure 3. The Focal Structures obtained form the Social Network of the DHNs

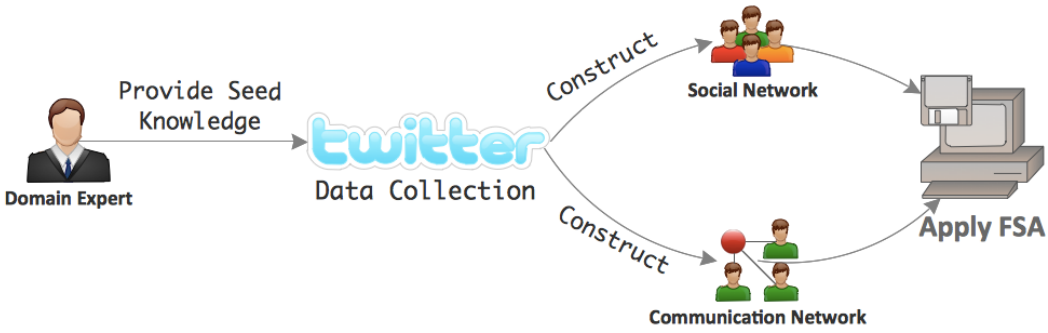

Figure 4. The Proposed Framework to Study Online Deviant Group Networks

which had the second largest entire graph count (379) referred to the threat of the hacktivist group, Anonymous, to attack the Saudi Government websites [6] as an expression of their support for a Saudi youth who was captured at the age of 17 participating in protest against the Saudi government and sentenced to death, and; 3) the hashtag "#OpBeast" [7], which had the third largest entire graph count (213) referring to the attack that was planned by Anonymous, against animal cruelty and depravity websites. Anonymous were encouraging people to tweet using this hashtag to spread awareness that they are against some of the websites that do not operate in the dark web (some .com websites) which provide services that show images of animal cruelty, feature animal depravity and bestiality. These represented key insights provided by our methodology, and warrants further exploratory and observational research into DHNs.

We then split the collected data into two (Agent X Agent) networks namely **social network** (*friends and followers*) and **communication network** (*mentions, and replies*) (See Figure 1). The communication network shown in Figure 1 contains 2,740 nodes and 3,445 edges. On the other hand, the social network contains 56,964 Nodes and 76,766 edges (not shown here due to its massive size).

By examining their communication network (Agent X Agent Network of users with mentions, replies, and tweets), we were able to identify the accounts they interacted with the most, the accounts that tweeted the most, and the accounts that helped spread their messages by retweeting it. Observing and monitoring this communication network of DHNs was key to understanding the content they were trying to spread e.g., a message or propaganda.

We then examined their social network (an Agent X Agent network of friends/followers of the hacking groups). This enabled us to identify the source of information accounts (accounts who have top out-degree centrality), the accounts that receive the information (accounts who have top in-degree centrality), and the accounts that work as bridges between the different parts of the network (accounts who have top betweenness centrality). Studying the social network was important to understand the roles of the nodes in the network (e.g., bridges, seeders of information, etc.) and the coor-

[6] #OpNimr: Anonymous fight to stop execution of Saudi youth. *AlJazeera*. 2015

[7] Anonymous launched #OpBeast against animal cruelty and depravity. *TechWorm*. 2015

dination strategies these groups followed to spread their message to their audience

3.2.2 FSA Findings

Here, we applied the FSA algorithm on the *communication network* and the *social network* to find the most coordinating **set of nodes** instead of **single node** analysis. FSA is a recursive modularity-based algorithm. Modularity is a network structural measure that evaluates cohesiveness of a network (Girvan & Newman, 2002). The FSA algorithm works in two steps:

Top-down division step: the algorithm will partition the network into subgraphs or sub-structures. This will obtain the candidate focal structures from the complex network by applying the Louvain method of computing modularity (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008).

Bottom-up agglomeration step: the FSA algorithm will stitch the candidate focal structures, i.e., the highly interconnected focal structure or the focal structures that have the highest similarity values are stitched together, then the process iterates until the highest similarity of all sibling pairs is less than a given threshold value.

The similarity between two structures is measured using the well-known Jaccard's coefficient (Şen et al., 2016; Jaccard, 1912) which results in a value between 0 and 1 (where 1 means the two networks are identical and 0 means the two networks are not similar at all). The stitching of the candidate focal structures was performed to extract structures with low densities (Şen et al., 2016). We used the development version of FSA in the UALR Center of Social Media and Online Behavioral Studies (COSMOS) lab[8](Şen et al., 2016).

We ran the FSA algorithm on the **communication network** (see Figure 2) which enabled us to identify the set of nodes that are communicating together at the highest frequencies. Running FSA here resulted in 8 FSAs (groups) which contain 58 nodes total (the number of nodes in each FSA is shown in Table 1, FSA with $ID = 1$ has the highest number of nodes i.e., 18 nodes). Figure 2 shows these focal structures, which are marked inside the blue circle in the figure on left. Upon zooming-in on this structure (displayed on the right) we found the set of nodes that communicate the most. Each FSA is distinguished by different colors. For example, the *Hacktivist groups* (@OpAnonDown) and (@CypherLulz) communicate together a lot more than the rest of the nodes in the network since they are in the same focal structure.

We also applied the focal structure algorithm on the **social network** (the friends and followers network) (See Figure 3) to find the set of nodes that according to their position in the network made them the most coordinating set of nodes instead of a single node (a focal structure has to act together to be powerful or effective). This resulted in 2 FSAs (set of nodes or group of nodes) which contained 46 total nodes (figure 3 on the left is the social network with the the FSAs, while the right side represents the a zoomed-in snap showing the 2 FSA nodes). An example is the hacking group *Think Tank group* and *the Cult of the dead cow group* are very powerful/effective if they act together as they are very well connected in the Twitter network (they are in the same focal structure).

[8]The FSA version we used is available online at http://www.merjek.com (guest account - username: merjek, password: merjek123)

3.3 Proposed Framework

We propose a framework (See Figure 4) that can be used to discover the relationship between the deviant groups e.g., hacking groups. The framework starts with seed knowledge (e.g., accounts or keywords used by the group) which is usually provided by domain experts. Second, data collection with Twitter API (in our case we used NodeXl). Third, the analysis of the collected network by examining the overall network, then the communication network (tweets, retweets, mentions or replies) then the social network (the friends and followers network). Fourth, the application of FSA analysis (Şen et al., 2016) to discover coordinating groups. Finally, the derivation of insights from the results obtained. A similar approach has been applied to collect and analyze data during many online events that were carried out by *deviant groups* such as ISIL (Al-khateeb & Agarwal, 2015b), and during the dissemination of propaganda where the Dragoon Ride Exercise conducted (Al-khateeb & Agarwal, Under review).

4. CONCLUSION & FUTURE WORK

In this work, we explored one instance of an online deviant group (i.e., DHNs) network on Twitter. We proposed a framework that can be used to understand DHN's and their methods of communication as well as what they are communicating. Such a framework would enable authorities to act tactfully and deter and/or reduce the damage done by cyber attacks. To note, this framework can be reapplied to other forms of deviant groups utilizing OSNs, such as violent extremist groups and cybercriminal organizations.

The authors are aware of the limitation of this framework as it was applied only to Twitter, although there are many other social media outlets. They are also aware of the limitation of the data collection constrained by the Twitter API, as it represents a small sample of the total population of Twitter. For these limitations, we are planning to study the cross-media affiliation of these groups, e.g., the blogs these DHNs use to disseminate their messages, and potentially discover hidden links between those groups through cyber forensic tools and methodologies. In addition, we plan to study other social media outlets used by these groups such as Facebook, Instagram and Tumblr.

The analysis helped in answering our stated research questions. More specifically:

1. Do DHNs use social media platforms to coordinate attacks? If so, what are their levels of sophistication?
   Discussion: DHNs do use social media to coordinate attacks and invite their followers to participate in them in a very sophisticated manner as shown in the two aforementioned cases.

2. Can their affiliations be tracked and monitored?
   Discussion: the affiliation of DHNs can be tracked and monitored through the usage of online social network analysis tools and techniques. This is part of what we performed in this study.

3. Can their coordination strategies and communication networks be inferred from empirical observation, data set collection, and analysis?
   Discussion: we believe by collecting data sets of different events from these groups we can leverage machine learning to help in predicting the deviant acts or their corresponding triggers/causes.

4. Can the most influential set of nodes be identified?

Discussion: the most influential set of nodes can be identified as our results showed using the FSA algorithm.

Additionally, we articulated the need to bridge the gap between social network analysis and cyber forensic techniques to uncover hidden relationships between deviant groups, and obtain forensically-relevant data. The results showed a strong connection between those different DHNs through their social network and also a high volume of communication seemed to occur between them. This suggests a sophisticated ability to coordinate deviant acts through their social media communication vectors. By applying FSA we were able to identify the most influential set of nodes in the observed social and communication networks. By using SNA we were able to identify the powerful nodes, bridges between different network components, the hashtags/keywords/bigrams they use, who is disseminating the most information, etc.

For future work, we plan to analyze the message contents of these groups, their sentiments, and how their followers' sentiments change over time. By monitoring their sentiments and communications, a predictive model of an attack or organized deviant act can be developed.

ACKNOWLEDGMENTS

This research is funded in part by the U.S. National Science Foundation (NSF) (award numbers IIS-1110868 and ACI-1429160), U.S. Office of Naval Research (ONR) (award numbers: N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412), U.S. Army Research Office (ARO) (award number: W911NF-16-1-0189), U.S. Air Force Research Lab (AFRL), and the Jerry L. Maulden/Entergy Fund at the University of Arkansas at Little Rock. We would also like to acknowledge the valuable contributions of the undergraduate students, who worked on the research projects supported by the U.S. National Science Foundation's (NSF) Research Experiences for Undergraduates (REU) program (award number: CNS-1359323). This research is also funded in part by the Elder Family Chair Endowment at the University of New Haven. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

REFERENCES

Al-khateeb, S., & Agarwal, N. (2015a). Analyzing deviant cyber flash mobs of isil on twitter. In *International conference on social computing, behavioral-cultural modeling, and prediction* (pp. 251–257).

Al-khateeb, S., & Agarwal, N. (2015b). Examining botnet behaviors for propaganda dissemination: A case study of isil's beheading videos-based propaganda. In *2015 ieee international conference on data mining workshop (icdmw)* (pp. 51–57).

Al-khateeb, S., & Agarwal, N. (Under review). Examining the use of botnets in propaganda dissemination case studies of the 2014 crimean water crisis and the 2015 dragoon ride exercise. *NATO Strategic Communication Center of Excellence (STRATCOM CoE)*.

Al Mutawa, N., Baggili, I., & Marrington, A. (2012). Forensic analysis of social networking applications on mobile devices. *Digital Investigation*, *9*, S24–S33.

Baggili, I., & Breitinger, F. (2015). Data sources for advancing cyber forensics: What the social world has to offer. In *2015 AAAI Spring Symposium Series.*

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.

Foundation, S. M. R. (2014). *NodeXl: Network Overview, Discovery and Exploration for Excel.*

Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, *99*(12), 7821–7826.

Hanneman, R. A., & Riddle, M. (2005). Introduction to social network methods. In (chap. 7). University of California Riverside.

Huber, M., Mulazzani, M., Leithner, M., Schrittwieser, S., Wondracek, G., & Weippl, E. (2011). Social snapshots: Digital forensics for online social networks. In *Proceedings of the 27th annual computer security applications conference* (pp. 113–122).

Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New phytologist*, *11*(2), 37–50.

Mulazzani, M., Huber, M., & Weippl, E. (2012). Social network forensics: Tapping the data pool of social networks. In *Eighth Annual IFIP WG* (Vol. 11).

Seddon, M. (2014, May). *How a british blogger became an unlikely star of the ukraine conflict and russia today.* Retrieved 2015-08-19, from http://bzfd.it/1qpuL2z

Şen, F., Wigand, R., Agarwal, N., Tokdemir, S., & Kasprzyk, R. (2016). Focal structures analysis: Identifying influential sets of individuals in a social network. *Social Network Analysis and Mining*, *6*(1), 1–22. Retrieved from http://bit.ly/1qS8Y4D doi: 10.1007/s13278-016-0319-z

Şen, F., Wigand, R. T., Agarwal, N., Mete, M., & Kasprzyk, R. (2014). Focal structure analysis in large biological networks. *3rd International Conference on Environment Energy and Biotechnology.*

Walnycky, D., Baggili, I., Marrington, A., Moore, J., & Breitinger, F. (2015). Network and device forensic analysis of android social-messaging applications. *Digital Investigation*, *14*, S77–S84.

Yuce, S., Agarwal, N., Wigand, R. T., Lim, M., & Robinson, R. S. (2014). Studying the evolution of online collective action: saudi arabian women's 'oct26driving'twitter campaign. In *International conference on social computing, behavioral-cultural modeling, and prediction* (pp. 413–420).

APPENDIX A

This appendix contains the results we obtained in our project and mentioned in Section 3.2.

Table 2. Top 10 domains and hashtags

Domains	Entire Graph Count	HashTag	Entire Graph Count
twitter.com	536	anonymous	392
ccc.de	389	opnimr	379
anonops.com	177	opbeast	213
bit.ly	111	intelgroup	200
buff.ly	96	opisis	177
youtu.be	81	cccamp15	125
gov.br	80	higsec	118
co.uk	76	alimohammedalnimr	117
pastebin.com	76	offline	94
rt.com	71	radioanonops	90

Table 3. Top 10 keywords and bigrams

Keyword	Entire Graph Count	Bigram	Entire Graph Count
rt	3256	tango,down	173
anonymous	351	rt, anonintelgroup	94
opnimr	381	saudi, arabia	79
down	308	rt, heidi_coon	65
amp	278	down, lulzsecroot	60
now	250	rt, youranonnews	58
isis	234	heidi_coon, opbeast	47
opbeast	211	rt, kaidinn	45
up	209	lt,3	43
intelgroup	209	alimohammedalnimr, opnimr	43