THE JOURNAL OF
DIGITAL FORENSICS,
SECURITY AND LAW

Journal of Digital Forensics,
Security and Law

Volume 11 | Number 4 Article 8

12-31-2016

# A Survey of Social Network Forensics

Umit Karabiyik
*Department of Computer Science Sam Houston State University*

Muhammed Abdullah Canbaz
*2Department of Computer Science and Engineering University of Nevada*

Ahmet Aksoy
*2Department of Computer Science and Engineering University of Nevada*

Tayfun Tuna
*3Department of Computer Science University of Houston*

Esra Akbas
*4Department of Computer Science Florida State University*

*See next page for additional authors*

## Recommended Citation

Karabiyik, Umit; Canbaz, Muhammed Abdullah; Aksoy, Ahmet; Tuna, Tayfun; Akbas, Esra; Gonen, Bilal; and Aygun, Ramazan S. (2016) "A Survey of Social Network Forensics," *Journal of Digital Forensics, Security and Law*: Vol. 11 : No. 4 , Article 8.
DOI: https://doi.org/10.15394/jdfsl.2016.1430
Available at: https://commons.erau.edu/jdfsl/vol11/iss4/8

# A SURVEY OF SOCIAL NETWORK FORENSICS

Umit Karabiyik[1], Muhammed Abdullah Canbaz[2], Ahmet Aksoy[2], Tayfun Tuna[3], Esra Akbas[4], Bilal Gonen[5], Ramazan S. Aygun[6]

[1]Department of Computer Science
Sam Houston State University, Huntsville, Texas, USA
umit@shsu.edu
[2]Department of Computer Science and Engineering
University of Nevada, Reno, Reno, Nevada, USA
mcanbaz@unr.edu, aksoy@nevada.unr.edu
[3]Department of Computer Science
University of Houston, Houston, Texas, USA
ttuna@uh.edu
[4]Department of Computer Science
Florida State University, Tallahassee, Florida, USA
akbas@cs.fsu.edu
[5]School of Information Technology
University of Cincinnati, Cincinnati, Ohio, USA
bilal.gonen@uc.edu
[6]Computer Science Department
The University of Alabama in Huntsville, Huntsville, Alabama, USA
aygunr@uah.edu

## ABSTRACT

Social networks in any form, specifically online social networks (OSNs), are becoming a part of our everyday life in this new millennium especially with the advanced and simple communication technologies through easily accessible devices such as smartphones and tablets. The data generated through the use of these technologies need to be analyzed for forensic purposes when criminal and terrorist activities are involved. In order to deal with the forensic implications of social networks, current research on both digital forensics and social networks need to be incorporated and understood. This will help digital forensics investigators to predict, detect and even prevent any criminal activities in different forms. It will also help researchers to develop new models/techniques in the future. This paper provides literature review of the social network forensics methods, models, and techniques in order to provide an overview to the researchers for their future works as well as the law enforcement investigators for their investigations when crimes are committed in the cyber space. It also provides awareness and defense methods for OSN users in order to protect them against to social attacks.

**Keywords**: Digital Forensics, Social Networks Analysis, Online Social Networks, Crime, Terror, Deception, Social Attacks

# 1. INTRODUCTION

Digital forensics, a branch of forensic science, deals with the investigation and recovery of digital information found in digital devices which are mostly found at crime/incident scenes. Early twenty-first century became the golden age of digital forensics because of the technological advances in today's world. Especially with convenient accessibility of smart devices, most data are now being stored and shared in digital forms such as pictures, diaries, calendars, videos, etc. Smart phones, tablets, computers, smart household devices. Moreover, wearable devices have already become part of our everyday life. The demand towards the usage of technological advances makes tremendous amount of data being stored and shared particularly in online social networks. This inevitable developments make any device a storage of potential evidence related to a crime or an incident.

Any type of information stored or transferred in digital form can be identified as digital evidence (Nelson et al., 2015). The data collected from social networks may contain invaluable evidence for a digital forensics investigator in both criminal and corporate investigations. For example, social media is one of the recent developments that attracts everyone regardless of their age, gender, socioeconomic status, etc. and it produces large amount of digital data for analysis (Raghavan, 2013).

Social network analysis (SNA) is a method of using network and graph theories in order to investigate and analyze social structures for different circumstances (Otte & Rousseau, 2002). From the digital forensics investigation point of view, the main goal of using SNA in forensically sound manner is to understand the relations (i.e., links or edges) between the actors (nodes) for variety of purposes such as solving criminal activities, preventing terrorist attacks, identifying social attacks, detecting deceptions, categorizing and matching social network accounts, etc. Mathematical models, techniques and tools are developed in order to improve understanding, preventing, detecting, and predicting potential criminal activities. Therefore, it is an absolute need for investigators and researchers to be familiar with such work in order to confront with related challenges. Figure 1 shows sample forensics studies using OSNs on crime and terror incidents around the world.

The vast usage of OSNs and availability of supporting application interfaces have attracted significant forensics analysis on OSNs. Furthermore, the number of offenses against the law where OSNs are utilized is also increasing everyday. It is necessary to understand and analyze crimes and attacks through online social networks and prevent criminal activities, detect malicious users, and solve criminal cases. Moreover, the safety of OSN users should be increased as much as possible. We have analyzed numerous papers related to threats mediated through OSNs. The diversity and numerousness of research studies along with significant similarities in terms of the ways these crime/terror networks operate, how social attacks are designed, and damages caused by these criminal activities and attacks aggravate feasibility of an effective forensic analysis as well as a comprehensive survey of studies. In the literature, we have not found a comprehensive coverage of topics that would help forensics analysts see the broad picture of cases and understand existing attacks and solve their cases. Our approach is not to list possible threats and cover relevant studies for each threat. For example, identity-theft, identity cloning, spam, child pornography, finding crime networks have been studied and solutions have been proposed for each type of crime or attack. Such cover-

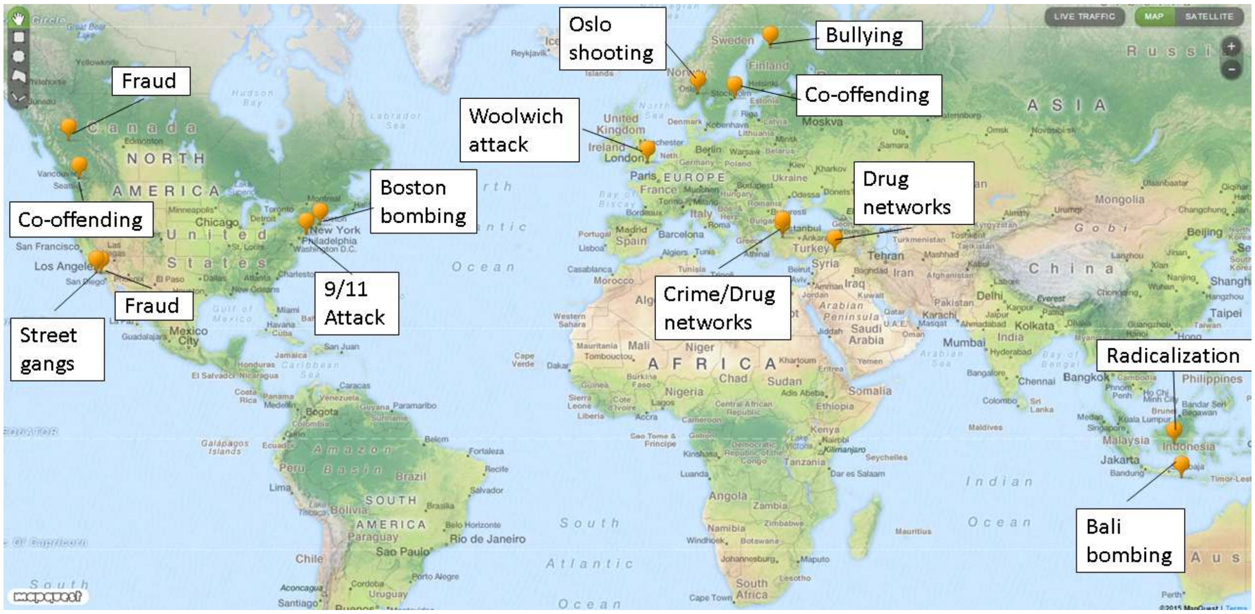

Figure 1. Map of sample crime analyses around the world

age has limitations for forensics analysts and typically requires new ways of understanding with the launch of a new type of attack.

The main goal of this paper is to provide an extensive review of social network forensics research that would help to understand various aspects of forensics research on OSNs starting from crime and terror network structures and their analysis to social attacks through OSNs, and the countermeasure methods against to these attacks. Particularly, we present relevant research in a way that necessary tools could be developed for forensics community, systematic methods could be designed for solving crimes (initiated through OSNs), and methods for protecting OSN users are applied effectively.

A survey on social network analysis is not complete without covering topics related to crime/terror networks and social attacks. Therefore we cover both aspects in this survey. We explain how OSNs can be analyzed to understand crime and terror network structures, how social attacks are designed, how OSN users could be protected, and how methods could be developed to help forensics analysts to detect, solve or prevent social attacks. The major difference between terror/crime network structures and social attacks through online social networks is that the first one is organized by a group of people whereas the second group of attacks could be initiated by a single person. To the best of our knowledge, this paper is the first comprehensive area survey of social network forensics with respect to aforementioned objectives. Therefore, the major contribution of our paper is the organization and coverage of diverse research studies on crimes and attacks, which could have similar damages or utilize similar methods, to help forensics community and protect OSN users. More specifically, our paper has the following contributions:

- a comprehensive overview of research studies on crime and terror network structures with focus on analyzing network structures, spatial analysis of crime networks, ranking individuals and relationships, identifying co-offending networks, similarity of crime networks, information diffusion related to terrorist attacks, and socio-economic relationships affecting crime and terror,

- a novel categorization and analysis of social attacks through online social networks based on proposed four levels of attacks: (i) crime or criminal activity (e.g., identity-theft, child pornography), (ii) type of attack (i.e., information disclosure or opinion/emotion influence) (iii) scheme of attack (social engineering and information discovery), and (iv) attack components (fake identity, fake messaging, fake relationship, and data crawling) and coverage of literature including children-oriented attacks with respect to this categorization,

- coverage of forensics analysis methods and protection according to our categorization of attacks including our proposed way of categorizing privacy preservation methods, handling fake identities (user categorization, profile matching, authorship analysis, detecting deception, defenses for de-anonymization), handling fake messages (preventing, detecting, demoting, credibility, improving resilience to infection), handling fake relationships, and searching traces of data, and

- a detailed list of open research issues and future direction along with suggestions for OSN users, OSN service providers, forensics analysts and law enforcement agencies.

This paper is organized as follows. The following section provides research studies related to understanding analysis of crime and terror network structures with respect to ranking individuals and relationships, co-offending relationships, spatial analysis of networks, similarity of crime networks, visualization of such networks, information diffusion not only for terror attacks but also for white-collar crimes, and influence of socio-economic relationship on crime and terror. Section 3 discusses social attacks through online social networks. Forensic analysis and defense methods for social attacks are particularly explained in Section 4. Finally, the last section introduces open research issues in social network forensics based on currently available methods and concludes our paper.

# 2. UNDERSTANDING CRIME AND TERROR NETWORK STRUCTURE

Understanding crime and terror network structures plays critical role for forensics investigations. The analysis of these structures include understanding the rank of individuals, relationships and roles of people, collaboration among individuals and different networks, spatial analysis based on mobility, and similarity between networks. Determining the underlying graphs may be used to crack down these networks and prevent possible attacks. Despite the vital importance of such analysis, currently available digital forensics tools (both commercial and open source) are not sophisticated enough to provide thorough analysis. On the other hand, there are many research efforts to answer this need individually by focusing on particular events and incidents.

In Table 1, we categorize relevant research on network structures along with other factors that influence crime and terror networks. Once the structures are identified, tools and techniques to further analyze them becomes necessary. Hence, in this section we discuss available methods, tools and techniques which are used to analyze network structures, perform ranking among the individuals in criminal networks, measure similarities between the criminal networks, and investigate diffusion information on these networks. Furthermore, we also aim to attract digital forensics researchers' and tool developers' attention to these available studies for their future tool developments on network analysis in a broader sense. In the next section we start with network structure analysis using certain metrics on individuals as well as their relations and behaviors in the crime and terror networks.

## 2.1 Network Structure Analysis

In order to understand the network structure analysis, it is crucial to grasp centrality metrics which are classified as the most known characteristics of the network analysis. The *Degree* centrality is the number of ties that a node has whereas the *betweenness* centrality is the number of shortest paths from all vertices to all others that pass through a specific node averaged over all pairs of node in a network. The *closeness* centrality is the sum of distances to all other nodes. These centralities may reveal information about links between individuals and their roles in network structures.

Morselli et al. (2007) investigate Kreb's terrorist network (V. Krebs, 2002) and the Caviar network (UCINET, 2017) in order to analyze the effects of degree, betweenness, and closeness centrality criteria, and how they can be used for the detection of terrorist networks and criminal enterprise networks. The Kreb's terrorist network data consists of 37 participants in which 19 of them are hijackers who executed the attacks and other 18 additional contributing criminals. On the other hand, the Caviar network includes 110 participants of a profit driven criminal enterprise network. The major difference between these two datasets is that the Kreb's terrorist network is described as a snake-like, sparse network showing substantial distance within the members of the group whereas the Caviar's Network is a cluster-like network with shorter distances between the members. In their study, terrorist activities are observed to be operated without a core centrality member or person, whereas criminal enterprise networks such as drug trafficking built around the central members. Such analysis can be used to understand the purpose of forming criminal groups.

Centrality metrics can be also quite useful when performing analysis on the resilience of a network. For instance, Piraveenan et al. (2012) introduce a metric to analyze the structural robustness of networks by measuring the change of size of the largest component with respect to the network node removals. This measure is suitable for random and sequential node removals. In an ideal robust network, node removal should decrease the size of the largest component linearly, whereas in real networks this decrease could be nonlinear. Therefore, the robustness coefficient can be defined as the ratio of the size of the largest component in this network to the size of the largest component in ideal robust network. After defining the coefficient, Piraveenan et al. (2012) demonstrated how it can be used to evaluate the real world network data. An analysis is done for *degree based ordering*, *betweenness centrality based ordering*, and *closeness centrality based ordering* to show that this measure can be valuable to choose a strategy to attack/defend a network. Degree based attack analysis is performed by removing the highest order node in a network whereas random attack analysis is performed by removing a node in random degree order.

The experiment results for different networks reveal that *cortical networks*, *neural networks* and *food webs* show the highest ro-

Table 1. Overview of studies on network structures and information diffusion

| Paper | Category | Application | Data Source | Location /Attack |
|---|---|---|---|---|
| Bora et al. (2013) | network structure | street gangs | Twitter | Los Angeles |
| Ozgul & Erdem (2012) | identify networks | drug network | | Diyarbakir |
| Ozgul & Erdem (2012) | co-offending | criminal & drug networks | | Bursa, Diyarbakir |
| Tayebi & Glasser (2012) | co-offending | | | |
| Hipp et al. (2013) | information diffusion | social tie & crime relationship | | |
| Morselli et al. (2007) | network structure | leadership | Krebs, Caviar | |
| Tayebi et al. (2014) | co-offending& spatial behavior | | Police Information Retrieval System | Metro Vancouver |
| Frank (2001) | co-offending | offender-crime reporting relationship | | Stockholm |
| Hipp (2010) | socio-econonomics | social distance-crime relationship | American Housing Survey | |
| Baker & Faulkner (2004) | social network communication | fraud | Fountain Oil & Gas company | Ventura County California |
| McBride & Caldara (2013) | criminal network visualization | | | |
| Semenov et al. (2013) | network structure | tracking activities | vk.com | Boston bomber |
| Burnap et al. (2014) | information diffusion | survival & volume of messages | Twitter | Woolwich, London, UK |
| Sundsoy et al. (2012) | communication &relations | | Telenor (phone calls) | Oslo |
| Wiil et al. (2010) | network structure | link importance & removal | | 9/11 & 2002 Bali night attack |
| Breiger et al. (2014) | groups connectivity, similarity | drug trading | Big Allied and Dangerous (BAAD-1) | |
| Spezzano et al. (2013) | network structure | person successor problem | Sageman's Alqaeda and Lashkar-e-Taiba | |
| Husslage et al. (2015) | network structure | ranking individuals | | Al Qaeda 9/11 |
| Schweinberger et al. (2014) | information diffusion | relief operations | | 9/11 Attack |
| Ozgul et al. (2011) | network similarity | modus operandi | | Istanbul |

bustness whereas the Internet networks are among the least robust networks. This implies that man-made rapidly evolving networks are more vulnerable than biological networks. Lastly the results show that (i) the networks are more robust to random attacks and vulnerable to targeted attacks, and (ii) the networks are least robust against betweenness centrality based attack and most robust against closeness centrality based attacks. Network characteristics provide invaluable perspectives to better understand the complexity of the crime networks. Two papers above may not directly reflect the nature of researches on forensics investigation; however, they provide unique angles to see already complex structures, namely terrorist and criminal networks.

In addition to the network analysis from the centrality metrics perspective, categorization of users by their behaviors, and the analysis of these categorization in online social networks can be helpful to extract criminal networks from social networks. For instance, Vigliotti & Hankin (2015) focus on detecting unidentified criminals by following a few known criminals on social networks. They implemented a model derived from binomial distribution and discrete time models for behavior classification (normal or abnormal) by checking the instances of the behavior occurrences. The experiments are performed on two datasets: (i) Twitter dataset containing 11K users followed in a period of 45 weeks and (ii) the VAST dataset containing 400 users' phone records over 10 days. A directed graph $(N, E)$ is constructed where $N$ represents the number of nodes and $E$ represents the edges between these nodes. A time interval can be defined as $[t_n, t_{(n+1)}]$, where $t_n$ is the beginning time of an event and $t_{(n+1)}$ is the end of the event. With respect to time series, relationship of the edges can be obtained easily by using the equations below.

$$N_{i.}(t) = \sum_{j,i\neq j} N_{i,j}(t) \tag{1}$$

$$N_{.j}(t) = \sum_{i,i\neq j} N_{i,j}(t) \tag{2}$$

$$N_{..}(t) = \sum_{j,i\neq j} \sum_{i,i\neq j} N_{ij}(t) \tag{3}$$

Equation 1 represents the outgoing edges from node $i$, Equation 2 represents the incoming edges to node $i$, and the Equation 3 is the count of both outgoing and incoming connections for node $i$. Considering time intervals for each event, the time series for each node can be obtained as $dN(t_1)$, $dN(t_2)$, $\ldots$, $dN(t_n)$ where $dN(t_k)$ represents the number of relevant communications between time $t_k$ and $t_{k+1}$. The probability of a communication at interval $t_1$ is defined as $P(dN(t_1))$. The number of communications over $n$ intervals can be specified using a random variable $B(n) = dN(t_1) + dN(t_2) + \ldots + dN(t_n)$ assuming that the presence of communication is sampled using Bernoulli. The probability of $r$ communications can simply be computed using the binomial distribution as in Equation 4:

$$\mathbb{P}(B(n) = r) = \binom{n}{r} \pi^r (1-\pi)^{n-r} \tag{4}$$

The authors use discrete time stochastic process to model the connections (e.g., telephone calls) in a network. The links (at least one communication over interval) between nodes are sampled using Bernoulli $(Be(\pi))$ while the frequency $(r)$ of links between two nodes over an $n$ intervals is modeled using Binomial distribution.

As a result of these experiments, it is shown that 36 suspects out of 400 users in VAST dataset are detected. As for the Twitter dataset, roughly 1,600 out of ∼11K Twitter users were detected as suspects. In both cases, the number of anomalous nodes stays below ∼10% of the total population. This work clearly shows, despite the privacy related concerns, that publicly available data on online social networks may be monitored in order to detect perpetrators and criminals by knowing the characteristics of known criminals on online social networks.

Links in a network deliver important insights on analysis of a crime and how it is committed. In addition to (Vigliotti & Hankin, 2015), Semenov et al. (2013) describe how social media could be helpful for law enforcement in order to track the traces before and after a crime is committed. The authors claim that potential traces can be found in social media sites in order to prevent crimes if the profiles are analyzed before the crime is committed. They specifically take the *Boston Bomber* case as their focus and present how activities can be traced. Data is collected from *vk.com* (Russia based social network) using an API provided by *vk.com*. *Boston Bomber* case is first analyzed qualitatively and quantitatively. The collected data from one of the suspects' profile show the dynamics on the terrorists' alleged social media accounts before and after the attack. They also present data regarding the origin of the activities on those alleged accounts with respect to the comments on suspect's profile and the number of contributing users by the country. Second, the case is analyzed from the social network analysis point of view. The collected data consists of 1,208,892 nodes in total including the terrorist friends, second degree friends, and third degree friends. In their analysis, they show the number of friends found in the first, second and third degree friend circles and the origins of these friends regarding to their countries. They also show how the suspect's friends stop following him after his name appeared on news. In the meantime, data also show that some the alleged suspect's friends appear in the network to support him by sharing pictures and posts of him claiming his innocence.

Most of the time, a single representation of a network structure may not be enough to grasp the complete story of the network where there can be hidden insights. Analysts and the investigators need an integrated suite of capabilities that can help them rapidly understand and glean insight from a network, visualize knowledge by interacting with data, and collaborate to draw conclusions. Hence, we now discuss a brief summary of visualization for crime and terrorist networks particularly on virtual networks.

Visualization of networks is studied in order to understand the cognitive recognition process of human subjects on various network representations. We are aware that there are other works which merely focus on visual representations of networks themselves without taking human perception into account (e.g., Chen et al. (2004); J. Xu & Chen (2005); C. C. Yang & Ng (2007); Didimo et al. (2011); Chen et al. (2003)). However, they are out of the scope of this work.

The presentation of data in a pictorial or graphical format makes a big difference. Visual representations such as charts and maps help understand information more easily and quickly. In a research conducted by (McBride & Caldara, 2013), a comparative study of the data representation in either a table based or graph based representation is presented and discussed in order to get more meaningful visual representation of a criminal network. A mathematical model is designed in order to demonstrate the criminal networks. Then, an experiment of three sessions is set up with 86 subjects. Subjects are shown some networks which are randomly selected by a computer from a set of 10 out of 16 networks. The networks are randomly generated to illustrate the typical degree distribution of the social networks. As an outcome of the experiments, it is stated that a graph representation leads to a faster disruption decision on the criminal networks; however, the table style representation may lead to a more accurate disruption choice. Even though there is a trade-off between these two approaches, using both can give the best results.

As the above works show, the analyses of network structures for both terror and criminal groups (including individuals involved) would provide law enforcement and forensics examiners quite a bit of knowledge and evidence about the incidents in order to answer their questions.

## 2.2 Spatial Analysis of Crime Networks

In addition to structural network analysis, spatial analysis is also critical for identification of criminal groups and their activities. Spatial analysis from this aspect aims at identifying and quantifying spatial relationships and distributions of crime patterns. In this section, we provide summaries of spatial analysis research focusing on crime networks such as identification of the radicalization (Mazumder et al., 2013), human behavior and mobility patterns (Bora et al., 2013) and prediction of the locations of upcoming criminal activities (Tayebi et al., 2014).

Online social networks can be used to identify radicalization using the content and geolocation of tweets of the users. Mazumder et al. (2013) state that online social networks provide large amount of information on what people feel about other people, places, events, and political activities, etc.. As a part of the Minerva project at Arizona State University, activity information of the twitter users has been collected to understand political activities of Indonesia. Based on these data, a heat map is generated for the radical activities in provinces of Indonesia by computing *radicalization index* and *location index* of each Twitter user from Indonesia. Rather than considering tweets individually, all data of a user are analyzed. A *degree of radicalism* is assigned to each user based on the content of his or her tweets. Since Twitter API provides geolocation of tweets, it is possible to track the location of a user based on his or her tweets. Hence, geolocation is used as another parameter. Considering tweets for the same user might have been posted in different provinces, a probability distribution estimate is used per user. Results show high accuracy in detecting radicalism in users' tweets. The higher the value in the radicalism index, the higher the radicalism is in the province.

Mobile devices, on the other hand, generate enormous amount of data from location services. These data help model human behavior and mobility patterns. The study in (Bora et al., 2013) aims to observe human movement to understand the relationships of geographical regions, neighborhoods, and gang territories by analyzing 10 million geographically tagged tweets from Los Angeles. Using a graph based representation of street gang territories as vertices, and interactions between them as edges, they train a machine learning classifier (i.e., Naive Bayes Classifier) in order to distinguish the rival and non-rival links. They were able to correctly identify 89% of the true rivalry network, which beats a standard baseline by about 30%. Looking at larger neighborhoods, they were able to show that the distance traveled from home follows a power-law distribution, and the direction of displacement (i.e., the distribution of movement direction) can be used as a profile to identify physical (or geographic) barriers when it is not uniform. Finally, considering the temporal dimension of tweets, they have detected events taking place around the city by identifying irregularities in tweeting patterns. This method can be easily adapted to create new tools for law enforcement to detect outbreaks led by gangs and other criminal actors in real time.

In addition to real time event detection, it is also possible to predict future criminal activities. In order to do that, spatial behavior of criminals can be analyzed to predict the whereabouts of their upcoming criminal activities. Tayebi et al. (2014) propose a probabilistic model on the well-known offenders' spatial behavior and an extended version of earlier random walk model derived by Short et al. (2008), *CRIMETRACER*, which brings functionality to generate the activity space associated with the offenders living in an urban area. In *CRIMETRACER*, random walk process is personalized by using co-offending information, crime trends of offenders, and also the crime event history of road segments in order to uncover the unknown spatial behavior of every single offender. The data is retrieved from the Police Information Retrieval System (PIRS) which particularly focuses on crimes in Metro Vancouver, where there is a road network composed of 64,108 road segments with an average length 0.2 km. In *CRIMETRACER*, the first task is to get the list of an offender's crime events chronologically. Then, it generates the offender's activity space by training the model with the first 80% of the crime data. The rest of the data (the crime locations in particular) are used for testing the model. In the evaluation, offenders with at least two different crime locations are considered, which corresponds to ∼10% of the offenders in the crime data set. The results of *CRIMETRACER* are evaluated with methods such as random walk, hot spots, proximity, offender-based collaborative filtering (CF), location-based CF, co-offending-based CF which are similar to the state-of-the-art methods for location recommendation. *CRIMETRACER* exceeds all other methods which are used in the experiments.

## 2.3 Ranking Individuals and Relationships in Crime Networks

As discussed above, nodes represent the active objects in the terror and crime networks. In many cases, nodes establish, maintain, and control the events occurring in the network. In this section, we explain research studies which particularly focus on analyzing the ranking of individuals and criminals (Sarvari et al., 2014; Husslage et al., 2015), importance of links between the nodes (Wiil et al., 2010), and the person successor problem (Spezzano et al., 2013).

Sarvari et al. (2014) suggest that information regarding the organization of a criminal community can be achieved by analyzing criminal social graph structures. Sarvari et al. (2014) aim to construct a large scale social graph from a small set of email addresses of some criminals. By using the Facebook profiles that were associated with these email addresses, they constructed a social graph of 43,000 nodes using 1,000 email addresses. Then, a large scale analysis was performed on this graph to identify profiles of high rank criminals, criminal organizations, and large scale communities of criminals. Finally, a manual analysis was performed on these profiles to detect many public criminal groups on Facebook. The data (email addresses) were collected from *BestRecovery* (V. Krebs, 2002), an online data theft service. Centrality measures and community detection techniques were used to determine criminals having a central position in the graph, subgroups and communities found in the network, criminals acting as brokers of collaboration and information in the network, and the ranking of criminals based on their importance and influence on the network. It was shown that the key members of the networks had high ranks with respect to every centrality measure. The similar results were observed with the *PageRank* dataset (Page et al., 1999) as well. This shows that highly connected members are positioned in the center of the graph. This also indicates that they have connections with other well-connected members. If we take the readily available social network websites and corresponding data collection tools into account, it would be indispensable technique for law enforcement to identify and eventually eliminate key figures in such criminal communities.

In addition to the identification of key players in terrorist networks, the terrorist networks can also be used for ranking individuals and ultimately preventing likely terrorist attacks. Husslage et al. (2015) introduce a new game theoretic model as an extension of (Lindelauf et al., 2011). In addition to the earlier work, this method also takes the operational strength of connected networks into account. The ultimate aim of the proposed model is two-fold: providing certain strategies to prevent possible terrorist attacks, and destabilizing the terrorist networks by ranking the individuals based on their importance in the terrorist networks. Authors used sensitivity analysis on the rankings gathered from the Al Qaeda's 9/11 dataset (Kean et al., 2004) (B. Krebs, 2013) for the purpose of testing the robustness of their model. During the analysis, they took possible additional information of the network members, variations in relational strengths, and the absence or presence of a small percentage links in the network into account. Their proposed game theoretic model was developed by using monotonic weighted connectivity game. This model is used to calculate the centrality measure of the network. In the rest of this work, the authors used sensitivity analysis methods to test their models' robustness under different circumstances including small variations in the weights, addition/deletion of a small portion of the links, and changes in the weights related to interactions. The results show that their model is robust in a sense that it is able to rank the individuals in the network despite small changes in the dataset. Considering the successful applications of game theoretic models into many real life problems particularly in security related issues (Zhang et al., 2013; Jiang et al., 2014; Delle Fave et al., 2014), we expect to see more of similar methods being applied to various digital forensics domains in addition to (Husslage et al., 2015).

Thus far, we have discussed the importance of the nodes in various network analysis methods. However, the links of a network may play a critical role as its nodes with respect to network efficiency and secrecy. Wiil et al. (2010) aim to determine which link rather than node is playing more crucial role in communication, and also whether or not removing the respective link affects the secrecy and efficiency of the network. After analyzing the current available strategies, the authors implemented a method to analyze the importance of links in a terrorist network based on a transportation network (which is later implemented in *CrimeFighter Assistant* (Wiil et al., 2010)) and other well-known terrorist attacks. The idea of this method/technique is originated from the social networking analysis concept. In this proposed concept, removal of an important link from the network is expected to result destabilization of the network, since removing the link would result in connection or communication loss between the connecting nodes. This work was claimed to be the first implementation of the secrecy and efficiency measures when analyzing the link importance. After the complete analysis on the transportation network, a balanced view is presented for analyzing the data of a terrorist network by giving importance to links. Later this algorithm was tested on the well-known attacks such as *9/11 network* and *2002 Bali night club bombing network*. This algorithm was implemented in the *CrimeFighter Assistant* and compared with other approaches. The authors show that the measure of link importance proposes ways to view the important links that affect the performance of the terrorist networks for their destabilization.

Now we discuss the link formation in a social network after a critical node is removed in addition to above study on link removal. We believe this method may also help law enforcement to determine successor of a terrorist leader in terrorist networks when particular nodes are removed. Spezzano et al. (2013) present theoretical models in order to study the Person Successor Problem (PSP) by analyzing person replacements in terrorist groups. They aim to reduce the lethality of the terrorist groups by predicting the possible replacements when an important figure (e.g., leader of the terrorist group) is removed. They also aim to predict the new structure of the network after removal is performed. This work takes the property and expertise of the individuals in terrorist networks into account when predicting the changes in the network. The authors developed two algorithms, STONE-Predict and STONE-Reshape. These algorithms use a probability density function based on a probabilistic model in order to predict the person replacement and the reshaping of the network. These algorithms are tested on three datasets. The first dataset is synthetic dataset created by the authors. The second and third datasets are Sageman's Al-Qaeda network Sageman (2004) and Lashkar-e-Taiba network Subrahmanian et al. (2012) respectively. The results show that STONE-Predict algorithm achieves 80% accuracy of predicting who will replace the removed person in the network when tested on those datasets. In testing, the information in real datasets were used as ground truth, and domain experts were consulted. This work shows that collection of good data regarding terrorist of crime networks (e.g., Project Caviar (UCINET, 2017)) and using proposed algorithms after analyzing the dataset may eventually help to diminish these networks by identifying the key figures and their successors.

## 2.4 Identifying Co-offending in Crime Networks

Collaborative crimes end up with bigger impact and hazardous results. That is why well-structured terrorist groups become threats known globally. Having a network with high density complicates the understanding of the organization and the control dynamics of a network. Identification of the co-offending characteristics of a crime network provides insights for defense and security purposes. In this section, we provide overview on identification of the features which lead to committing crimes by Ozgul & Erdem (2013), partnership modeling and analysis of criminal networks by Tayebi & Glasser (2012); Frank (2001).

Identifying features that lead to committing crimes in a collaboration is important for forensics research. Ozgul & Erdem (2013) studied the identification of the crime features which are important for committing crime together. In their methodology, they first collected the co-offending connections of the criminals which represent the previous cooperation. Second, the *spatio-temporal* and *modus operandi* are used to identify the criminal networks. Third, they combined latter two outcomes to conclude whether it can help for detection of current criminal networks. Finally, demographic similarities of the criminals are used to observe the current networks. Two criminal network datasets from Turkey were used in their experiments: Bursa and Diyarbakir. Bursa dataset consists of 6,114 crimes, 478 criminals in 85 experimental criminal networks; on the other hand, Diyarbakir dataset contains 40 drug networks with about 56 crimes and 154 criminal records. The authors conclude that the history of the criminal connectivity plays a very important role in the current criminal networks. On the other hand the type of crime may differ but does not depend on the demographic characteristics.

Moreover, analyzing partnerships between criminals or criminal networks can help to understand co-offending networks. Tayebi & Glasser (2012) proposed a co-offending network analysis which can be handled benefiting from the social network analysis with the use of machine learning techniques. Their methodology is divided into four sections: modeling the crime data and extraction of the co-offending network, a group detection technique that is an extension of the clique percolation method, crime determination process that produces the same characteristics of the crimes, and an evaluation model that can analyze the groups in a time period. In the data modeling, a crime data is modeled as a tripartite hyper graph $H(N, E)$ where each node in $N$ is categorized as actors ($A$) such as criminals, victims, suspects, etc., incidents ($I$) that are the reported crimes, and resources ($R$) that are used in a crime such as vehicle, weapon, etc. A co-offending network is derived from the nodes connected to each other for the crimes they are involved together. Discovery of a crime group is derived by an algorithm that first collects the offender groups from the data, computes the activity and criminality scores based on the crimes committed by members, determines the resources benefited by the members, identifies the groups that are likely to be crime organizations, and updates the group evaluation results for each time frame. Dataset used in this project consists of roughly 4.4 million reported offenses, ∼4 million unique individuals, and 39 different subject groups. Five datasets in different sizes (with 9K to 21K data entries) are chronologically divided and used. However, only offenses with more than one offender are considered. As a result, from ∼4.4 million crime records with 150K co-offending networks, 20K of the offender groups are detected in which 1,800 of them were active. It is stated that a continuous crime partnership is observed and most of them do not endure for long periods.

In addition to partnership of offenders, relationship between offenders and victims can also be modeled in order to analyze and understand the network infrastructure. Frank (2001) discusses a statistical network model that can help to predict the secondary participation of crimes and understand the structure of co-offending youth networks. The juvenile crime data in Stockholm is used such that a bipartite graph model of the data is built from assumptions about crime reporting and offender detection for a specific type of crimes. A mathematical modeling of a basic crime model, related crimes and co-offending crime models are given as a matrix representation. In the given representations, a connectivity is defined as an adjacency matrix to come up with a simple probabilistic model for crime reporting and offender identification. Statistical inference (in particular, the number of co-offenders and the total number of offenses) is discussed. The author shows that it can be possible to present the numbers of offenders of different activities and estimate the total number of offenses by the model parameters via the numbers of crimes of different sizes.

## 2.5 Measuring Similarity of Crime Networks

Criminals and terrorists may form alliances, operate as an underground network, and share operational motivations. By analyzing the similarities of crime networks, it may be possible to determine the purpose of a newly detected crime networks. In this section, we deliver summaries on detection of the criminals based on their demographic features (Ozgul & Erdem, 2012), two-mode network analysis for understanding the formation and involvement of the terrorist groups in drug trading (Breiger et al., 2014), a similarity metric extraction based on development of the crime and the behaviors of the criminals (Ozgul et al., 2011).

Demographic features such as kinship can be effective for detection of the secondary (other) criminals in the network. Ozgul & Erdem (2012) propose an Extended Social Detection Model (XSDM), which aims to detect criminals by their demographic features such as ethnic origins, kinship, etc. Their model is a successor model of the Social Detection Model (SODM) (Ozgul et al., 2010). A set of new features such as neighborhood information which represents the criminals who live in the same location are included to the new model. Their method is composed of the following steps: (i) link creation, (ii) populating link weights, (iii) graph representation generation, (iv) detecting criminal networks, and (v) evaluating the detection results by using precision, recall, and f-measure calculations. They have performed their experiments on Diyarbakir Drug Network data collected from Diyarbakir, a province in the southeastern Turkey. The dataset consists of 2,552 individuals, and 221 previously known drug dealing networks (DDN). The experiments were conducted on different packages such as R, RODBC, RBGL, igraph, SNA, etc. Out of 221 DDNs, 81 criminal networks were detected. As a result, by using the proposed XSDM model, 0.88 precision, 0.39 recall, and 0.51 f-value results are obtained which are better than the predecessor model, SODM. High precision and respectively low recall indicates that their method is reliable when a criminal network is detected, but it may miss a lot of other criminal networks.

The social networks may also be analyzed on whether if they serve to a specific purpose such as drug trading. Breiger et al. (2014) aim to analyze how terrorist groups are involved in drug trading by generalizing two-mode network analysis method. They used a limited set of 12 distinct properties (6 binary variables and their complements) shown in Table 2 to analyze terrorists' drug trading involvements. They used the data from one of the most extensive public datasets called as Big Allied and Dangerous (BAAD-1) (Asal et al., 2009). This dataset contains terrorist activities and attributes of 395 terrorist groups that were known to be active in 65 countries between 1998 and 2005. The analysis model presented in this paper is an extension (dual-mode) to Ragins Qualitative Comparative Analysis (QCA) techniques (Ragin & Rihoux, 2009). The authors compared their findings with another method called Barycentric correspondence analysis (Greenacre, 1984). The analysis results for different test cases show different accuracies depending on the number of parameters used in the configuration. The highest accuracy rate that an analyst is predicting the drug trading achieved in one configuration is 78% while random guess only provides 9% accuracy since the database shows only 35 out of 395 groups are known to be involved in drug smuggling. With the best configuration which covers the groups' connectivity and similarity based on the given parameters, the highest hit rate of 78% could be achieved. As for the confidence intervals

Table 2. Limited set of 12 distinct properties

| **Variable/Complement** | **Description** |
|---|---|
| **DRUG (D)/drug (d)** | Drug trade involvement (yes/no) |
| **CBRN (C)/cbrn (c)** | Chemical, biological, radio-logical, or nuclear weapon usage (yes/no) |
| **DEGR (D)/degr (d)** | Network degree level (high/low) |
| **TERS (T)/ters (t)** | Control level of territory (strong/weak) |
| **ETHN (E)/ethn (e)** | Founding ideology that emphasized ethnicity (yes/no) |
| **SIZE (S)/size (s)** | Size of the network (large/small) |

for consistency and coverage measures, this work achieves 95% confidence interval when the highest accuracy is achieved.

The terrorist networks can be compared based on the way they commit or organize crimes. Ozgul et al. (2011) aim to develop a similarity metric for terrorist networks in Istanbul, Turkey. This work is based on the ways that crimes are committed and behaviors of terrorist groups with respect to their modus operandi. The authors developed crime ontology based similarity measure (COSM) model. In this model, they used Direct Acyclic Graph for developing crime ontology. After creating the ontology, links are weighted in the network. After all the links are calculated, terrorist network similarity scores are determined based on the model. In order to test their model, authors used Istanbul terrorist networks dataset. This dataset contains attacks committed by 40 terrorist groups in 2005, 2006, and 2007. In their experiments, authors show the similarity scores which are calculated using COSM. The authors evaluated their model using COSM as well as other similarity measure such as Cosine similarity and Jaccard similarity. They report that COSM creates better clustering based on the similarity scores than the other two methods. They also validated their results with domain experts (i.e., law enforcement). However, experts criticized COSM just solely being dependent on a crime and modus operandi similarity. It was recommended to expand COSM's capabilities by adding more attributes such as location and date/time preference of terrorists.

## 2.6 Information Diffusion

The online social networks can be used as online community watch to stop or solve crimes due to fast data propagation in the network. The connectivity and fast information sharing in social networks may be helpful for stopping or solving crimes. According to (Zafarani et al., 2014), there exist four general types of information diffusion. These are herd behavior, information cascades, diffusion of innovation, and epidemics. Herd behavior is related to when individuals act upon actions of all others. In information cascades, people only observe their neighbors. Diffusion of innovations is helpful in terms of observing how innovations spread across a population. Finally, epidemic models are similar to diffusion of innovations where infection is used instead of adoption.

In sympathetic situations, information tend to spread more attentively. Emotions in such cases help people more in getting involved with the problem. Information diffusion on Facebook with respect to carjacking was investigated by (Dingli, 2012). In case that owner learns that his or her car is stolen, the owner does everything which can help to recover the vehicle. Besides reporting the incident immediately to the police, the owner informs his or her own network of friends about the misfortune. Facebook's Graph API (Facebook, 2005) was used in order to investigate a stolen car case and how people create the appeal in Facebook and what information is shared. As every object in Graph API is represented by a unique ID, Mr. L's profile unique ID is used in order to collect data from his wall's properties via the Graph API. A part of his car's plate number is used to search for posts relating to the stolen car case. The results show that the post which propagated most was shared 30 times, and more than 3 comments were made on the photo. On other hand, more than 37% of the users passively followed status messages of their friends and reacted on them by writing comments. Due to the well-connected structure of the nodes in social networks as well as very locally clustered connectivity, it can be said that the vast majority of connections span a short distance by looking at how data propagates. Also, some other findings are (i) posting in specific time periods has a huge impact on the durability of data propagation (how long a message gets attention) and (ii) user-generated posts generate much more interactions in terms of 'shares' than the automated posts.

Another natural scenario where people tend to reach information about their loved ones is in the case of terrorist attacks. After the announcement of such tragic events of which people feel emotional and urge to make sure the ones they love have made it through such events. From this extend, information diffusion for the after effects of a terrorist attack in Woolwich, London, UK was investigated by (Burnap et al., 2014). After a terrorist attack, the size (or volume) of information flow on a topic as well as how long (survival) this topic keeps interest among people are analyzed by studying social, temporal, and content measures. The data was derived from the Twitter based on the retweeting action related to the Woolwich terrorist attack. In this work, different predictive methods are used and compared. For example, Zero Truncated Negative Binomial (ZTNB) regression method (King, 1989) is used because of the plotted distribution gathered using dependent size measure where the Cox regression technique (StatsDirect, 2017) is used in order to model survival. The principal component analysis is also used to reduce the dimensionality of the predictors like social, temporal, and content factors of the tweet. Based on the information diffusion, five hypotheses were proposed to predict the size and survival of information flow. The hypotheses are based on sentiment and tension, online and offline media association, linking content features such as URLs and hashtags, etc. Data collected from Twitter streaming API about Woolwich tweets are subjected to pre-processing and re-coding prior to modeling. In the pre-processing and re-coding stage, a number of tweets are extracted and analysis is performed based on the retweeting action while filtering cases having retweeting size less than five times. This extracted data is modeled using the predictors of size and survival based on the aforementioned five hypotheses.

The analysis revealed significant changes between the independent and dependent variables of size and survival of information flow adjoining a terrorist attack. It has been shown that the main factors for predicting the size and survival of information flow in the event were temporal, social, and content factors. While offline media was a better predictor for the size of information flow, the tension in tweets was determined to be better indicator of survival of information flow. Interestingly, it was concluded that social factors provided the largest amount of variance for the size, and content factors showed the predictive flow of information for the survival. Individual perception varies depending on the interest and burst of enthusiasm which leads to rapid increase or decrease in information propagation as sometimes interest may fade at early stages of tweets which eventually may reduce its lifetime. The lifespan of tweets can be increased by using URLs or hashtags with combination of positive expression and low tension for events like terrorist attacks.

Since social network platforms such as Facebook and Twitter are not the only ones to perform communication, other communication media can also be studied for information diffusion. One of the alternatives is mobile phone operators. Communication between users of such operators can also be analyzed for studying information diffusion.

(Sundsoy et al., 2012) study how the usage of mobile devices increases at the time of an occurrence of an unexpected event such as 22 July Oslo Bombing. The main focus of this paper is to show people's behavior in terms of their communication preferences when they need to reach out their core social networks when it is urgent. The phone data retrieved from only single phone operator in Norway called Telenor is analyzed. Each person in the network represents a node and the number of calls between the nodes represents the degree of closeness of one node to another. The stronger the relation is the more number of calls are expected. This data is analyzed to observe how the number of calls increased during the bombing. It also shows that, as expected, people tend to call their closest relations first, and then call other relations. The data is also analyzed for geographical spreading in the area when the bombing occurred. The results show that the call activation closer to the explosion area increases 4 times when compared to an hour before the bombing. The data does not contain the actual location of the user at the time of the incident, instead it contains the postal code of the subscriber who may not be at the incident scene. This misleading location information might have caused detection of many people who are far from the incident calling many other that are also far from the incident. However, it is hard to verify this result without knowing actual locations of callers and people being called.

Organizational response of the participants and/or coordinators of relief campaigns to a terrorist attack is another important topic. Studying their impact on overall effectiveness on recovery can be analyzed to improve forensics resilience and preparedness. In Schweinberger et al. (2014), authors study the aftermath effects of the disaster that happened in the U.S. on September 11th, 2001 on the World Trade Center. It focused on the inter-organizational networks that emerged in response to the attacks to clarify few questions for helping develop a plan for future responses to such similar unexpected disasters. The three important questions are (i) determining the organizations participating in the disaster response and why/how they contribute, (ii) assessing whether these organizations were given the coordinator role or emerged as coordinators, and (iii) the ability to absorb disturbances to the disaster response. Bayesian framework is utilized to provide answers to the above questions. This framework allows organizations to collaborate and take respective characteristics (covariates) into account to determine whether there is an excess of transitivity using the Bayesian score test. The dataset is generated from the rescue and relief operations by the September 11, 2001 attacks Bevc (2010). It includes 717 organizations with two attributes: the scale of operations categorized as local, state, international, and national, and the type of organization as collective, governmental, nonprofit, and profit. Based on the data collected from different sources, it was evident that small organizations were more collaborative in the disaster response. Network redundancy can help in absorbing the disturbances that reduce the ability to coordinate in the disaster response. So, the evaluation of the network redundancy in the form of transitivity was of prime interest. By evaluating the results performed by the Bayesian framework, it was concluded that small organizations dominated the inter-organizational network and a few dominating organizations behaved as formal coordinators.

There are advantages and disadvantages of having emergent and formal coordinators in disaster response. While they may help to spread information as a point of contact, they may also become bottleneck for the disaster response. As a result, it was concluded that formal coordinators cannot take all burden of disaster response. They should be aware of emergent coordinators and provide necessary resources for them. While network redundancy may reduce disturbances to the response, the key coordinators should be protected and functional for proper to progress of the response. At last, Bayesian framework proved to be simple and outperformed other computational models with less computational time with and without considering covariates. We believe that a similar framework could also be used for resilience and preparedness of forensics investigations as the first responders may collaborate to response to such catastrophic events when they occur.

## 2.7 Influence of Socio-Economic Relationships on Crime and Terror

Drawing a false or unjustified connection between financial aspects and terrorism may lead researchers to incorrect decisions. Therefore, understanding the causes of terrorism in socio-economical relationships are essential if an effective strategy is to be crafted to fight against it. In many cases, the relationship between the crime and the criminal is directly related to financial, demographic, and ethnic background of the criminals. Thus, we distinguished this section to include socio-economical relationships of the criminal and terrorist activities. We provided summaries and brief insights with respect to; (i) crime rates affected by ties among the neighborhoods and the crime and disorder distribution of the individuals along with their mobility between the neighborhoods Hipp et al. (2013); Hipp (2010), (ii) investigation on loss of capital with the influence of social networks Baker & Faulkner (2004).

In earlier studies, criminologists show that network ties among neighborhood residents may impact crime rates. Hipp et al. (2013) propose to simulate network ties speculatively and construct structural network in lack of links between pairs in a social environment. They spatially locate the households in a city. Then, they employ spatial interaction functions and simulate a network of social ties among these residents. After building the simulated network environment, they compute network statistics. Later on, these statistics are further analyzed in order to extract the notions of cohesion and information diffusion which underlie theories of networks and crime.

This study focuses specifically on five cities in the United States (i.e., Buffalo, Cincinnati, Cleveland, Sacramento, and Tucson) for which they aggregated data on the actual occurrence of reported crime events between 2000 and 2002. Their combined census data consisted of information about the number of households in a block along with the number of persons in each household. A limitation of this data was the induction, a degree of uncertainty regarding to the number of persons in these units, which they have placed the extra persons randomly throughout the block. Their results show that these network statistics are robust predictors of crime levels in several cities. The results on information flow present a strong negative effect on crime rates in the models aggregated to small units (blocks), but a weaker effect in models aggregated to block groups.

Demographic characteristics and regional closeness can be correlated in order to observe social patterns and criminal activities. Hipp (2010) proposes a new social distance measure based on the demographic characteristics of people who live in a region. The aim of this work is to analyze the crime and disorder distribution of individuals and mobility between the neighborhoods. The subsample of *American Housing Survey (AHS) - U.S. Census Bureau* (2013) was used to evaluate the social distance between 11 residents of more than 650 blocks for three time periods. Social distance measure is derived from economic differences, education, race/ethnicity, and the life course factors as:

$$sd_{ij} = \frac{1}{K}\sum_{k=1}^{K} |(x_{ik} - x_{jk})\Phi_k| \qquad (5)$$

where $K$ is the number of social factors, $x_{ik}$ and $x_{jk}$ are the values of factor $k$ for each individual $i$ and $j$, and $\Phi_k$ is the salience of factor $k$. After populating a matrix of social distances based on Equation 5, Hipp (2010) analyzed the results using a hierarchical model which consists of two levels: (i) individual parameters and (ii) the micro-neighborhood measures. The social distance had a significant effect on crimes and disorder such that more distant individuals of the society are showing more tendency to get involved in a disorder or a crime.

Influence of social networks on loss of capital has attracted the attention of economic sociologists and white-collar criminologists with opposite views. While economic sociologists claim that social ties benefit the investors, white-collar criminologists believe that social ties yield high probability of loss of capital. Baker & Faulkner (2004) studied a test case on a legitimate business that committed fraud. An oil and gas enterprise, Fountain Oil & Gas Company (Fountain) with its 230 investors, is used as a case study. Data was gathered from the District Attorneys Office of Ventura County, California, and the receiver appointed by the bankruptcy court. The data include detailed records of bank accounts, financial transactions, and lists of all properties (including investor, amount, well, and date) for all investors of Fountain. Later they surveyed the investors regarding whether they lost or earned money and the reason of using social networks (if they used). Total of 72 investors were interviewed which corresponds to 31% of the data. As a result, they concluded that people who do not use social ties (79% probability of loss) or rely on pre-existing social ties or prior investors without using due diligence deal with higher probability of loss (49%) than people who utilize within-network exchange with due diligence (14%). In this study, pre-existing social ties played a beneficial and protective role for investors.

Existing social ties, demographic characteristics, and living nearby expands the human interactions on both sociologically and criminally. Socio-economical aspects on crime and terror reflect invaluable insights on understanding the spread of crime and terror within the neighborhoods, friendships and social networks. Aforementioned studies that are discussed in this section provide an interesting perspective which has to be expanded and investigated more deeply.

### 2.8 Summary

With the increase of terrorist activities around the globe nowadays; it becomes quite clear that researchers should spend more efforts on understanding both organized crime and terrorist networks. The crime and terrorist networks are getting more complex while the diversity of these violent groups are increasing. Moreover, wide and easy availability of the Internet provides convenient connectivity not only for the peaceful citizens of the globe, but also provides a common ground for the crime groups by enabling them to establish worldwide connectivity for their outreach, recruitment and propaganda (Weimann, 2004; Farwell, 2014; Archetti, 2015; Tuttle, 2016). Created by innocent intentions, the online social networks became the playground for these groups where a broad portfolio of criminal activities are created, shared and operated. Moreover, members of these crime groups became experts and professionals within the networks. In order to understand current challenges and develop countermeasures for these criminals and their activities, one should put tremendous amount of effort on understanding the underlying infrastructures of the terrorist networks on the online social networks.

Previous research studies conducted by J. N. Shapiro (2005), Tucker (2008), Bjelopera (2012) and Csermely et al. (2013) have shown that the formation of crime groups has changed from centralized structure to cell networks where a small number of people know each other in a huge crime network where these people act more or less autonomously by the help of digital communication technologies. They have implemented their own networks within the social networks. Moreover, they created their own experts for encryption and secure communication.

The studies we have put together in this section deliver both complementary and diverging insights on the infrastructure analysis of the activities of organized crime and terrorist networks from various perspectives. We believe understanding of these studies will create awareness as well as help to develop technological advancements for possible countermeasures for law enforcement agencies.

## 3. SOCIAL ATTACKS THROUGH ONLINE SOCIAL NETWORKS

We define social attacks mediated through OSNs as attacks that benefit from information available or extracted from OSNs and might result in loss of property, damage to property, loss of life, loss of reputation, financial loss, and criminal activity.

Attacks through OSNs are critical as users may not be aware of the actual person whom they interact with. Moreover, information posted or spread through OSNs may adversely affect people's behavior in their daily life. While increasing recognition and reputation through OSNs, it is possible that some sensitive information is also available to attackers. For example, re-identification and de-anonymization attacks are possible for especially identity-theft purposes (Gao et al., 2011). It is important to understand and categorize these attacks, develop defenses and protection methods before undesired outcomes occur, and identify attackers in case of criminal activities. There are responsibilities of OSN users, OSN service providers, third-party application developers, and friends and relatives of these OSN users to thwart attacks. We explain possible types of attacks in this section.

## 3.1 Anatomy of Attacks through Online Social Networks

We start with explaining how attacks through online social networks are typically analyzed and then we provide our way of understanding and categorizing these social attacks for forensics analysis.

### 3.1.1 Related Work on Categorization of Attacks

There has been significant amount of work related to online social networks. In the literature, in general, the security issues and possible threats are listed and categorized based on their similarities, outcomes, or methods used by these attacks. Joe & Ramakrishnan (2014) discuss a survey of various security issues in online social networks and list a number of security issues resulting from image tagging, email spam attack, social phishing, and sharing day-to-day activities. It is also possible to see other ways of categorizing the list of threats or security issues in the literature.

Attacks are categorized into privacy breaches, viral marketing, network structural attacks, and malware attacks by Gao et al. (2011). Privacy breaches may result from the service provider, other users, or third party applications. Fire, Goldschmidt, & Elovici (2014) categorize threats into classic, modern, combined, and children-oriented threats. Classic threats include malware, phishing attacks, spammers, cross-site scripting (XSS), and internet fraud. Modern attacks include clickjacking, de-anonymization attacks, face recognition, fake profiles, identity clone attacks, inference attacks, information leakage, location leakage, and socware. Combined attacks try to exploit both classic and modern attacks to deploy a more sophisticated attack. Children-oriented attacks target children.

OSN attacks are also categorized into forging nodes/identities and forging social links/connections by (Zhang et al., 2010). Attackers might be an insider from the OSN or outsider to disrupt the OSN. Insider attackers might include OSN service provider, third party application provider, another user, or anyone who gets access to the network structure.

### 3.1.2 Categorizing Attacks for Forensics Analysis

There are numerous types of attacks through OSNs. Some attacks are sophisticated and launched as a series of different types of attacks. Categorizing attacks through OSNs is challenging due to similarities between the ways those attacks are designed and damages caused by them. This also aggravates developing a systematic approach of dealing with these attacks. We start with explaining an OSN in a basic terminology and then categorize attacks. We believe this way of categorizing attacks helps forensics analysts analyze attacks and solve cases.

Attacks benefit from the information available on or extracted from OSNs, and use the services of OSNs. As discussed earlier, an OSN is a graph of vertices and edges where vertices indicate users and edges usually indicate relationships between users. Attackers may try to manipulate users by creating fake user identities and relationships between users. Each OSN user has her or his own social profile which could be visible partially, fully or not visible at all. Attackers may try to extract unavailable information about a user profile by analyzing other attributes of the user, messages of the user, neighbors of the user, and the user profile on other OSNs. OSN users may post or share messages in this network. It should be noted that the content of OSNs is also visible to non-members of OSNs at a level. Messages posted by attackers may contain fake information or illegal content and may not only influence registered OSN users but also passive users of the OSN. OSN service providers provide APIs to increase the usability of their services and publicize anonymized network structure. These APIs or services could also be used by attackers to launch their attacks. Search tools provided by OSNs to find users or search messages may also help attackers target specific population in OSNs and build malicious relationships between OSN users.

Our categorization of social network attacks is based on the OSN structure and information conveyed through OSN networks. After surveying literature on attacks through OSNs, we believe that the attacks should be organized into four levels for an effective and methodical analysis of these attacks:

1. crime or criminal activity,
2. type of attack,
3. scheme of attack, and
4. attack components.

These four levels of attacks are usually mixed in the literature and all labeled as attacks. This makes the analysis of attacks complicated; hence the forensics analysis.

***Crime or Criminal Activity.*** This level typically involves the outcome of an attack. Identity-theft, physical threat, burglary, and child pornography are examples of crimes where online social networks could be used.

***Types of Attacks.*** There are two types of attacks that could result in a criminal activity based on the goal of an attack:

1. information disclosure attack and
2. opinion/emotion influence attack.

In information disclosure attack, the attacker aims to collect personal information about the victim, whereas in opinion/emotion influence attack the attacker tries to influence the user's opinion or emotion about a person, a product, or an event to act in a specific manner. Information disclosure attacks includes both real-world and cyber-space attacks such as profile-squatting and reputation slander using identity theft, phishing, personalized spamming, stalking, bullying, and corporate espionage (ENISA, 2007). In these attacks, attackers may target a specific part of the population based on age and sex. Especially, children are likely to be victims of such attacks. It is possible that opinion/emotion influence attack can take place after information disclosure attack. Opinion/emotion influence attacks could be performed by spreading messages to targeted or untargeted nodes in the network.

***Schemes of Attacks.*** Schemes of attacks correspond to ways of launching various types of attacks with the purpose of information disclosure or opinion/emotion influence which could lead to a criminal activity. We categorize schemes of attacks as social engineering (e.g., spam, phishing) and information discovery (e.g., de-anonymization).

***Attack Components.*** The schemes of attacks utilize four attack components:

1. fake nodes/identities,
2. fake relationships,
3. inappropriate/fake messages/posts, and
4. data crawling.

The attackers may utilize the functionalities of OSNs to create fake identities and build fake relationships with honest users. Fake messages/posts involve publishing or uploading false information on OSNs. However, inappropriate content of a post may also be problematic and considered as a crime (e.g., sexual assault streaming (Van Der Galien, 2017b,a; Moroney et al., 2016)). Data crawling techniques may involve analyzing the profile of a user in a single or multiple OSNs, checking relationships with other users, and analyzing posted messages to collect information about the user. Social engineering scheme is likely to create fake nodes, relationships and messages, whereas information discovery scheme may extract information without creating fake nodes and relationships.

Fake identity attack component is used in identity-cloning attacks. In other words, identity clone attack is actually a scheme of attack that utilizes fake nodes. Information disclosure attacks may be performed by creating fake identities and relationships in the network. For example, direct or indirect information available regarding whereabouts of a person (e.g., classes taken by a student) may help stalkers to determine where the attacker is using data crawling attack component leading to information disclosure attack (Gross & Acquisti, 2005). If a person is away from home and posts messages from a vacation spot, such information gives hint for burglars.

We suggest analyzing criminal activities by asking the following questions:

1. Is the crime result of information disclosure attack, opinion/emotion influence attack or both?
2. What type of scheme is used to launch such attack?
3. Which attack components are used for developing the scheme of attack?

We believe that once these questions are answered effectively, forensics analysis could be performed adequately.

We organize the rest of this section as follows. We firstly cover social engineering scheme and then explain information discovery scheme. We have dedicated a special section for children-oriented attacks. The categories and sub-categories of papers studied in the following sections are provided in Table 3.

## 3.2 Social Engineering Scheme

Social engineering scheme is a critical type of attack as the attacker manipulates the victim to provide the information to the attacker or to act in a specific manner. Social engineering scheme could be used for both information disclosure and opinion/emotion influence attacks. Social engineering scheme may utilize fake nodes, fake relationships, and fake messages. After providing general social engineering attacks, we briefly look into fake identity/messaging, and spam/malware/social bot attacks.

### 3.2.1 General Social Engineering Attacks

Social engineering attacks are categorized into phishing, dumpster diving (attack using discarded items), shoulder surfing (information obtained by viewing over someone's shoulder), reverse social engineering (attacks where victims are lured to contact attackers), waterholing (setting a trap for the victim to visit a compromised web site), advanced persistent threat (usually internet-based espionage attacks), and baiting (malware infected storage medium to be found by victims) (Krombholz et al., 2015). Such attacks have been successful in compromising Google's internal system in 2009 (Zetter, 2010), breaking the RSA security token system in 2011 (RSA, 2011), compromising Facebook in 2013 (Schwartz, 2013), obtaining NY Times employees credentials in 2013 (Perlroth, 2013), and obtaining Paypal's credit card numbers using phishing

Table 3. Overview of studies on social attacks and protection

| Paper | Category | Sub-category |
|---|---|---|
| Irani, Balduzzi, et al. (2011) | social engineering | reverse social engineering |
| Huber, Mulazzani, Weippl, et al. (2011) | social engineering | fake messaging&insecure communication |
| Goga et al. (2015) | social engineering | identity cloning |
| Jin et al. (2011) | social engineering | identity cloning |
| Bilge et al. (2009) | social engineering | profile cloning |
| Garg & Nilizadeh (2013) | social engineering | spam |
| Brown et al. (2008) | social engineering | spam |
| Misener (2011) | social engineering | social bot |
| W. Xu et al. (2010) | social engineering | malware |
| Heymann et al. (2007) | social engineering | spam |
| Viswanath et al. (2014) | social engineering | fake identity |
| Guillory & Hancock (2016) | social engineering | deception |
| Squicciarini & Griffin (2014) | social engineering | deception |
| Griffin & Squicciarini (2012) | social engineering | deception |
| Alowibdi et al. (2014) | social engineering | deception |
| Yan et al. (2011) | social engineering | malware |
| Wagner et al. (2012) | social engineering | social bots |
| Nash et al. (2013) | social engineering | information diffusion |
| Myers et al. (2012) | social engineering | information diffusion |
| Weng et al. (2013) | social engineering | information diffusion |
| Dong et al. (2011) | social engineering | fake relationships |
| (Backstrom et al., 2007) | information discovery | de-anonymization |
| Wondracek et al. (2010) | information discovery | de-anonymization |
| Ding et al. (2010) | information discovery | de-anonymization |
| Nilizadeh et al. (2014) | information discovery | de-anonymization |
| Narayanan & Shmatikov (2009) | information discovery | de-anonymization |
| Sweeney (2002) | information discovery | neighborhood |
| Chester & Srivastava (2011) | information discovery | neighborhood |
| Dey et al. (2012) | information discovery | attribute disclosure |
| M. Li et al. (2014) | information discovery | location disclosure |
| Irani, Webb, et al. (2011) | information discovery | identity disclosure (multiple OSNs) |
| Z. Yang et al. (2014) | information discovery | sybil |
| Seigfried-Spellar et al. (2012) | children-oriented | child pornography (laws) |
| Wolak et al. (2010) | children-oriented | online predators |
| Seigfried-Spellar (2013) | children-oriented | child pornography |
| Huitsing et al. (2012) | children-oriented | bullying |
| Vermande et al. (2000) | children-oriented | bullying |
| Murphy (2012) | children-oriented | children (detecting) |
| Penna et al. (2010) | children-oriented | online gaming |

emails (SocialEngineer, n.d.). Viral marketing attacks may also use social engineering scheme and come in the form of spams, phishing and account attacks. Since significant information is available on online social networks, phishing attacks are four times likely to be successful if it appears to be coming from a known person rather than an unknown person (Jagatic et al., 2007). Designing attacks through fake messages and relationships increases the success of such attacks.

While in traditional social engineering, an attacker may directly communicate with the victim, in Reverse Social Engineering (RSE), the victims are tricked into contacting the attackers (Irani, Balduzzi, et al., 2011). Irani, Balduzzi, et al. (2011) categorize reverse social attacks into recommender-based, demographics-based, and visitor tracking-based attacks. In recommender-based RSE, the goal of the attacker is to influence the recommender of the online social network to be a possible candidate (friend) for the victim. In demographics-based RSE, the attacker creates a fake profile for establishing friendships based on profile similarities (e.g., dating sites). In visitor tracking-based RSE, the attacker visits the profile of the victim. When the OSN displays information about who has visited the victim's profile, the victim may visit the attacker's profile and contact the attacker.

Social engineering attacks can also be categorized as targeted and untargeted (Irani, Balduzzi, et al., 2011). In a targeted attack, the attacker has a specific user as a victim whereas in an untargeted attack the attacker tries to reach as many users as possible. Moreover, attacks can also be categorized as direct and mediated (Irani, Balduzzi, et al., 2011). In direct attacks, the attacker may post messages where the victim can see directly. In mediated attacks, the attacker uses an intermediate agent to bait the user to himself or herself. Irani, Balduzzi, et al. (2011) analyzed possible victims by creating fake profiles. For example, an attractive female photo on a profile is a bait for young people for seeking relationships.

To analyze social engineering attacks, the entities (e.g., environment, attacker, trick and victim) need to be identified and their roles should be investigated (Algarni et al., 2013). Algarni et al. (2013) examine social engineering threats on social networking sites and inquire the questions of "which entities exist and how do they effect social engineering in social networking sites?" It is concluded in their study that the success of social engineering attacks is affected by the characteristics of four main entities: the online social network (the environment), the social engineer (the attacker), the plan and technique (the trick), and the online social network user (the victim).

### 3.2.2 Fake Identity and Messaging Scheme

It is possible for attackers to create fake identities and messages in OSNs. Fake identity could be used for both information disclosure and opinion/emotion influence, whereas fake messages are typically used for opinion/emotion influence. Fake identity attacks create unreal fake nodes and enhance the attack with fake relationships. When attackers want to influence other users by posting false content, they use fake messaging attack component. Some attack schemes could be directly based on attack components.

Deception through providing partially incorrect information in profiles is a special type of fake identity scheme. As DePaulo et al. (1996) describes, deception is a fact of the life. Deception is an act with an intent to mislead others without getting noticed by them, and deception became especially critical when examining social media services in which the borders between victims and deceiving others are mostly hazy (Alowibdi et al., 2014). Users may provide partially true information as well as false data about themselves. Some users may lie about their critical attributes such as gender, age (Alowibdi et al., 2014). Attackers may utilize specific attributes to establish their attacks. One of the major challenges for deception analysis is determining the actual motivation of deception and separating unintentional incorrect information from intentional false information to get benefit from other users. When information provided by OSN users is not true, it is important to determine the reasons for false information. Users may provide false information since personal information could be accessible by undesired people. Another reason is that users would like to have good social images. The most critical false information representation is deceiving others to get unfair benefits or abuse others. Such detection of deception is critical for forensics analysts.

***Fake Messaging.*** Spread of fake news through online social networks caused a man to set fire at a restaurant which he thought that it was a center of child-sex trafficking (Kang & Goldman, 2016). Google and Facebook have started to take steps to deal with fake news since the end of 2016 (Wakabayashi & Isaac, 2017), and only time will tell whether if it will be a success. Validating truth of messages is a very complicated issue. At least, OSN service providers may put effort in labeling messages related to crimes.

Human interactions may also influence how messages are spread in OSNs. Weng et al. (2013) analyze the Yahoo! Meme dataset between 2009 and 2010, where they try to observe different strategies that users adopt for interacting with others. The authors characterize users and use different parameters which are approximated by Maximum-Likelihood Estimation (MLE) (Cowan, 1998). Triadic closure was considered as the major preferred method for forming connections among users. Triadic closure is used when user A follows user B and user B follows user C. Overtime, it is observed that user A follows user C as he or she keeps observing messages from user C. In their paper, the authors observed that, in the early stages, triadic closure is very dominant. However, in later stages, it was observed that information diffusion in the network plays a major role in causing people to interact with each other.

It is important to analyze how information is spread in OSNs to deal with fake message diffusion. Myers et al. (2012) present a model where users can reach information either through social network links or through external sources. Then, their model is applied to URL emergence in Twitter network and evaluated over one month of traces from Twitter. The authors analyze the effect of external sources over information diffusion among people. It is observed that 71% of the information in Twitter is caused by information diffusion, whereas the rest of the information comes from external sources.

OSN providers should provide additional control mechanisms for checking the authenticity of messages. Huber, Mulazzani, Weippl, et al. (2011) devised friend-in-the-middle (FITM) attack (when privacy settings are at high levels to avoid spams) by basing their attack on insecure communication between the user and OSN. They have developed friend injection to become a member of a target network, application injection to obtain profile content, and social engineering to exploit gathered information. They highly recommend that OSN service providers use HTTPS instead of HTTP for all communication exchanges to thwart FITM attacks.

***Identity Cloning.*** Impersonation which is a type of identity cloning attack disrupts the trust among OSN users. Identity cloning attacks can be categorized into celebrity impersonation attacks, social engineering attacks, and doppleganger bot attacks (Goga et al., 2015). In celebrity impersonation attack, the attacker imitates the profile of a famous person by attaining his/her followers. In social engineering attack, the attacker exploits victim's friends to obtain sensitive information about the victim. The doppelganger bot attack is impersonation of an ordinary person and likely to be used to influence other users by providing reviews or other information. Since the profile is realistic, it is not detectable by sybil or spam defenders. They analyze profile similarity, social neighborhood overlap, time overlap between accounts, and differences between accounts to detect impersonating accounts using support vector machines (SVM) classifier.

Attackers developed enhanced methods of identity cloning by carefully setting privacy attributes. Jin et al. (2011) propose detection of faked identities based on attribute similarity in profiles and similarity of friend networks. When attacker creates a fake identity, the attacker may not only duplicate actual victim data but may also benefit from privacy settings. For example, a fake profile may keep the birth date and college as private while the victim profile keep them as public which makes the fake identity more realistic. Moreover, their method also considers the possibility of attacker trying to add a subset of actual friends of the victim, recommended list of friends, excluded set of friends, or combination of these. They apply a number of similarity measures on profiles and friend networks for detecting fake identities. However, the validation is not an easy task.

Identity cloning attacks could also be categorized whether identity cloning is based on the same OSN or multiple OSNs. Bilge et al. (2009) categorize attacks from identity cloning attacks as same-site profile cloning and cross-site profile cloning attacks. In same-site cloning, an attacker duplicates the profile of a user and sends messages to become friends with that user's friends in the same OSN. In cross-site profile attacks, an attacker creates a profile of the victim in another OSN where the victim is not registered yet, and the attacker sends messages to become friends with the victim's friends.

***Profile Deception.*** There are various forms of deceptions. While Turner et al. (1975) categorize deception into lies, exaggerations, half-truths, secrets, and diversionary responses, Utz (2005) takes another approach in computer-mediated communication and categorizes deception into category deception (gender switching), attractiveness deception, or identity concealment. Since true identity of social media users may not be known, it is possible that innocent people may be deceived or abused through misbehavior of deceptive users. Predicting attributes of social media users may help to protect regular users as well as identify perpetrators.

As social networking sites provide more convenient ways of communications and relationships,they also make the deception easier. Walther (1996) points out that people take advantage of computer-mediated communication environments to enhance their self-presentation and inflate other people's view about themselves. In addition, Buller & Burgoon (1996) propose Interpersonal Deception Theory, which defines deception as an intentional act in which senders knowingly transmit messages that aim to foster a false belief or interpretation by the receiver. On the other hand, according to some other researchers, deception is not always intentional (e.g., self-deception) (Gardner & Martinko, 1988), (Jones et al., 1962), (A. Li & Bagger, 2006). Regardless of the deception type, successful forensics investigation requires detection of deception.

Guillory & Hancock (2016) study how people use deception to enhance self-presentation. In their experiment, they investigate how users' social connections affect the way they use deception in their resume profiles. For instance, people feel more comfortable to lie in an environment where others do not know them. However, it is less likely for people to lie in an environment where they have several friends due to the fear of being caught. Moreover, the authors also predict that the users would be reluctant to lie about objective information such as education or work places for enhancing their self-presentation since such objective information could be verified as either truthful or deceptive by their network connections. On the other hand, subjective information such as interests, hobbies, etc. is more difficult to be verified as deceptive. Thus, people would lie more about such subjective information. They have setup experiments where undergraduate students are split into three groups to create their resumes for a consultant position. First group was asked to create their resumes in an MS Word document. Second group was asked to create a completely private resume on LinkedIn that only user and researchers could access. The third group was asked to create a public LinkedIn resume. Just after all the resumes were created, the researchers revealed the actual purpose of the study. The students then were asked to identify all lies on the resumes and asked to provide the truthful version of them. The results of this work show that the number of lies that students told did not differ among those three groups. On the other hand, the type of lies differed among the three groups. The students in the first group tend to lie more about objective information, e.g., responsibilities, than the students in the second group. On the other hand, the students in the third group lied more about subjective information (interests, hobbies). This study shows that forensics investigators need to develop methods or search tools to verify subjective information. This may eventually help them detect criminals/perpetrators who commit crimes via lies.

Identifying cases when online users are likely to deceive or provide false information may help to determine credibility of information. Squicciarini & Griffin (2014) investigate the problem of deception in online sites by analyzing the posts of a forum to determine why and how honest users engage in deceptive activities. A game theoretic approach was proposed to understand users' behaviors based on three cases: (i) peer pressure, (ii) a potential reward, and (iii) comfort level. It is shown that users will play a coordination game to minimize social stress by choosing a level of deception. Prior to examination of forum posts, they conducted an online survey on undergraduate students who have many social network interactions. The purpose was to get an insight on when, how and why users deceive. It was asked whether they would withhold information, tell the truth, or lie about their age, job, GPA, location, phone, etc. The results of the survey show that users' tendency to deceive for data types is not mainly for privacy concerns, but it is highly correlated with the desire to portray a successful social image. Furthermore, it is shown that lying in social networks and withholding information are two different actions. In addition, it is shown that deceit actions are influenced by inner-circle users. In the analysis of the forum posts, 3.7M posts of 21K users were filtered down to 1,400 users who had at least one infraction. 356 users had at least 2 infractions. The infractions are identified by the website and other users' comments about the user's activities. One of the conclusions of their research is that deceptive users are not only hard to identify but they are also influential and affect other users to do deceptive actions. Based on this study, it can also be argued that criminals may affect other people with or without criminal inclination to commit crimes. Therefore, similar study may also be adapted to search for such promotive users.

Social sites request certain information to be provided by their users upon the registration process. Users usually try to minimize the amount of information they provide which they think might reduce their identification on the internet. This however, raises security and privacy issues. Griffin & Squicciarini (2012) develop a deception model which uses game theory to detect users' willingness to release, withhold or lie about their information. In their work, a preliminary model of deception is designed using a game theory. In this model, it is intended to show the complexity of information revelation in OSNs. Misrepresentation in three cases are also examined. These cases are truthful information disclosure, withholding information, and deception. Data was collected from users through surveys. The information that was aimed to help users gain (i) privacy awareness, (ii) attitude toward information withholding and practices, and (iii) attitude toward lies and misrepresentation. The tendency of a user to lie seems to be highly correlated with users' desire to show a good social image as discussed above (Squicciarini & Griffin, 2014), and the tendency is not very related to users' privacy concerns.

### 3.2.3 Spam, Malware and Social Bot Attacks

For categorization of users into spammers, bots, cyborgs, and fake users based on their behavior, the interested readers may study the survey by (Tuna et al., 2016). In this section, we provide a brief overview of how attacks of such users can be launched as spams, malware, and bots rather than focusing on user behavior. Spread of fake messaging through information diffusion also helps success of such attacks.

***Spam Attacks.*** Social engineering attacks can be performed through spams, and there could be correlation between the network attackers or spammers and financial and cultural features of a society (Garg & Nilizadeh, 2013). The main type of attack for spams is opinion/emotion influence. Social spammers may not only launch attacks to spread ads but may also disseminate pornography, viruses or befriend victims, steal personal information (Bilge et al., 2009), and destroy a system's reputation (Lee et al., 2010). Fake messages are the main attack component for spam attacks and can be strengthened by fake identity and relationships.

Social engineering attacks through propagating spams on OSNs such as Twitter and Facebook or through websites such as Craigslist is very common. Even though there are some analysis tools and filtering techniques, in general it is not enough to prevent such spams. Thus, it is crucial to detect the spam messages and spammers. Garg & Nilizadeh (2013) investigate if financial, structural and cultural features of a society explain the actual reasons behind Craigslist-based crimes across 30 American cities. The authors focused on *cars+trucks* category of Craigslist with two subcategories: *by-owner* and *by-dealer*. Furthermore, flagged advertisements by Craigslist were considered, and the ads were classified according to the city. Shapiro-Wilk test (S. S. Shapiro & Wilk, 1965) and null hypothesis techniques suggest that the number of people in the cities are not related with the number and percentage of flagged ads. Also, income per capita is substantially related with the online scam exposure and it suggests that when education level and income increase, the percentage of exposure to online scams also increases.

Context-aware spam is defined as a spam that is highly likely to be clickable due to authentic social connections (Brown et al., 2008). They are usually categorized as relationship-based (using only friendship information), unshared-attribute (using an attribute from one of the parties in the relationship), and shared attribute (using a visible attribute to both parties in the relationship) based attacks. Brown et al. (2008) show that people with private profiles are vulnerable to spam as much as other users.

Similar to spams, especially fraud diffusion may economically hurt many people. Nash et al. (2013) combine social network analysis and diffusion theory to study a fraud that spread in British Columbia, Canada, defrauding 2,285 investors for loss of 240 million dollars. The diffusion of fraud is studied from the point of view of victims who invested in the fraudulent scheme. Illegal innovations are typically studied with respect to the offenders adopting novel crime techniques, not crime victims. The purpose of their study is two-fold. First, they assess whether there is empirical evidence of diffusion in the spread of the fraud. Second, they examine the nature of the relationship between victims and the individuals who influenced them to invest in order to determine the relative importance of opinion leaders who influenced their friends and family into investing, and industry professionals and mass media channels in spreading the fraud. By analyzing the fraud from a social network perspective, they show how the fraud diffused through short chains from multiple anchor points, including victims who unknowingly became agents of diffusion in their own victimization network. The data for this study are drawn from a victim survey of the investment and securities fraud orchestrated by Eron Mortgage Corporation. To identify the number of friends and family who nominated specific individuals as a source of influence, they used in-degree centrality. They have shown that an illegal innovation spreads through word of mouth, change agents, opinion leaders, and network bridges. This study confirms the importance of applying social network methods to diffusion research, including diffusion of criminal innovations in order to take the technical skills of criminals into account.

***Malware and Social Bot Attacks.*** Malware attacks may create accounts, join groups, and post messages on OSNs. Malware software can be used as an attack using worms such as Koobface (W. Xu et al., 2010) and cross-site forgery (Fitzgerald, 2009). Malware attacks can post unwanted messages on the walls of users. Fake messaging is the main tool for such attacks. Similarly, fake identity and relationships may increase the probability of being prone to malware attacks. Social bots may attack OSN users to spread information or to influence opinions of users (Misener, 2011).

## 3.3 Information Discovery Scheme

In information discovery scheme, the major goal is to identify users using the network structure and available information from user profiles. Information discovery scheme is suitable for information disclosure attacks and could be used for identity-theft purposes. OSN users may not be aware how much information could be revealed by just sharing a small amount of information. For example, it is possible to identify demographics information, face photo identification, or social security number for identity theft purposes (Gross & Acquisti, 2005). 5-digit ZIP code, gender, and date of birth information together is satisfactory to identify around 87% of the U.S. population using 1990 U.S. census data (Sweeney, L., 2000), whereas this percentage has gone down to 63% for 2000 U.S. census (Golle, 2006), which is still quite high. Data crawling-based attacks may use information from a single OSN or multiple OSNs and may target specific attributes of victims.

The information discovery scheme may use the available information from OSNs or may create fake nodes and relationships to reveal the identities in a network. We briefly discuss de-anonymization, neighborhood, and attribute closure in this section under information discovery scheme.

#### 3.3.1 De-anonymization using Network Structure and Similarity

Attacks to reveal the privacy of users by identifying their connections are categorized as active and passive attacks (Backstrom et al., 2007). In an active attack, the attacker creates a number of accounts and a subnetwork and tries to communicate with users. When this small network becomes the target network, the locations of victim users can be identified by using the connections in that network. In passive attack, the attacker forms a coalition with a number of users and use their neighborhood information to compromise the privacy of users in their network. Wondracek et al. (2010) use (browser) history stealing to identify group memberships and show that group memberships are satisfactory to uniquely identify or de-anonymize an individual.

De-anonymization attack can also be categorized into mapping-based and guessing based approaches (Ding et al., 2010). While mapping-based approaches use the background knowledge of known users with their IDs to map vertices in released anonymized data, guessing methods may target a user in released data to determine one or more of known users by matching released data against known users. Nilizadeh et al. (2014) show that existing de-anonymization methods can be enhanced by first dividing the network into communities, mapping communities, enriching seeds within communities, and global propagation by mapping remaining nodes. Their divide-and-conquer method is able to map 40% of nodes for Twitter dataset of 90K nodes with 20% edge noise and 16 seeds whereas the base Narayanan-Shimatikov technique (Narayanan & Shmatikov, 2009) could almost not map any.

Sybil attacks are an example of attacks that could be used as a social engineering scheme as well as information discovery scheme. In sybil attacks, the attacker may launch accounts (from as small as seven to thousands depending on the purpose) for de-anonymization (information disclosure) or manipulating voting results (opinion/emotion influence) (Gao et al., 2011). Z. Yang et al. (2014) found that 80% of sybil nodes do not necessarily form its own network and they just try to build relationships with normal users. Their analysis shows that edges between sybil nodes are formed unintentionally by the time. They also found out that sybil nodes without explicit social ties may act together to spread spam.

#### 3.3.2 Neighborhood Attacks

Neighborhood attack is a way of re-identification of a victim by obtaining some information about the neighbors of a victim using anonymized networks which are expected to preserve the privacy of users. In the k-anonymity model, it is expected that an individual cannot be identified with a probability more than $1/k$ in an anonymized network (Sweeney, 2002). Sweeney (2002) proposes a neighborhood extraction and coding method to satisfy k-anonymity with small anonymization cost by changing labels and adding edges.

While most of the studies are done to prevent the attacks on identifying user, analyzing attribute disclosure attack that targets sensitive information has also been needed. Chester & Srivastava (2011) provide $\alpha$-proximity measure to capture the susceptibility of a network for attribute disclosure attack. The social network is considered as an undirected, simple, and vertex-labeled graph.$\alpha$-proximity measure is defined based on the label distribution of vertices in the graph. An algorithm is designed to make the graph $\alpha$-proximal that converts a vertex-labeled graph $G$ into an $\alpha$-proximal graph with a number of additional edges. Basically, their main goal is to make sure that by gaining an access to a specific part of the network an attacker should not be able to predict sensitive information. For example, assume that network nodes are labeled with respect to having some disease or not. If there are sub-graphs having mostly diseases, it will be easy to predict that members of that neighborhood have disease if there is an access to that neighborhood in the graph. To make sure that attackers cannot predict that type of sensitive information, the label distribution of a neighborhood in a graph with labeled nodes should be similar to the entire graph.

Some information from user or neighbor profiles could indirectly identify specific attributes of OSN users. Dey et al. (2012) show that it is possible to estimate the age of users on Facebook by analyzing friend's ages, ages of friends of friends, and so on. They show it is possible to estimate ages with an error of a few years for highly private users who do not share friend list in public. Using high school graduation years of users and their friends reveal helpful information about the birth years of users.

#### 3.3.3 Attribute Disclosure Attacks

Some attributes can be predicted based on the interactions and services provided by OSNs. Location-based social networks that help to find friends with spatial proximity have other types of vulnerabilities and the most important of them is that the location of users can be revealed. M. Li et al. (2014) show that using 30 volunteers and 3-week data, their attack method can detect top 5 locations of a user with high accuracy. The attacker may generate 3 fake anchor locations and applies iterative trilateration method using distances to three locations. The attacker further applies space partition to increase its accuracy.

Since users may have accounts on multiple OSNs, information available (or social footprint) from multiple OSNs may reveal the identity and attributes of a person. Irani, Webb, et al. (2011) show that while physical identification can be successful around 34% using information from a single social network, the physical identification attack may be successful around 90% when information from six or more profiles is used. They also state that the similar success percentages are obtained for password-recovery attacks. Narayanan & Shmatikov (2009) show that it is possible to de-anonymize thousands of users by using information from multiple OSNs which users have registered. They show that users who are verified to have accounts on both Flickr and Twitter can be identified with 12% error rate by just analyzing the network structure without creating dummy sybil nodes.

### 3.4 Children-Oriented Attacks

Children may be victims in social networks or online games. Child abuse and child pornography (CP) is one of the major crimes today, and the online communication does not necessarily protect minors. It is useful to recall the development of CP laws in the US and in the world (Seigfried-Spellar et al., 2012). After that, we present threats for children by briefly looking at profiles of online predators (Wolak et al., 2010; Seigfried-Spellar, 2013), cyberbullying (Huitsing et al., 2012), detecting children (Murphy, 2012), and risk of online games (Penna et al., 2010).

### 3.4.1 Legal Aspects

Seigfried-Spellar et al. (2012) review the development of child pornography (CP) laws and criminalizing by US constitutions and report the issues of the laws as well as the techniques to block such uses and distributions. Since 1950s obscene pornographic materials were accepted illegal by the courts in the USA. And CP is an obscene expression of speech and is not protected by the First Amendment. In 1970s, the congress declared that CP is shifted from "cottage" to "organized abuse." Child Protection Act 1984 considered CP as "lascivious" and illegal when it focuses on the clothed genital region of children despite the lack of nudity. In 1990, some courts decided that private possession of CP is also illegal- not sharing does not make it legal. In 1996, virtual images of children are also considered as CP. CP currently carries maximum 30 years with minimum 15 years of imprisonment. Internet providers who do not willingly report CP incidents are fined $50K for the first time and $100K for subsequent incidents. They are responsible for reporting to NCMEC (National Center for Missing and Exploited Children). The European Union involves 27 countries that prohibit the CP, and there is an argument on whether filtering the sites by tools or removing the whole websites is the right solution.

### 3.4.2 Threats for Children

The harm after using online activities is categorized as harm from content (disseminating pornographic or harmful content about the child), harm by contact (the child targeted for participating in sexual abuse, being photographed, and disseminated), and harm by conduct (uploading harmful material or physically meeting with an adult) by (UNICEF Office of Research, 2011). These threats for children may involve online predators, risky behavior, and cyberbullying (Fire, Kagan, et al., 2014).

***Online Predators.*** While identifying criminals for OSN-initiated crimes, it should be noted that these criminals do not necessarily use fake identities such as representing themselves as children. Wolak et al. (2010) analyzed internet-initiated sex crimes and found out that the victims who were teenagers aged 13 to 17 knew that they were communicating with an adult, could lead to a physical meeting and engage in a sexual activity. Hence, internet-initiated sex crimes are not necessarily a result of online deceptive information. In another study, Seigfried-Spellar (2013) reports some statistics from previous studies on different preferences of child pornography consumers, and investigates measuring the preference of these consumers. Previous statistics by Internet Watch Foundation reports that 69% of child victims are under 10 years old, 24% are under 6 years old, and 4% are under 2 years old. Canadian center for child protection examined about 35K websites. Over half of the images were from under 8 years old. 68.9% of them were sadistic and 83% included females. A survey question was created, and volunteers (CP users) from social networks and websites were invited to rate their content preference on child pornographic images: toddler, poses, violence, etc. Respondents rated their opinions as (i) strongly do not prefer, (ii) do not prefer, (iii) indifferent, (iv) prefer, and (v) strongly prefer. The collected survey data was limited to 2 individuals, which cannot provide statistically significant conclusions. However, there were notable descriptive differences in the types of images preferred by these CP users. In general, one of the respondents appeared to prefer pornographic images of teens, regardless of whether the child was posing in a nonsexual or sexually provocative manner, while the other one preferred a wider-range of sexually explicit image content.

***Cyberbullying.*** Cyberbullying is not rare, as an online survey shows that 12% of parents state that their kids are exposed to cyberbullying based on data from 24 countries (Ipsos, 2012). Unfortunately, cyberbullying may result in committing suicide (Dean, 2012; Pearce, 2013). The use of online social networks at K-12 schools may be analyzed to understand and prevent psychological attacks such as bullying. Huitsing et al. (2012) examine whether the association between victimization and psychological adjustment is moderated by the classroom network position of bullies and victims. They test the effects of the social structure of bullying and victimization networks to see how it is correlated with two psychological adjustment variables: depression and self-esteem. In order to assess the structure of these networks, they use measures of centralization. Degree centralization is used to measure the importance of a student in the network. If the centralization of a classroom is high, then it is more likely that only a few students are central. The authors first examined how victims behave if bullying is common in that classroom. If the victimization is not a common event, then the victims tend to blame themselves. They collected the data via internet-based questionnaires from 78 Finland schools, involving 429 classrooms and a total of 8248 students in grades 3-5 (mean ages in 10-12). They measured depression and self-esteem of the students by asking them questions. Participants responded on a 5-point Likert-type scale. They used *victimization*, *sex*, and *age* as the independent variables describing the students. As for the independent variables describing the classrooms, they used classroom averages of victimization and bullying. In order to collect this data, they used peer nominations. To calculate the classroom centralization of victimization and bullying, they used the normalized degree variance. The degree variance is used to measure the heterogeneity of students. Analyses showed that boys were reported to be somewhat more victimized than girls, whereas girls were somewhat more depressed than boys. No sex differences were found for self-esteem or age. It was also found that boys and younger children were less depressed than girls and older children.

Vermande et al. (2000) also show that such central social network structures of classroom are typical bullying and victimization networks. They examine the dispersal of bullying and victimization in classrooms by examining the centralization of victimization and bullying in the classroom.

***Detecting Children by Images.*** Available images in profiles may help online predators detect children. Murphy (2012) shows that humans can predict the ages of children and distinguish the juvenile and adults; and it is not just common sense but it is decided by preconscious condition of human experience on estimating on the facial features and proportions. Studies show that people can estimate ages of subjects which are in 20-54 with 2.39 years deviation. Old faces age slower than young faces. Thus, estimating ages of old people is more error-prone. In other words, predicting a young person's age is easier than of an old one (e.g., 5 years old will not be guessed as 15 years old, but 35 years old person could be guessed as 45). In this study, several surveys were deployed. The first part of the survey had some images from websites and the task was to identify whether the image belonged to an adult or a kid. At least 90% of respondents predicted correctly. In the second part users had to order (sort) the photos of a person when the person was between 6 years and 16 years old. For two sets of photos, the accuracy is between 74% and 99% for 167 respondents. The respondents commented that they used the size of shoulders and proportion of facial features to decide. Another survey was completed by 107 people and the goal was to estimate the ages of 47 subjects whose ages were between 0 and 25. Survey respondents were quite accurate on predicting the ages. However, the prediction accuracy decreased while reaching the adolescence and it increased again after adulthood. Majority of estimations for under age of 18 was higher than the real age, which implies that if the criminal examiner has a reasonable doubt that the image may not be under age, then the image should not be considered for criminal charges.

***Risk of Online Gaming Networks.*** Protection of children is very important as they are usually interested in playing online games with people who may portray themselves as someone else. Penna et al. (2010) present an approach about safety for children who are playing online games. Their approach proposes to monitor and detect relationships forming with a child in online games. If an offline meeting has been arranged with the child or it has the potential to occur, system will give an alert according to relationship. The system was installed on the child's computer with the local Windows account name provided by parents. Communications and event data were stored in a database. Some indicators and tokens of physical locations or meetings being arranged were searched on the stored data. If messages were detected as suspicious, then system gave alert to the parents via email or other report output. Test data was collected from the researchers' personal game play experience (World of Warcraft: WOW) and online communications (1,191 hours over 3.5 years). The prototype database included 118,230 chat messages and 922 aliases. Real data was used to assist in the creation of the framework including strings, rules, and regular expressions. In their experiments, the prototype successfully achieved its objectives by outputting suspicious scenarios and highlighting them with high ranks.

## 3.5 Summary

There are many types of OSN-initiated attacks covered in the literature. Categorization of attacks in the literature focuses on the similarity of attacks and damages. Such categorization methods have limitations for forensics analysts since these methods do not help forensic analysts develop a systematic way of solving cases. We have proposed categorization of attacks into four levels: crime or criminal attacks, type of attack, scheme of attack, and attack components. Our categorization helps to separate crimes which could be outcome of information disclosure or opinion/emotion influence and could be launched using social engineering and information discovery schemes. These schemes utilize fake nodes, fake relationships, inappropriate/fake messages/posts, and data crawling.

Forensics analysts should be aware of the fact that attackers can establish various forms of attacks. Knowing existing attacks and learning how to prevent damages caused by these attacks may help law enforcement personnel take necessary precautions for new types of attacks. Especially, understanding children-oriented attacks, developing an effective legal framework for OSNs, and monitoring activities of children will help to protect minors in our society.

# 4. FORENSICS ANALYSIS AND DEFENSE METHODS FOR SOCIAL ATTACKS

Protecting OSN users and identifying criminal activities and criminals are important for forensics analysts. The defense methods should make sure that identities, relationships, and messages are authentic. Moreover, users should not reveal private information carelessly and OSN service providers should ascertain that user information is not revealed by de-anonymization attacks. The three methods suggested for dealing with spams by (Heymann et al., 2007) can be generalized for developing defense methods for social attacks: *prevention*, *detection*, and *demotion*. Prevention methods should guarantee that the identities belong to real people, relationships among identities are real, and messages that are posted are verified if needed. Moreover, OSN service providers should not allow information leakage and users should be informed about possible unintentional data closure. Detection methods may allow creation of new identities, relationships, and posting messages. Once these entities are created, the detection system tries to validate them. The demotion method may rank identities, relationships, and messages and assigns a score or rank so that users can protect themselves. While the strength and frequency of security measures such as Captchas may help to deal with some of the attacks (Bilge et al., 2009), user awareness is critical for dealing with attacks (Gao et al., 2011) as there is a trade-off employing security measures and user friendliness.

In order to protect users from social attacks, the security of OSN users should be improved by privacy preservation techniques. In this section, we categorize privacy preservation techniques into wrapper-based, access control, and methods that increase user awareness. We explain specific methods that will help forensics analysts identify fake identities, fake messages, and fake relationships and protect users from malicious users, messages, and relationships. Finally, we provide a brief overview of studies on searching traces of OSN data from devices of OSN users to support forensics investigations.

## 4.1 Privacy Preservation Methods

Private information includes personal sensitive information, the relationships of users, locations that users visit, the location of user homes, temporal data such as time-of-day, a specific date, or a special occasion about users. Attacks based on data crawling methods may reveal private information about users. Quantifying, measuring, and evaluating privacy are some of the most difficult tasks in social networks. Measures for for determining disclosure of privacy (Y. Wang & Nepali, 2013) and analysis of the security behavior of users (keeping patches up to date, cautious website visits, making regular backups, and using anti-virus programs) are needed. Wash & Rader (2011) used mischief and vandal models for protection, and crime and burglar models led to prevention of access control.

Management of sharing private information appropriately is not an easy task and how legal frameworks and OSN service providers address privacy breaches or improper use of personal data should be studied. Then, based on the needs, different types of tools can be implemented to detect and protect users' privacy. After analyzing privacy preservation methods in the literature, we categorize them into three groups: wrapper-based methods, access-control methods, and methods that increase user awareness. Wrapper-based methods ba-

Table 4. Overview of studies on forensics analysis and defense methods

| **Paper** | **Category** | **Sub-category** |
|---|---|---|
| Sayaf et al. (2013) | privacy preservation | access control |
| I. Singh et al. (2012) | privacy preservation | wrapper-based |
| Baden et al. (2009) | privacy preservation | wrapper-based |
| K. Singh et al. (2009) | privacy preservation | wrapper-based |
| Yamada et al. (2012) | privacy preservation | access control |
| Hu et al. (2013) | privacy preservation | access control |
| Paradesi et al. (2012) | privacy preservation | access control |
| Burkholder & Greenstadt (2012) | privacy preservation | user awareness |
| White et al. (2013) | privacy preservation | user awareness |
| Y. Wang & Nepali (2013) | privacy preservation | user awareness |
| Z. Chu et al. (2010) | fake identity | user categorization |
| Viswanath et al. (2014) | fake identity | user categorization |
| Peled et al. (2013) | fake identity | profile matching |
| Vosecky et al. (2010) | fake identity | profile matching |
| Johansson et al. (2013) | fake identity | profile matching |
| Orebaugh & Allnutt (2010) | fake identity | authorship analysis |
| Orebaugh et al. (2014) | fake identity | authorship analysis |
| Orebaugh et al. (2014) | fake identity | gender detection |
| Douceur (2002) | fake identity | sybil |
| Yu et al. (2006) | fake identity | sybil |
| Yu et al. (2008) | fake identity | sybil |
| Cao et al. (2012) | fake identity | sybil |
| Wei et al. (2012) | fake identity | sybil |
| Alvisi et al. (2013) | fake identity | sybil |
| Heymann et al. (2007) | fake messaging | prevention, detection, demotion |
| Zubiaga & Ji (2014) | fake messaging | credibility of messages |
| Sikdar et al. (2013) | fake messaging | credibility of messages |
| Wagner et al. (2013) | fake messaging | user susceptibility |
| Yan et al. (2011) | fake messaging | reaction to infection |
| Yan et al. (2011) | fake messaging | network sanitation |
| Fire, Kagan, et al. (2014) | fake relationship | scoring friends |
| Mutawa et al. (2012) | data tracing | Facebook, Twitter, MySpace |
| Awan (2015) | data tracing | Facebook, Twitter, LinkedIn |
| H. Chu et al. (2014) | data tracing | Facebook, Google Map |
| Cahyani et al. (2016) | data tracing | DropBox, Google Drive, OneDrive |

sically use an additional service and the user utilizes this service while interacting with the OSN. Access control methods determine how and who should access the data of OSN users. Methods that aim to increase user awareness inform users how much information is available about themselves. Next, we discuss these groups in detail.

### 4.1.1 Wrapper-based Methods

Most OSNs today provide some form of privacy controls so that users can protect their shared content from other users. However, these controls are typically not sufficiently expressive and/or do not provide fine-grained protection of information. I. Singh et al. (2012) introduce a new privacy control-group messaging on Twitter with users having fine-grained control over who can see their messages. Specifically, they demonstrate that such a privacy control can be offered to users of Twitter without having to wait for Twitter to make changes to its system. They have designed and implemented Twitsper, a wrapper around Twitter that enables private group communication among existing Twitter users while preserving Twitter's commercial interests. Their design preserves the privacy of group information (i.e., who communicates with whom) both from the Twitsper server as well as from undesired Twitsper users. Furthermore, their evaluation shows that their implementation of Twitsper imposes minimal server-side bandwidth requirements and incurs low client-side energy consumption.

Persona (Baden et al., 2009) enables users to choose personal information to be stored on a decentralized storage system and combines attribute-level encryption with public key cryptography. Lockr (Tootoonchian et al., 2009) (i) separates social networking content from the functionalities of the OSN, (ii) avoids re-use of social data by the OSN for unintended purposes by providing digitally signed social relationships to access social data, and (iii) enables two strangers with a common friend to confirm their relationship using a message encryption based on a social relationship key without exposing themselves to others. FaceCloak (Luo et al., 2009) is another method that hides private information, encrypts it, and stores on a separate server while using alternate fake information to the OSN while providing access to only trusted or authorized users of a person. NOYB (None Of Your Business) (Guha et al., 2008) is a wrapper for OSNs to support privacy of users by encrypting user data and creating ciphertext similar to legitimate data where only authorized users can decrypt and decode the data.

There is no guarantee that these wrapper-based methods do not abuse the information provided. To benefit the functionalities of a third party application, a user may grant access to her personal information regardless of the information is actually needed or used for intended purposes. K. Singh et al. (2009) developed xBook, a wrapper around a third party application, that controls whether information is provided to external entities based on the policies or not. However, it does not check whether information is actually needed by the application or not.

### 4.1.2 Access Control Methods

The conflicts in OSNs may attract different types of information leakage. Yamada et al. (2012) show that friend-list recovery, profile recovery and post recovery attacks are possible. Typical OSNs apply permit-take-precedence meaning that grant authorizations have higher priority than denial rather than denial-take-precedence where denial of access has higher priority than grant access. Such attacks may be performed using friends list or posting messages on the walls of friends.

Multiparty access model (Hu et al., 2013) is proposed for preserving private data by categorizing OSN users as owner, contributor, stakeholder, and disseminator. Whenever an access request is made, the policy evaluation is performed with respect to the policies of relevant parties and the final decision is based on by their decision, sensitivity voting schemes or different types of conflict resolution such as threshold-based, strategy-based or for dissemination control.

It is also important to know how information about a person is revealed as an outcome of search results. Paradesi et al. (2012) point out the problem of the privacy of social networks and provide an interface, Policy Aware Social Miner (PASM), which makes it possible for the users to create policies to guide the consumers how searches about them should be executed. People may share a content within a specific context which is not harmful, but when it is viewed together with other shared content this may picture the person in a different way or may be harmful to this person. In their system, a data consumer enters the user's Facebook username, keywords to search, and the purpose of the search such as employment, commercial, financial, and medical. Users can keep their privacy restricting access, refutation, and filtering. Users may restrict access to their profiles by no-employment, no-commercial, and no-medical options. Refutation allows users to refuse the accuracy or the ownership of information. Users may also use filtering to hide posts having specific words. While this interface can be effectively used for Facebook, it can also be used for Twitter, LinkedIn, Google+, or other social networks.

It is important to compare how technical methods and legal frameworks work in case of improper use of private information on OSNs. Sayaf et al. (2013) address the privacy preservation of users in social software by comparing technical methods (e.g., services provided by OSNs) and legal frameworks (i.e., laws) is important . Access control methods (ACMs) and accountability methods are chosen as the technical methods. The Directive 95/46/EC of the European Parliament and of the Council *'on the protection of individuals with regard to the processing of personal data and on the free movement of such data'* is used to compare with the chosen technical methods. In terms of privacy control, they analyze the control of content, control of audience, linking to the user, and protection of data. Certain approaches of access control that aim at enabling users to control who can access their data are reviewed. ACMs provide more control over the data on personal exemption use than laws; however, ACMs are difficult to use. Nevertheless, ACMs do not restrict enough how OSN providers and third-parties use the information. Accountability and audit approaches may identify misconduct by analyzing the logged interactions of the user. The technical and legal frameworks are not necessarily consistent and it is suggested that ethical framework can be achieved after integrating both legal and technical frameworks.

#### 4.1.3 Methods for Increasing User Awareness of Privacy

The main goal of increasing user awareness is to inform how much information about the user is revealed so that the user can take necessary precautions. For example, Heatherly et al. (2013) show that private information leakage can be reduced by removing some details from the profiles and removing some friendship links. Brown et al. (2008) suggest that friends of users should also make their profiles private. Using images instead of text for sensitive attributes or blocking contextual information from non-friends are some suggestions that could be implemented by

OSN service providers. Measuring the privacy of a social network (Y. Wang & Nepali, 2013) or protecting privacy by concealing users privacy (I. Singh et al., 2012; Sayaf et al., 2013) may require tools to be implemented for protection of users' privacy. Moreover, Irani, Webb, et al. (2011) suggest that account recovery services should indicate how strong account recovery questions are.

Private information (e.g., address, relationships, visited locations) of many users in online review sites are inadvertently disclosed with their reviews. Such private information about a person may be attacked from their online reviews. If the attacker knows the identity of a user, private information disclosed in reviews can be combined with information from other sources. Burkholder & Greenstadt (2012) aim to determine whether such detailed private information is disclosed or not. Types of unstructured and structured information made public by online review sites are characterized, and a privacy check tool is developed. Three keyword categories were used for keyword matching to detect privacy disclosure: *relationship*, *location*, and *temporal*. Relationship keywords are used to catch potentially inadvertent disclosures of relationships or information tied to relationships. Potentially inadvertent disclosures of the location of a user or user's home are identified by location keywords. Temporal keywords are used to find potentially inadvertent disclosures such as location of a user at a given time or how often the user goes to a specific place. Words in the reviews are then compared with keywords in the given keyword list. The Natural Language Toolkit (NLTK) (Steven et al., 2009) is used to scan each review for name recognition. Keyword matches and recognized names are counted per review. The total number of matches for a given keyword category and the total number of matches for individual keywords are computed. The number of named-entity matches, actual named-entity counts, and the percentage of keyword matches by category are provided in their analysis. In their experiments, they used Amazon, Netflix, Yelp, OpenTable, and TripAdvisor as online review sites. For each online review site, 10 items were selected for review scrape and analysis. Items were selected based on popularity because it is assumed that more popular items would have a greater number of lengthy reviews. Using this tool, potentially sensitive review text is annotated by keyword matching and named-entity recognition, and awareness of the privacy threat in online review sites is increased through examples and statistics.

Rather than analyzing attacks on real social networks, a model of an online social network can be built and analyzed against attacks. White et al. (2013) developed a model of an online social network and simulated an attack towards this model. A game theoretic approach is used for the attack to analyze the actions an attacker and how much damage the attacker may cause to the system. The simulation results help to figure out the weakest points in the system. The authors explain four different levels of information sharing: optimal, under-shared, over-shared, and hybrid states. Their game theoretic approach also helps users determine their optimum level of sharing.

Alternative way of identifying how much information could be leaked is to develop quantifiable measures for preserving privacy. Y. Wang & Nepali (2013) proposed three privacy index measures based on three different privacy measurement functions.

- Weighted Privacy Index (w-PIDX) measures an entity's privacy based on known attribute list weight.
- Maximum Privacy Index (m-PIDX) measures an entity's privacy based on

the maximum attribute impact factor of all the known attributes.

- Composite Privacy Index (c-PIDX) measures an entity's privacy based on composite privacy measurement factors.

The efficacy of the PIDXes is evaluated in various user groups by different testing scenarios. In an actor model, an actor is a social entity (e.g., people, organization, etc.) in a social network. Actor has certain characteristics represented as attributes. Each attribute has a different impact on privacy called as Attribute Privacy Impact Factor (APIF). In addition, each attribute is assigned a sensitivity value indicating the sensitivity of information revealed by that attribute. They use probability to describe hidden information and virtual attributes as a reflection of information visibility. Virtual attribute describes a group of attributes behavior and their impact on privacy. A virtual attribute may have significant impact on the privacy of an actor. Among all the attributes in an actor model, if certain attributes are known, privacy might be disclosed. Privacy index is used to describe an entity's privacy exposure factor based on the known attributes. It is computed using privacy impact of known attributes. If it is lower than given threshold $T$, privacy is considered to be preserved.

Preliminary attributes are extracted from social networking sites' personal profile and privacy settings. 20 attributes including name, SSN, birth date, education, etc. are selected for testing, and a privacy impact factor is assigned to each attribute. The three PIDXes are evaluated when known attributes change incrementally for different user groups. While w-PIDX is not good for privacy ranking, m-PIDX is not good at measuring privacy increment change. Tests and analysis show that composite privacy index measures best the privacy for social network actor model.

User privacy has become an important concern in online social networks (OSNs). The users should be notified of inadvertent privacy disclosures by both client-side and server privacy check tools. However, finding the best privacy preservation technique is not a simple task.

## 4.2 Handling Fake Identities

Since true identities of users at social network sites may not be known, it is important to protect users by identifying real and fake user accounts. For example, friendship requests may not always come from a real person. There are different types of users such as fake users, bots that generate automatic messages, and cyborgs (human-assisted bots or bot-assisted humans) (Z. Chu et al., 2010) and spammers (M. Singh et al., 2016; A. H. Wang, 2010). Criminals such as stalkers, sex offenders, etc. may also take the advantage of OSNs and hide their true identities. Therefore detecting these types of users is important for taking secure actions and as well as performing forensics investigations.

Strong Captchas are recommended to avoid automated identity cloning attacks (Bilge et al., 2009). When new accounts are created OSNs may require valid IDs. In case the system is able to identify candidate fake identities, those users could be requested to answer questions from his or her friends properly (Schechter et al., 2009). Forging identity attacks could be minimized if a person's identity is determined with respect to his/her contacts or whom he/she communicates in the network (Zhang et al., 2010). Alternative methods may include analyzing activities of fake identities or their response to the validation process (Jin et al., 2011).

It is important for forensics investigators to identify fake profiles. The first step of fake identity detection is separating human users from bots and cyborgs. Once human users are detected, their true identity may be revealed using multiple online social networks and analyzing contents of messages for authorship. In this section we discuss the studies focusing on categorization of users on OSNs, profile matching by comparing multiple networks, and identifying the true authors of online postings. We briefly look into detection of profile deception. Since some attacks such as de-anonymization attacks are launched by sybil nodes, it is critical for systems to limit the number of fake identities that could be created to thwart such attacks.

### 4.2.1 Categorizing Users

The users may be exposed to different types of interactions on online social networks. In Twitter, there are some tweets posted by automated programs known as bots. In addition, cyborgs have recently emerged on OSNs. In order to assist human users in identifying whom they are interacting with, Z. Chu et al. (2010) focus on the classification of human, bot, and cyborg accounts on Twitter. The authors analyze a collection of over 500K accounts to see the differences among human, bot, and cyborg in terms of tweeting behavior, tweet content, and account properties. According to their analysis, a classification system is proposed which includes the following four parts: (i) an entropy-based component, (ii) a machine-learning-based component, (iii) an account properties component, and (iv) a decision maker. The entropy-based component calculates the entropy (and corrected conditional entropy) of inter-tweet delays of a Twitter user. A lower entropy indicates periodic or regular timing of tweeting behavior, a sign of automation, whereas a higher entropy implies irregular behavior, a sign of human participation. The machine learning component is used to determine whether the tweet content is a spam or not based on the learned text patterns. The account properties component checks properties such as URL ratio, tweeting device makeup, etc. and generates a real-number value for each property. These three components create a set of features that are used to train a classifier. A classification score is assigned to each class and the class having the highest score is used as the indicator of a bot, cyborg, or human. In the experiments, 1K cyborgs, 1K humans, and 1K bots are used for validation. The sensitivity values are obtained as 82.80%, 93.70%, and 94.90% for cyborg, bot, and human categories respectively. The authors also conclude that bots are not misclassified as humans by their system and vice versa.

Behavior-based anomaly detection techniques can also be used for detecting fake identities (Bilge et al., 2009). Viswanath et al. (2014) apply principal component analysis on temporal, spatial, spatiotemporal and combination of these features for detecting anomalous behavior and use it for detecting fake, compromised, and colluding identities without using a priori labeling while achieving low false-positive rates. They have achieved 66% accuracy of detecting anomalous behavior regarding fake, compromised and colluding identities with 0.35 false positives corresponding to 94% of misbehavior.

### 4.2.2 Matching Profiles across Multiple Networks

Nowadays, people have a separate account for each different social network. So, entity resolution for online social networks is important to match profiles of the same person across multiple OSNs using available information in their profiles and networks. In other words, two different user profiles from two different OSNs can be matched based on the features extracted from user profile and networks. There are some studies about profile matching and while some of them just use profiles of users (Peled et al., 2013), some of them use friend networks.

Peled et al. (2013) provide a four-step algorithm for entity resolution based on user profiles: data acquisition, feature extraction, training set construction, and building the model. In the data acquisition stage, irrelevant data are filtered out after crawling user profiles. Feature extraction stage extracts features for comparing two profiles from two different OSNs. Three different types of features are created: *name based*, *general user information based*, and *social network topological features*. Name-based features target name similarity and use the following list of name similarity measures: Soundex, difference, longest common sub-string (LCS), compression, Damareau Levenshthein, Jaro Winkler, n-gram, VMN (Vosecky et al., 2009). General user information based features are compared using locations distance, current employer similarity (n-gram, Damareau Levenshthein, Jaro Winkler), Jaccard, semi vector space model similarity, and vector space model of full profile similarity. Social network topological features include the number of mutual (common) friends and the number of mutual friends of friends. When building the training set in the third stage, users who are members of both networks are identified according to their names by performing a cross reference check between the collected users' profiles. Next, each matched pair of profiles is checked manually using the profile photos to see whether they belong to the same real person. A label (match or a non-match) is given to each checked pair. The remaining users whose names did not match are used to create negative pairs. Then a model is constructed by using the training dataset in the last stage. In the experiment stage, 16,561 user profiles from Facebook and 15,430 from Xing are used. Among the collected profiles, 464 pairs of users are detected as the same user (with the same first and last names) in Xing and Facebook. AdaBoost, Rotation Forest, Random Forest, Logistics Model Tree (LMT), LogitBoost and Artificial Neural Networks are used to build the classifier model. According to the performed experiments in this work, it is observed that even though the name based features are the most important ones, the best results are obtained when 27 features are combined. The best score was obtained using LogitBoost with 0.98 AUC (area-under-the-curve).

To extend the profile matching method, a structural approach is proposed to identify a user based on their friend network in (Vosecky et al., 2010). When some part of a user profile is missing or unavailable, friend network can be useful for profile matching. The total number of mutual friends is determined between two users' friend lists on two different domains. The friends' names are matched and a similarity score is computed using a match name function. All similarity scores above 0.75 are added to obtain a total friends overlap score. Structural approach is combined with profile matching approach. To evaluate their methods, they use a dataset crawled from Facebook and StudiVZ. A multi-layer perceptron (MLP), a logistic regression method (LogitBoost) and an Adaptive Boosting (AdaBoost) are applied for classification and their performance is compared. AdaBoost yields particularly the highest accuracy (93%) in their experiments.

While many people have accounts on multiple online social networks, some people also use several accounts on a single social media site with different user names. It is important to detect users who use multiple aliases for catching terrorists and extremists as they may lead people to commit crimes (see the case of owner of many personae Joshua Goldberg (Justice.gov, 2017)). Johansson et al. (2013) present four different matching techniques to identify users with multiple aliases.

These are (i) string-based, (ii) stylometric-based, (iii) time profile-based, and (iv) social network based matching techniques. String-based matching is based on alias names using Jaro-Winkler edit distance. Time-profile based matching is based on the publishing time of the posts. The time of day is discretized into equal-sized intervals. A feature vector that represents the number of posts per hour is created. Euclidean distance is used to measure the difference of two vectors. Stylometric matching is based on the statistical analysis of writing style such as the length of words and sentences, the number of letters, digits, punctuation and function words. Cosine similarity is used to measure similarity between these feature vectors. Their social network based matching is based on thread or friend information. They try to see how two aliases have similar networks using Jaccard similarity of neighbors, which is the ratio of the number of common neighbors to the number of union of neighbors. For experimental analysis, they used data composed of 9 million documents from Irish website forum (https://www.boards.ie). The aliases who have more than 60 posts are investigated. For analysis, basically two versions of each user are built where one version has the odd numbered posts and the other version has the even numbered posts. If the selected alias appears within top-N rankings ($N = 1$ or $N = 3$), alias matching is considered as success. The experiment results show that time profile-based matching has higher accuracy than stylometric-based matching. The combination of both matching techniques (time profile and stylometric) gives significantly better accuracy results than applying them individually. The accuracies for 50, 250, and 1K users were 70%, 55%, and 43%, respectively. The accuracy increased to 80%, 70%, and 56% for 50, 250, and 1K users, respectively, when top-3 aliases were considered for success.

### 4.2.3 Authorship Analysis

Messages posted on social media give important information about their authors. Each person has distinguishing stylometric features as his or her writeprints. Authorship analysis of messages may be performed for cyberforensics investigations. A variety of computer-aided statistical methods are used to analyze text to determine the most plausible author of a piece of text in authorship identification. Orebaugh & Allnutt (2010) present an Instant Messaging (IM) authorship analysis framework to determine the identity of an author. The proposed framework is composed of data pre-processing, feature extraction, and classification stages. In the data pre-processing stage, the messages are parsed and irrelevant data such as metadata and noise are filtered out. The logs are split into conversations based on the number of messages per conversation. Conversations are fed to the feature extraction module. Instant Messaging feature set consists of 356 features composed of 200 lexical, 154 syntactic, and 2 structural features. The lexical features consist of 20 character-based, 3 word-based, 59 abbreviation frequency, and 42 emoticons frequency features. The syntactic features are composed of 146 function word frequency and 8 punctuation frequency features. The structural features consist of the average number of words and characters per message. For classification, C4.5 decision tree, Naive Bayes, SVM, and k-nearest neighborhood classifiers are evaluated. SVM provided better and consistent results as the number of authors increased from 2 to 19. The best accuracy was 84.4% when all features are used with SVM for 25 authors with 50 messages per conversation.

Behavioral biometrics are acquired over time and can be used to recognize or verify the identity of a person. Orebaugh et al. (2014) use behavioral biometrics-based instant messaging writeprints as cyberforensics input. Author writeprints are created with stylometric features extracted from IM messages and statistical methods are used to analyze and evaluate the writeprints. Similar to the study by Orebaugh & Allnutt (2010), their framework includes developing a stylometric feature set, pre-processing the data, creating writeprints and creating PCA (principal components analysis) visualizations of writeprints. They use a 356-dimensional vector as the feature set including lexical, syntactic, and structural features. Conversations are created as a set of messages $\{M_1, \ldots, M_p\}$ by splitting logs in the writeprint extractor module. These conversations are used to define stylometric features. For each set of messages $\{M_1, \ldots, M_p\}$ of each supplied author $(A_n)$, a writeprint $(W_x)$ is given as the output. A writeprint is a $t$-dimensional vector, where $t$ represents the total number of features. PCA models are used for dimension reduction on the writeprints after normalization and standardization. Gnuplot is used for visualization. In their experiments, personal IM conversation logs with and without author information are used for evaluation. Their plots with different settings for datasets with author information show separate groupings for each author. Plots for unknown author show some separation among the authors.

#### 4.2.4 Detecting Deception

Deception became particularly critical topic when examining social media services where interactions between victims and deceivers are not fully controlled. Alowibdi et al. (2014) analyze gender deception for Twitter since Twitter profiles do not have *gender* attribute. They calculate a gender score and male trending factor on twitter data by using username, first name, and 5 profile colors as feature vectors with hybrid classifiers based on Naive Bayes and Decision trees. The color features are composed of background color, text color, link color, sidebar fill color, and sidebar border color. If a user provides Facebook profile link, the gender of that user is extracted from the Facebook profile. The genders extracted from Facebook profiles are used as ground-truth in order to build the model for gender prediction. Out of 192K profile data collected, to balance male-female distribution, randomly 87.3K male and 87.3K female accounts were selected and then investigated. Gender prediction was performed based on first names (f), usernames (u), and color features (c). Gender prediction accuracies by first name, username, colors and all-combined are 82%, 70%, 75%, and 85%, respectively. Male trending factor is calculated for each user: $m = (w_f * s_f + w_u * s_u + w_c * s_c)/(w_f + w_u + w_c)$ where $w_f$, $w_u$, and $w_c$ are weights of the three gender indicators (first names, usernames, and the 5 color characteristics combined). $w$ weight values represent the difference of accuracy from baseline, $w_f = 82\% - 50\% = 32\%$. A sensitivity value for each factor ($s_f$, $s_u$, and $s_c$) is calculated with respect to the proportion of female vs. male having that feature. For instance, the name "Mary" has a high sensitivity value close to 1.0 for females. Collected male trending values are divided into 5 categories by standard deviation and variance as in Table 5. Weak female and weak male groups are considered as deceptive groups. After manually checking 5% of the data, 24 out of 133 weak male and 11 out of 188 weak female users were determined to be quite deceptive. In addition to gender detection, identifying other attributes such as age, geo-location, and profession has been covered in Tuna et al. (2016).

Table 5. Five groups depending on the male index

| | |
|---|---|
| Strong female | $0 \leq m \leq \mu - 2\sigma$ |
| Weak female | $\mu - 2\sigma \leq m \leq \mu - \sigma$ |
| Weak male | $\mu + \sigma \leq m \leq \mu + 2\sigma$ |
| Strong male | $\mu + 2\sigma \leq m \leq 1$ |
| Neutral | *otherwise* |

#### 4.2.5 Defense Methods for De-anonymization

Trusted certification, resource testing and recurring costs are some methods to deal with de-anonymization and sybil attacks. In trusted certification, where only certified users can join the network, may completely eliminate the sybil attack (Douceur, 2002). Resource testing analyzes network structure, network bandwidth, computational and storage resources, and the number of IP addresses associated for nodes corresponding to real users. Recurring validation such as Captchas while creating new nodes may minimize the creation of sybil nodes (Gao et al., 2011).

SybilGuard (Yu et al., 2006), SybilLimit (Yu et al., 2008), SybilRank (Cao et al., 2012) and SybilDefender (Wei et al., 2012) are some methods to thwart sybil attacks. SybilGuard uses "quotient cuts" (i.e., removing small number of attack edges would lead removing large number of nodes from the graphs) between trustable nodes and sybil nodes by removing a small set of edges (Yu et al., 2006). While SybilGuard limits the number of sybil nodes, it may accept $O(\sqrt{n}logn)$ nodes which could be large if the network is large. Moreover, it assumes that the social network is fast mixing without validation. SybilLimit method shows that actual social networks are mixing within 10 to 20 nodes despite having social communities. SybilLimit reduces the number of accepted nodes by a factor of $\Theta(\sqrt{n})$. SybilGuard first identifies a sybil node and then applies sybil community detection algorithm to identify the sybil community. The number of attack edges could be limited by allowing users to rate their relationships or using an activity network graph that indicates interactions among users rather than only relationships. SybilRank detects sybil nodes in 3 stages: (i) propagating trust via $O(logn)$ power iterations, (ii) ranking nodes based on degree-normalized trust, and (iii) annotating some of the fake nodes. Their main contribution is to be able to rank nodes where lower values indicate the likelihood of being a sybil node. Alvisi et al. (2013) list four commonly used properties of structural graphs: popularity (degree of a node), small distance property (the maximum distance between any nodes in a social graph), clustering coefficient (related to the number of friendships of a user to the number of friendships among friends), and conductance (the ratio of the number of outgoing edges from the graph to the number of edges in the graph). The conductance property makes harder than other properties to act as undetected while not being fully fool-proof.

Do these de-anonymization defense methods make OSNs safe? In some scenarios, researchers develop de-anonymization techniques and then develop possible defense strategies against those types of techniques. Alvisi et al. (2013) point out the Maginot syndrome for OSNs. In other words, while building strong defenses against sybil attacks, sophisticated attacks can be designed that exploit the weaknesses of these defenses. OSNs might be more susceptible for sophisticated attacks or undetected attacks than considered.

## 4.3 Handling Fake Messages

Content-based analysis of messages and postings is needed to detect illegal/inappropriate postings and fake messages and to avoid deception. Since OSNs also allow posting images, videos, and audio, OSN service providers should detect harmful messages and postings and remove or block them in real time. Accounts could be hacked (Zangerle & Specht, 2014) and OSN service providers should detect account hacks by analyzing messages. Moreover, users might be infected by malware and methods are needed to limit the propagation of malware and spam as well as harmful messages. The spread of malicious messages should be prevented if possible. Otherwise, they should be detected, and doubtful messages should be demoted by the system. OSN service providers may also analyze the credibility of messages. Lastly, we briefly overview how the resilience of OSNs could be improved to avoid the harmful results of malicious messages.

#### 4.3.1 Prevention, Detection and Demotion

The countermeasures for spams are grouped as detection, demotion, and prevention by (Heymann et al., 2007). These countermeasures could be generalized to all types of messages. *Detection-based methods* try to identify spams and remove them by methods such as text classification, user behavior analysis, link analysis, manual user identification, and manual moderator identification. Brown et al. (2008) and W. Xu et al. (2010) suggest that OSN service providers may setup fake accounts or “decoy friends” to detect context-aware spam, worms and propagation of possible worms. OSN service providers may limit the number of memberships of accounts to restrict possible fake accounts. *Demotion-based methods* try to lower the rank of spams. *Prevention-based methods* add barriers for spams by changing interfaces or limiting user actions by incorporating Captchas, account fees, proof of work, pay per action, community size limits, and hidden or partitioned input interfaces. Digitally signed messages or browser alerts may be helpful to thwart phishing attacks (Chou et al., 2004).

#### 4.3.2 Analyzing Credibility of Messages

Determining the truth of messages is a challenging task. False messages can easily be spread through OSNs and may result in financial loss or fatal outcomes. OSN service providers should develop effective methods to analyze the credibility of messages.

In the literature, there are some research studies that aim to detect footprints of fake messages and posts. Zubiaga & Ji (2014) study the credibility of information shared on social networks, particularly after natural calamities such as Hurricane Sandy stroke the US. In such serious moments, it is important to have reliable and trustworthy information shared on the Internet. In this study, Zubiaga & Ji (2014) analyze the accuracy of credibility perceptions on different features of witness pictures posted on Twitter during Hurricane Sandy’s impact and aftermath in the East Coast of the United States in October 2012. They analyze credibility ratings provided by Amazon Mechanical Turk2 (AMT) workers on different features of the tweets with those pictures. They evaluate their assessments by identifying real and fake pictures from those features, and then compare with the results obtained by experts. They determine factors and features for assessing credibility. They identify (i) features that can *improve accuracy of credibility* perceptions such as author details currently not readily available on Twitter’s feeds, (ii) features that may *harm accuracy*, such as writing and spelling in tweets, and (iii) other factors that influence users’ perceptions (e.g., repeated exposure of the same hoax received from different authors leads users to mistakenly getting convinced about its veracity).

A total of 32 pictures were analyzed for the Hurricane Sandy. For each picture, 6 ratings were obtained: authority, text plausibility, picture plausibility, corroboration, presentation, and the whole tweet. Since the authors were already given final assessments for each picture by professionals, they were interested in analyzing not only how features were related to each other, but also in studying the accuracy of AMT workers when rating each feature (i.e., verifying tweets). The results show that the whole tweet helped to correctly distinguish real tweets from fake ones in terms of recall.

Another way to check credibility is to analyze messages during and after an event. Sikdar et al. (2013) provide a method on how to construct reliable ground truth values or identify credibility for the microblogging sites such as Twitter. They collected two different datasets from Twitter using the Streaming API (Twitter, 2017) for messages on Hurricane Sandy during and after the hurricane. The first dataset consists of a large group of people and many of them may not know each other. Also, since this dataset was collected while the hurricane was developing, there exists some uncertainty about the hurricane in the first dataset. On the other hand, the users of the second dataset were more connected, and they were more knowledgeable about the event. The second dataset was collected by using the keyword "#occupysandy," which is a relief effort to distribute resources. Therefore, the second dataset is more credible than the first dataset. In order to perform the analysis they collected samples of tweets from each dataset. They conducted two surveys using MTurk users to evaluate the credibility of messages with 6 levels of credibility opinion (can't-answer, strongly non-credible, moderately non-credible, neutral, moderately credible, and strongly credible). The first survey includes messages, source image, and number of retweets. The second survey includes messages only.

In these surveys, they also ask users to annotate the tweets regarding the credibility of the message newsworthiness, and the credibility of the user. For the ground truth selection, they look at the correlation between credibility judgments in both surveys. They find that there is no correlation between credibility of messages in both surveys. However, the credible newsworthy messages of the first and second surveys are highly correlated. They also find out credibility of messages, credibility of users, and newsworthiness of tweets are highly correlated in the first survey.

Psychologists have shown that human behavior deviate from its normal behavior when a person lies. Bhaskaran et al. (2011) focus on the eye movements to detect the deceits. In their experiments, 132 human subjects were interviewed using two types of questions: (i) questions involving basic conversations and (ii) critical questions which involve reward or punishment. These interviews were video recorded. The subjects also filled out questionnaires to indicate if they lied or not. In order to collect the data for their experiments, the eye pupil regions were detected using image processing techniques to learn the behavioral features. The authors used Bayesian model of eye movements in the basic conversation period to learn the normal behavior of a person. The rest of the interview was divided into time slots and each slot was evaluated using the trained model by computing its log-likelihood. Deviations from normal behavior were tested using the log likelihoods of each time slot. Deceit and non-deceit videos showed very distinctive patterns. In the testing part of the process on the 40 subjects, they had 82.5% accuracy in deceit detection.

Understanding the dynamics of deception and deceptive behaviors in online social networks is both crucial and significantly challenging. It requires new tools, softwares, and new perspectives in applied analyses to improve the findings. The expansion in the online social media world continues with the help of diverse set of tools which aim to ease their usage for the end users. Hence, aforementioned methods, tools and models are particularly important for law enforcement and forensics investigators to detect certain activities (e.g., deception, fake messages, spam) which may eventually evolve to criminal ones.

#### 4.3.3 Improving Resilience to Infection

To avoid infection of user systems by malicious malware or messages, it is important to check the susceptibility of users to such infection. If user systems are infected, the infected users may inform neighbors as soon as possible or the OSN provider may develop the network in a way that malicious messages are quarantined before being propagated to other parts of the network.

***Susceptibility of Users to Infection.*** It may be possible to determine the susceptibility of users for social bot attacks and whether users are infected or not (Wagner et al., 2012). Wagner et al. (2012) extract a wide set of linguistic, network (hub and authority score, in- and out-degree, clustering coefficient) and behavioral (conversational variety, conversational balance, conversational coverage, lexical variety, lexical balance, topical variety, topical balance, informational variety, informational balance, informational coverage, temporal variety, temporal balance, temporal coverage) features to determine the susceptibility of users. They found out that susceptible users tend to use OSN for a conversational purpose and communicate with many different users using more social words and show more affection than non-susceptible users on Twitter. This shows that the active users of OSNs are vulnerable to social bot attacks than passive users.

***Reaction to Infection.*** The attacker may consider the number of neighbors as well as the number of neighbors to be active after being infected for effective malware propagation (Yan et al., 2011). User-oriented defense methods are categorized as active and reactive (Yan et al., 2011). In active defense, when a machine recovers from infection, it informs neighbors and the neighbor getting this message becomes immune when it becomes active. If a neighbor transmits this message to its other neighbors then it is considered an active defense otherwise it is reactive defense. Detecting malware as quickly as possible is critical; however, active defense is still effective even with low detection rate.

***Network Sanitation.*** OSN service providers may analyze URLs in messages from or to sanitized nodes selected based on the degree of a node, active nodes, or activities and make sure that the messages from/to these nodes are malware free (Yan et al., 2011). In another approach, Yan et al. (2011) propose preventive containment where the graph is partitioned into islands and it is guaranteed that malware does not propagate from one island to another island.

### 4.4 Handling Fake Relationships

OSNs introduce different types of relationships, trust strength and interaction intensity (Zhang et al., 2010). The relationships could be bi-directional as friendship relationships or one-directional as in follower or fan relationships. Trust strength indicates how much a user trusts another user on a general or specific topic. Interaction intensity involves both quantity and quality of interactions between users. Social link forging (fake relationship) attacks can be avoided or minimized by the trust strength and interaction intensity of relationships, and moreover, since OSN users may also meet face-to-face, the true identities of (malicious) users are likely to be revealed. The trust path between users can help to determine the position of a node.

OSN service providers should help to find trustable friends. Dong et al. (2011) provide a method of finding secure friends by letting users utilize virtual IDs and digital signatures for authentication, applying proximity prefiltering by eliminating profiles that are less likely to be friends and enhancing this scheme with homomorphic cryptography to validate social coordinates and proximity results.

Fake relationships can be minimized by increasing user awareness. Bilge et al. (2009) suggest that users should be more alert when accepting friendships. On the other hand, OSN service providers may look at the relationships between users before suggesting friendships beyond basic email look-up (Irani, Balduzzi, et al., 2011). Similar to detecting spam messages, a honeypot account can be set up to detect requests from other users (Irani, Balduzzi, et al., 2011). Irani, Balduzzi, et al. (2011) suggest that the usage of Captcha for incoming friendship requests could be an additional barrier for reverse social engineering attacks.

OSN service providers should enhance OSNs by providing tools that help users accept or reject some friendship or relationships requests. Such tools will improve user's security and privacy. A Social Privacy Protector (SPP) software (Fire, Kagan, et al., 2014), which is developed for Facebook, includes three protection layers to enhance the user privacy. In the first layer, the software identifies a friend of a user who might pose a threat. Then this friend's exposure is restricted to the user's personal information. A credibility score is assigned to each friend in their friend list according to the strength of the connection between the user and his/her friends. The strength of each connection is computed based on different features such as the number of common friends and the number of pictures tagged in together. A simple heuristics or a more sophisticated machine learning algorithm is used to compute the credibility score. Friends with lower scores have the higher likelihood of being fake profiles. In the second layer, Facebook's basic privacy settings are expanded based on different types of social network usage profiles. The user's internet activity and the number of applications installed on the user's Facebook profile are observed. The number of installed applications that have access to user's private information is provided as a warning to the user. The last layer includes the HTTP Server to analyze, store, and cache software results for each user. The HTTP server is a part of the SPP responsible for connecting the SPP Firefox Add-on to the SPP Facebook application. Also, some statistics are provided about Facebook user privacy settings, which were obtained by the SPP Add-on. These statistics demonstrate how Facebook users' personal information is exposed to fake profile attacks and third party applications. Experiment data is collected on each SPP's user and his/her links. The links of the user include different types of Facebook friends: (i) friends who are recommended with restrictions by the application due to a low connection-strength score, (ii) friends who were restricted by the user, (iii) friends who were not restricted by the user, (iv) friends who were not recommended with restrictions by the application, and (v) friends who were deliberately chosen to be restricted. Then, the collected data is used to learn more about Facebook user's privacy settings, and classifiers are built using collected data to identify Facebook profiles with higher likelihoods of being fake.

## 4.5 Data Tracing for Forensics

In many criminal cases, searching forensics artifacts in the traces of social networks data is essential for the success of an investigation. Hence, it is studied in a variety of contexts: collecting and creating new data representations from different data sources by crawling different websites (Huber, Mulazzani, Leithner, et al., 2011) or tracking the logs in system files (Turnbull & Randhawa, 2015). Due to the common usage of mobile devices for social network access, mobile devices have also been studied for recovering traces of data for various activities (Mutawa et al., 2012; H. Chu et al., 2014; Awan, 2015; Norouzizadeh Dezfouli et al., 2016; Cahyani et al., 2016).

### 4.5.1 Searching Traces of Data

Many online social networks provide APIs for data collection. However, these available APIs and other tools that are used for crawling data from online social networks cannot guarantee the granularity and efficiency of the data for proper analysis. Social Snapshots (Huber, Mulazzani, Leithner, et al., 2011) is developed to collect social network data with the aim of harvesting more data compared to the other available tools such that it can be utilized more effectively for searching and analyzing online evidence. It is designed as a hybrid system of a web browser and an online social network mediator application having six modules: *social snapshot client*, *automated web browser*, *social network mediator*, *hijacking*, *digital image forensics*, and *analysis*. *Social snapshot client* module is responsible to trigger the process by using user credentials. *Automated web browser* module takes care of simple communication with the social network. *Social network mediator* application is responsible of harvesting data over the targeted online social networks API. *Hijacking* module is a basic network sniffer package that collects legitimate HTTP cookies of an online social network. *Digital image forensics* module is responsible to match images gathered from online social networks with their original source. Finally, *analysis module* is a parsing module that separates and cleans the collected data. Social Snapshots is evaluated with the participation of 25 human volunteers. Facebook data has been collected through the mediator application via using the Graph API. The mediator application collected 9,800 Facebook account elements per test subject on average. 238 profiles have been crawled and on average, 22 phone numbers, 65 instant messaging accounts, and 162 email addresses were collected. The datasets of Social Snapshots consist of profile information such as user data, private messages, photos, etc. as well as the associated metadata such as internal timestamps and unique identifiers which are quite significant for digital forensics investigation and security research.

As mentioned earlier, accessing social networks via mobile devices is very common, thus tracing data in mobile devices may provide help to track activities of people for forensics analysis. Mutawa et al. (2012) focus on conducting forensic analyses on three widely used social networking applications: Facebook, Twitter, and MySpace. The tests were conducted on three popular smartphones: BlackBerrys, iPhones, and Android phones. In a very similar study, Awan (2015) conducted forensic analyses on Facebook, Twitter and Linkedn by using four type of smarthpones: BlackBerry, iPhone, Android and Windows phone. In both studies, the goal is to find the artifacts of social network applications on these devices for forensic analyses. The tests consisted of installing the social networking applications on each device, conducting common user activities (e.g., uploading photos,

Table 6. Social Network and mobile application data extraction for mobile forensic analyses

| | **Mutawa et al. (2012)** | **Awan (2015)** | **H. Chu et al. (2014)** | **Cahyani et al. (2016)** |
|---|---|---|---|---|
| Device / Operating System | IOS(IPhone), Android (Samsung), Blackberry | IOS(IPhone), Android (Samsung), Windows (Nokia), Blackberry | PDA | Android (Samsung), Windows (Nokia) |
| Application | Facebook. Twitter, MySpace | Facebook, Twitter, LinkedIn | Facebook, Google Map | Dropbox, Google Drive, OneDrive |
| Information collected | **IOS - Facebook** (user and contact details, profile picture URLs, photo uploads, posted comments, friends with active chat sessions), **Android - Facebook** (user and contact details,profile picture URLs, photo uploads, created albums, pictures viewed with app, mailbox/chat messages),<br>**IOS - Twitter** (for user and people followed: user names, profile picture URLs, tweets posted with time stamps) **Android - Twitter** (for user and people followed: user names, profile picture URLs, posted tweets and photos, other activity information such as which device is used to tweet),<br>**IOS - MySpace** (User name/password, posted comments with timestamps) **Android - MySpace** (user name/password, cookies & cache files) | **IOS&Android - Facebook** (Profile name, contact details, images)<br>**Windows - Facebook** (chat, videos, images)<br>**IOS&Android - Twitter** (profile name, followers, tweets)<br>**Windows - Twitter** (tweets)<br>**IOS&Android - LinkedIn** (profile name, contact details)<br>**Windows - LinkedIn** (Contact details) | **Facebook** (email)<br>**Google Maps** (geo-location) | **Android - Dropbox** (documents, sounds, images, videos)<br>**Android - Google Drive** (documents, sounds, images, videos)<br>**Android - OneDrive** (sounds, images, videos)<br>**Windows - Dropbox& OneDrive** (images)<br>**Windows - Google Drive** (NA) |
| Comment | No data could be gathered from BlackBerry Devices | No data could be gathered from BlackBerry Devices | | In windows device, only image files were successfully uploaded |

posting comments, and sending emails) using each application, acquiring a forensically sound logical image of each device, and then performing manual forensic analysis on each acquired logical image. The forensic analyses were aimed at determining whether activities conducted through these applications were stored on the device's internal memory. If so, the extent, significance, and location of the data that could be found and retrieved from the logical image of each device were determined. The results show that no traces could be recovered from BlackBerry devices. However, iPhone, Android and Windows phones store a significant amount of valuable data that could be recovered and used by forensic investigators.

Cloud storage apps and and common usage of these apps in mobile devices make researchers to perform forensic analyses on these devices. Cahyani et al. (2016) studied the role of mobile forensics in terrorism investigations involving the use of cloud apps: Dropbox, Google Drive and Microsoft OneDrive. These cloud storage apps may be used to store incriminating evidence while communication apps may be used to exchange voice and video messages. By using forensic techniques, one could potentially recover information such as chat logs, multimedia files, contact lists, and geo-tagged data, which can then be used to determine the chain of events, and identify their associates. A dataset comprises 37 files made up of document, audio, picture, video and executable files uploaded through these apps in two devices with different operating systems: android and windows phone. These files and their artifacts on these devices were collected using a combination of physical, logical and manual acquisition methods. As Table 6 shows, most of the data were downloadable by client apps. In additon to that, all sender-receiver account artifacts were gathered by accessing the related local SQLite database files in these mobile devices. Usernames, passwords, and the list of shared files shared could be extracted. Finally, H. Chu et al. (2014) worked in different environments and extracted geo-location info recorded by Google Maps and Facebook application from a personal digital assistant (PDA) device. Data are gathered from the image of the memory while the device is on.

In the traditional digital forensics investigations, examiner almost always analyzes the log files collected from computer and network systems. In addition to the analysis of user created data in mobile devices discussed above, analysis of log files produced by mobile devices is also significantly important for social network investigations. If a log of operations by a user is maintained properly, this log could be analyzed for suspicious data access. Turnbull & Randhawa (2015) implemented a tool, ParFor (Parallax Forensics), which provides a unified representation of different data sources to support a higher level reasoning. The hierarchy of files, directories and file systems, user accounts and system information, system events, and user events are examined to create an automated rule based extraction system. For example, if the login and logout time slots are kept for a user, then the files which seem to be changed by this user out of this time slots can be reported as a suspicious activity.

### 4.5.2 Discussion

Successful information collection is one aspect of searching traces of data. However, there are cases when too much insignificant or unnecessary information could be collected from mobile devices. Huber, Mulazzani, Leithner, et al. (2011) propose to create a timeline of the metadata they gathered from Social Snapshot tool so that they can be used in forensic analyses ( e.g., userA uploaded an image, userB commented, userA responded and shared a video, etc.). This would notably help if the goal is to track time based events. However, generic queries without a time frame would require too much effort to analyze the metadata especially for actively used social network profiles. There is a need to filter out some of the information from the metadata which will not really effect the analyses, but will speed up the manual examination.

Despite significant studies collecting social network information from mobile devices, the upgrades and updates made by OSN service providers and operating systems of mobile devices make reusability of previous work challenging. Huber, Mulazzani, Leithner, et al. (2011) could get phone numbers and contact details for Facebook profiles by using Facebook graph API and other crawling websites. Mutawa et al. (2012) and Awan (2015) were able to get different social network usernames/passwords, contact lists, and some related activities with their timestamps from the internal memory of a smartphone. H. Chu et al. (2014) managed to track the user geolocation coordinates with timestamps related to a facebook account from a PDA device. Cahyani et al. (2016) spotted the files and extracted sender-receiver information from cloud apps in mobile devices. Turnbull & Randhawa (2015) were able to get only MSN messenger contact lists from the user directories located in hard disks. These studies show that even though a lot of information can be gathered from mobile devices, forensic examination of smartphones for Social Network applications is still challenging due to constant updates in social network software, smartphone software (operating system), and hardware updates. As Mutawa et al. (2012) report, the results for different types of applications for different models of smartphones may change dramatically after updates by the operating system (OS) or the social network application. For instance, SnapChat application used to keep pictures on the device in an encrypted form. However, decryption key was hard coded in the application. It was possible to decompile the application and retrieve the key in order to decrypt the files to access the pictures. In the recent version of the application however, encryption key is not located in the application, hence it is currently infeasible to decrypt the pictures. We would then suggest to distinguish the OS and hardware first, and then repeat the experiment for each case separately rather than collecting a result for a specific device or OS. If the experiment can be done repeatedly with different models of the devices and different versions of mobile OSs, weakness and strength of a particular device and OS can be identified for further investigations.

## 4.6 Summary

It is important to protect OSN users from attacks and develop tools for forensics analysis to detect criminal activities and identify criminals. OSN service providers usually employ defense mechanisms through authentications (e.g., Captchas), security and privacy settings, internal mechanisms to detect fake profiles or attacks, and reporting malicious users or behaviors. Developing effective methods for protection and deploying discouraging strategies for attackers is critical for trustable OSNs.

Forensics analysts may use authorship analysis, profile matching, and user categorization techniques to identify fake users on OSNs. Fake messages could be analyzed by assigning credibility scores and demoted, blocked and removed before they cause actual harm. OSN service providers may inform about the susceptibility of users and develop methods for quarantining sections of the network to avoid propagation of malicious messages. OSN service providers may also consider assigning scores to OSN users and such scores could be used for accepting or rejecting friendship requests by users. These scores could also be helpful to identify fake identities for forensics analysts. Forensics investigators may analyze the traces of OSN data for collecting information about criminals. Those traces may reveal the profiles used OSNs and the possible terror networks.

We have added a few more recommendations to those suggested by (Fire, Goldschmidt, & Elovici, 2014) for OSN users:

1. not disclosing unnecessary personal information,
2. proper management of privacy and security settings,
3. rejecting friendships from strangers,
4. removing unnecessary or idle friendships,
5. installing internet security software,
6. uninstalling third party applications,
7. informing about malicious users,
8. not disclosing your location,
9. not trusting your OSN friends and their friends,
10. being alert to abnormal behavior of friends and relatives using OSNs, and
11. monitoring online activities of your child.

Based on our analysis of research studies, OSN service providers should

1. avoid or limit auto-creation of fake identities,
2. increase anonymity of the network (e.g., by deploying fake accounts),
3. increase the security of messaging,
4. increase the strength and frequency of security measures (e.g., Captchas),
5. validate users with IDs or phone numbers beyond email,
6. increase user awareness (e.g., informing about how much privacy is disclosed),
7. develop effective and simple access control methods,
8. develop measures for trustability of nodes, relationships, and messages (e.g., deploy methods to analyze anomalous behaviors),
9. incorporate user validation techniques to help users to confirm friendships,
10. deploy methods for reporting malicious activities,
11. analyze content of messages and posts for inappropriate material, and
12. label messages whether verified or not and provide links from external resources.

Forensics investigators need tools for

1. categorizing OSN users (i.e., bots, cyborgs, human users),
2. distinguishing fake identities and identifying true owners of fake profiles,
3. profile matching within a single OSN or across multiple OSNs,
4. authorship analysis for OSNs,
5. detecting fake messages,
6. analyzing credibility of messages,
7. determining authenticity of relationships, and

8. searching traces of relevant OSN data to solve cases.

Among the proposed defense techniques, wrapper-based privacy preservation techniques are designed with good will for honest users; however, they provide another level of shield for attackers. These defense techniques should not strengthen the strategies of attackers. Another issue is that it is not possible to trust the defense methods as they are usually detection based methods and hard to verify their performance or accuracy since it is not known how many fake identities, relationships, and messages exist in real OSNs. Typically, defense methods consider possible ways of attacks and are designed to protect from those types of attacks. Margirot syndrome (discussed above) could be a possibility for OSN defense methods.

# 5. OPEN RESEARCH ISSUES AND CONCLUSION

One of the most important novelties of the new millennium is online social networks with the advanced and simple communication technologies through easily accessible devices such as smartphones. In the past, one of the major limitations of forensics research was availability of data for forensics research. With publicly available online social network data and supporting APIs from online social network sites, there have been vast amount of data to be processed and analyzed. Since OSNs are available worldwide, they enable sharing of knowledge and experiences for protecting against terrorist attacks, criminal acts, and deception at different parts of the world.

In this paper, we provide an overview of some techniques that are useful for social network forensics. We have covered identifying network structures, determining the ranks of individuals in these networks, spatial behavior of crime/terror networks, determining the similarity of terror networks, information diffusion, a socio-economic factors leading to crimes and joining crime and terror networks. We have analyzed social attacks through online social networks and organized them into four levels: criminal activity, type of attack, scheme of attack, and attack component. We have explained how attacks are launched using various schemes of attacks and attack components. With the availability of affordable devices (such as $35 Amazon Kindle), children enjoy the entertainment and information provided by the Internet. Nowadays, most online games also provide a chat messaging system built in them and children may be exposed to people with indecent behavior. We have analyzed a few papers for the protection of our children. For protecting OSN users and forensics analysis, we explained privacy preservation methods, user categorization, authorship analysis, profile matching, credibility of messages, detecting deception, proposing safe relationships, and finding traces of OSNs on mobile devices for forensics analysis.

Despite numerous and various approaches for analyzing crime, terror networks and social attacks, these do not deter malicious people to abuse OSNs and their users. Our major goal in this paper was to present the available approaches for forensics analysis to prevent crimes, discourage malicious people abusing these online social networks, solve crimes and detect criminals and increase user awareness of possible risks of using OSNs, and recommend actions for forensics investigators, OSN service providers, and OSN users. Fundamental components of forensics analysis are likely to be the same but techniques should evolve as OSNs improve and provide more functionalities. We list possible issues with current approaches and propose future research as follows:

1. *Temporal Analysis of Crime and Terror Networks.* Many research studies aim analyzing network structure and similarities of such networks as mentioned in Section 2. There are a few studies focusing on spatial analysis of these networks. However, temporal evolution of these networks have not been studied in depth. Some major crime and terrorist networks are using OSNs for recruitment. The temporal growth analysis of these networks should be analyzed and necessary precautions and analysis should be done before these crime and terror networks grow.

2. *Accuracy of Detection.* Forensics analysts should be aware that the accuracy of detecting criminals, networks, fake identities, social attacks, etc. could be misleading. We do not have complete, accurate and ground-truth information about attacks and networks that utilize OSNs. The analysis is usually performed around known attacks and networks. Therefore, the accuracy of analysis is mostly related to the precision of detection and it may not consider recall. In other words, there may be many undetected or unknown attacks, fake identities, and networks on OSNs. Some research studies consider possible attack scenarios and verify whether OSNs are stable or not towards such attacks. Attackers may come up with their own schemes of attacks. Attackers are assumed to have different behaviors, connections, and profiles which could be separated from regular users. It is also possible for sophisticated attackers to imitate a real person. Forensics analysts should be cautious when relying on accuracy of detection methods.

3. *The Maginot Syndrome.* It is likely that the defense systems for OSNs might be useless for sophisticated or undetected attacks. Developing enhanced and powerful defense methods may not make OSNs safe. Especially considering the fact that some possible attack scenarios are developed by researchers, there is a high risk of vulnerability of OSNs for undetected and sophisticated attacks.

4. *Trusting OSNs and Fake Identities.* There are users who do not trust OSNs as well as users who are likely to believe what they see on OSNs. It is hard to gain complete trust of users in presence of many possible attacks. Important attacks are built around fake identities. The following suggestion could be considered by OSNs for users who do not want to use their real identities but still would like to utilize services of OSNs. OSNs may allow users to create accounts with verifiable information but represent them in different ways. Such representations along with a note or sign should be clearly visible to other users such that information is not true profile or information of the person. This resembles virtual card numbers offered by credit card companies where credit card companies issue a temporary card number for limited use and the credit card companies can verify such information. This may help to resolve some issues coming through fake identities. It is known that not all fake identities are for deception or attacks. In these cases, actual user information should be validated before creating those identities.

5. *OSN-generated Recommendations.* Recommendations by OSNs also help attackers. Users should be notified of how recommendations are generated.

Especially, users should be informed about OSN recommended relationships. For example, OSN recommends user B to user A and then user A may want to connect with user B. OSNs should let user B know that this interaction happened with the help of their recommendation system.

6. *Certifying Bots and Third-party Applications.* Bots or programs that collect information or interact with users must be certified by the OSN. Users should be able to determine bots serving for enhancing functionalities of OSNs as well as third-party programs.

7. *Considering OSN Trolls Disseminating Violence.* It is quite common that Internet trolls target OSN users in order to lead them to criminal activities. Currently available methods related to user identification and authorship detection could be used to create new tools for law enforcement's disposal in order to hunt OSN trolls causing criminal and terrorist activities.

8. *Building a Consortium of OSN Service Providers.* There are attacks designed using multiple OSNs. OSN service providers should form a consortium and determine what types of facilities and services should be provided to forensics investigators. Such consortium may also determine analyzing accounts on multiple OSNs and information disclosure using multiple OSNs.

9. *Increasing Security Features of OSNs.* OSN service providers must enhance security of messaging and validation of creating new accounts. Especially, for avoiding creation of fake nodes, strong Captchas are highly recommended. OSN service providers should analyze identity cloning, fake identity creation, detect spams and malware, and inform relevant users.

10. *Instant Notification of Malicious Acts and Posts.* OSN service providers must enable tools for instant notification of criminal activities. Law enforcement agencies should have branches for dealing crimes through OSNs and should cooperate with OSN service providers for emergency cases.

11. *Training OSN Users.* OSN service providers may provide sample scenarios and tests about how much information is revealed based on current privacy and security settings, acceptance of friendships from strangers, and publishing careless location information. OSN service providers should encourage installing internet security software. It would be interesting to give scores to OSN users based on how alerted they are to possible attacks or attribute closure. Similar measures may also be provided how much information is visible about themselves. OSN service providers should provide methods about actions to be taken when abnormal behavior or posts are detected by users.

12. *Scoring Identities, Relationships and Messages.* OSN service providers should give scores for OSN users (identities), relationships, and messages. OSN service providers should provide mechanisms to demote or prevent fake (false) identities, relationships, or messaging. Honest OSN users should be promoted. Methods that are applied for online product reviews could be adapted in the short run, and specific OSN methods could be developed in the long run. OSN service providers should try to verify content of messages or let users ver-

ify content of messages by providing links to external resources. Unverified messages could be labeled as 'unverified' so that users may not completely trust the content of messages. OSN service providers should develop tools for content-based analysis of multimedia posts and streaming.

13. *Accountability in Criminal Cases.* There should be legal frameworks for OSN service providers on how to act and what to provide to law enforcement professionals. Laws should be enhanced to protect rights of OSN users. OSN service providers should act accordingly while protecting their users.

14. *Not Protecting Attackers through Defense Methods.* We have observed that some techniques that are developed to protect the privacy of OSN users actually provide another level of protection for attackers. While security methods are developed and used, OSN service providers should also analyze how those methods could be used by attackers as well.

15. *Developing Methodological Approaches for Solving OSN Crimes.* Forensics analysts should be provided a general handbook of steps taken for specific crimes through OSNs. In this paper, for social attacks, we suggest studying attacks at four levels starting from the crime at the highest level, the type of attacks at the second level, scheme of attack at the third level, and attack component at the bottom level. After having answers for these four levels, forensics investigators should develop a systematic approach regarding how to deal with each possible type of crime.

We believe that social network forensics will be one of the hot areas very soon as the required tools become available. We organized the papers in a way that researchers who are interested in social network forensics have a starting point to explore and develop their research methodologies for their domains. With this, we also aim to help digital forensics tool developers in order to create state of the art tools via provided approaches. There were many other research studies that could benefit social forensics research. We have decided to keep the topics focused rather than disturbing readers' attention with related areas. Researchers who are interested in social network forensics are highly recommended to explore these related research areas.

# REFERENCES

Algarni, A., Xu, Y., Chan, T., & Tian, Y.-C. (2013). Social engineering in social networking sites: Affect-based model. In *Internet technology and secured transactions (icitst), 2013 8th international conference for* (pp. 508–515).

Alowibdi, J. S., Buy, U., Yu, P. S., Stenneth, L., et al. (2014). Detecting deception in online social networks. In *Advances in social networks analysis and mining (asonam), 2014 ieee/acm international conference on* (pp. 383–390).

Alvisi, L., Clement, A., Epasto, A., Lattanzi, S., & Panconesi, A. (2013, May). Sok: The evolution of sybil defense via social networks. In *2013 ieee symposium on security and privacy* (p. 382-396). doi: 10.1109/SP.2013.33

*American Housing Survey (AHS) - U.S. Census Bureau.* (2013). https://www.census.gov/programs-surveys/ahs/.

Archetti, C. (2015). Terrorism, communication and new media: explaining radicalization in the digital age. *Perspectives on Terrorism*, *9*(1).

Asal, V., Rethemeyer, R. K., & Anderson, I. (2009). Big allied and dangerous (baad) database 1-lethality data, 1998-2005. *Codebook. Project on Violent Conflict (PVC), University at Albany, State University of New York (h ttp://the data. harvard. edu/dvn/dv/start/faces/study/Study Page. xhtml.*

Awan, F. A. (2015, Dec). Forensic examination of social networking applications on smartphones. In *2015 conference on information assurance and cyber security (ciacs)* (p. 36-43). doi: 10.1109/CIACS.2015.7395564

Backstrom, L., Dwork, C., & Kleinberg, J. (2007). Wherefore art thou r3579x?: Anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of the 16th international conference on world wide web* (pp. 181–190). New York, NY, USA: ACM. doi: 10.1145/1242572.1242598

Baden, R., Bender, A., Spring, N., Bhattacharjee, B., & Starin, D. (2009, August). Persona: An online social network with user-defined privacy. *SIGCOMM Comput. Commun. Rev.*, *39*(4), 135–146. doi: 10.1145/1594977.1592585

Baker, W. E., & Faulkner, R. R. (2004). Social networks and loss of capital. *Social Networks*, *26*(2), 91–111.

Bevc, C. (2010). *Working on the edge: A study of multiorganizational networks in the spatiotemporal context of the world trade center attacks on september 11, 2001* (Unpublished doctoral dissertation). Department of Sociology, University of Colorado, Boulder.

Bhaskaran, N., Nwogu, I., Frank, M. G., & Govindaraju, V. (2011). Lie to me: Deceit detection via online behavioral learning. In *Automatic face & gesture recognition and workshops (fg 2011), 2011 ieee international conference on* (pp. 24–29).

Bilge, L., Strufe, T., Balzarotti, D., & Kirda, E. (2009). All your contacts are belong to us: Automated identity theft attacks on social networks. In *Proceedings of the 18th international* *conference on world wide web* (pp. 551–560). New York, NY, USA: ACM. doi: 10.1145/1526709.1526784

Bjelopera, J. P. (2012). *Organized crme: An evolving challenge for us law enforcement.* DIANE Publishing.

Bora, N., Zaytsev, V., Chang, Y.-H., & Maheswaran, R. (2013). Gang networks, neighborhoods and holidays: Spatiotemporal patterns in social media. In *Proceedings of the 2013 international conference on social computing* (pp. 93–101). Washington, DC, USA: IEEE Computer Society. Retrieved from `http://dx.doi.org/10.1109/SocialCom.2013.21` doi: 10.1109/SocialCom.2013.21

Breiger, R. L., Schoon, E., Melamed, D., Asal, V., & Rethemeyer, R. K. (2014). Comparative configurational analysis as a two-mode network problem: A study of terrorist group engagement in the drug trade. *Social Networks*, *36*, 23–39.

Brown, G., Howe, T., Ihbe, M., Prakash, A., & Borders, K. (2008). Social networks and context-aware spam. In *Proceedings of the 2008 acm conference on computer supported cooperative work* (pp. 403–412). New York, NY, USA: ACM. doi: 10.1145/1460563.1460628

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication theory*, *6*(3), 203–242.

Burkholder, M., & Greenstadt, R. (2012). Privacy in online review sites. In *Security and privacy workshops (spw), 2012 ieee symposium on* (pp. 45–52).

Burnap, P., Williams, M. L., Sloan, L., Rana, O., Housley, W., Edwards, A., ... Voss, A. (2014). Tweeting the terror: modelling the social media reaction to the woolwich terrorist attack. *Social Network Analysis and Mining*, *4*(1), 1–14.

Cahyani, N. D. W., Ab Rahman, N. H., Xu, Z., Glisson, W. B., & Choo, K.-K. R. (2016). The role of mobile forensics in terrorism investigations involving the use of cloud apps. In *Proceedings of the 9th eai international conference on mobile multimedia communications* (pp. 199–204). ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Cao, Q., Sirivianos, M., Yang, X., & Pregueiro, T. (2012). Aiding the detection of fake accounts in large scale social online services. In *Proceedings of the 9th usenix conference on networked systems design and implementation* (pp. 15–15). Berkeley, CA, USA: USENIX Association.

Chen, H., Atabakhsh, H., Petersen, T., Schroeder, J., Buetow, T., Chaboya, L., ... others (2003). Coplink: Visualization for crime analysis. In *Proceedings of the 2003 annual national conference on digital government research* (pp. 1–6).

Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: a general framework and some examples. *Computer*, *37*(4), 50–56.

Chester, S., & Srivastava, G. (2011, July). Social network privacy for attribute disclosure attacks. In *Advances in social networks analysis and mining (asonam), 2011 international conference on* (p. 445-449). doi: 10.1109/ASONAM.2011.105

Chou, N., Ledesma, R., Teraguchi, Y., &

Mitchell, J. C. (2004). Client-side defense against web-based identity theft..

Chu, H., Yang, S.-W., Hsu, C.-H., & Park, J. H. (2014). Digital evidence discovery of networked multimedia smart devices based on social networking activities. *Multimedia Tools and Applications*, *71*(1), 219–234. Retrieved from http://dx.doi.org/10.1007/s11042-012-1349-9 doi: 10.1007/s11042-012-1349-9

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2010). Who is tweeting on twitter: Human, bot, or cyborg? In *Proceedings of the 26th annual computer security applications conference* (pp. 21–30). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/1920261.1920265 doi: 10.1145/1920261.1920265

Cowan, G. (1998). *Statistical data analysis.* Oxford University Press.

Csermely, P., London, A., Wu, L.-Y., & Uzzi, B. (2013). Structure and dynamics of core/periphery networks. *Journal of Complex Networks*, *1*(2), 93–123.

Dean, M. (2012, October). *the story of amanda todd.".* The New Yorker. Retrieved from http://www.newyorker.com/online/blogs/culture/2012/10/amandatodd-michael-brutsch-and-free-speechonline.html

Delle Fave, F. M., Jiang, A. X., Yin, Z., Zhang, C., Tambe, M., Kraus, S., & Sullivan, J. P. (2014). Game-theoretic patrolling with dynamic execution uncertainty and a case study on a real transit system. *Journal of Artificial Intelligence Research*, *50*, 321–367.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of personality and social psychology*, *70*(5), 979.

Dey, R., Tang, C., Ross, K., & Saxena, N. (2012, March). Estimating age privacy leakage in online social networks. In *2012 proceedings ieee infocom* (p. 2836-2840). doi: 10.1109/INFCOM.2012.6195711

Didimo, W., Liotta, G., Montecchiani, F., & Palladino, P. (2011). An advanced network visualization system for financial crime detection. In *Pacific visualization symposium (pacificvis), 2011 ieee* (pp. 203–210).

Ding, X., Zhang, L., Wan, Z., & Gu, M. (2010, Sept). A brief survey on de-anonymization attacks in online social networks. In *2010 international conference on computational aspects of social networks* (p. 611-615). doi: 10.1109/CASoN.2010.139

Dingli, A. (2012). Using social networks to solve crimes: A case study. *International Journal of Virtual Communities and Social Networking (IJVCSN)*, *4*(2), 18–29.

Dong, W., Dave, V., Qiu, L., & Zhang, Y. (2011, April). Secure friend discovery in mobile social networks. In *2011 proceedings ieee infocom* (p. 1647-1655). doi: 10.1109/INFCOM.2011.5934958

Douceur, J. R. (2002). The sybil attack. In *Revised papers from the first international workshop on peer-to-peer systems* (pp. 251–260). London, UK, UK: Springer-Verlag.

ENISA. (2007, November). *security issues and recommendations* *for online social networks.* Position Paper; https://www.enisa.europa.eu/publications/archive/security-issues-and-recommendations-for-online-social-networks/at_download/fullReport.

Farwell, J. P. (2014). The media strategy of isis. *Survival*, *56*(6), 49–55.

Fire, M., Goldschmidt, R., & Elovici, Y. (2014, Fourthquarter). Online social networks: Threats and solutions. *IEEE Communications Surveys Tutorials*, *16*(4), 2019-2036. doi: 10.1109/COMST.2014.2321628

Fire, M., Kagan, D., Elyashar, A., & Elovici, Y. (2014). Friend or foe? fake profile identification in online social networks. *Social Network Analysis and Mining*, *4*(1), 1–23.

Fitzgerald, N. (2009). *new facebook worm - don't click da button baby!".* blog, Nov. 2009, [online] Available: http://fitzgerald.blog.avg.com/2009/11/new-facebook-worm-dont-click-da-button-baby.html.

Frank, O. (2001). Statistical estimation of co-offending youth networks. *Social Networks*, *23*(3), 203 - 214. Retrieved from http://www.sciencedirect.com/science/article/pii/S0378873301000405 doi: http://dx.doi.org/10.1016/S0378-8733(01)00040-5

Gao, H., Hu, J., Huang, T., Wang, J., & Chen, Y. (2011, July). Security issues in online social networks. *IEEE Internet Computing*, *15*(4), 56-63. doi: 10.1109/MIC.2011.50

Gardner, W. L., & Martinko, M. J. (1988). Impression management in organizations. *Journal of management*, *14*(2), 321–338.

Garg, V., & Nilizadeh, S. (2013, May). Craigslist scams and community composition: Investigating online fraud victimization. In *Security and privacy workshops (spw), 2013 ieee* (p. 123-126). doi: 10.1109/SPW.2013.21

Goga, O., Venkatadri, G., & Gummadi, K. P. (2015). The doppelgänger bot attack: Exploring identity impersonation in online social networks. In *Proceedings of the 2015 acm conference on internet measurement conference* (pp. 141–153). New York, NY, USA: ACM. doi: 10.1145/2815675.2815699

Golle, P. (2006). Revisiting the uniqueness of simple demographics in the us population. In *Proceedings of the 5th acm workshop on privacy in electronic society* (pp. 77–80). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/1179601.1179615 doi: 10.1145/1179601.1179615

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.*

Griffin, C., & Squicciarini, A. (2012). Toward a game theoretic model of information release in social media with experimental results. In *Security and privacy workshops (spw), 2012 ieee symposium on* (pp. 113–116).

Gross, R., & Acquisti, A. (2005). Information revelation and privacy in online social networks. In *Proceedings of the 2005 acm workshop on privacy in the electronic society* (pp. 71–80). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/1102199.1102214 doi: 10.1145/1102199.1102214

Guha, S., Tang, K., & Francis, P. (2008). Noyb: Privacy in online social networks.

In *Proceedings of the first workshop on online social networks* (pp. 49–54). New York, NY, USA: ACM. doi: 10.1145/1397735.1397747

Guillory, J. E., & Hancock, J. T. (2016). 6 effects of network connections on deception and halo effects in linkedin. *THE PSYCHOLOGY OF SOCIAL NETWORKING*, 66.

Heatherly, R., Kantarcioglu, M., & Thuraisingham, B. (2013). Preventing private information inference attacks on social networks. *IEEE Transactions on Knowledge and Data Engineering*, *25*(8), 1849–1862.

Heymann, P., Koutrika, G., & Garcia-Molina, H. (2007, Nov). Fighting spam on social web sites: A survey of approaches and future challenges. *IEEE Internet Computing*, *11*(6), 36-45. doi: 10.1109/MIC.2007.125

Hipp, J. R. (2010). Micro-structure in micro-neighborhoods: a new social distance measure, and its effect on individual and aggregated perceptions of crime and disorder. *Social Networks*, *32*(2), 148–159.

Hipp, J. R., Butts, C. T., Acton, R., Nagle, N. N., & Boessen, A. (2013). Extrapolative simulation of neighborhood networks based on population spatial distribution: Do they predict crime? *Social Networks*, *35*(4), 614–625.

Hu, H., Ahn, G. J., & Jorgensen, J. (2013, July). Multiparty access control for online social networks: Model and mechanisms. *IEEE Transactions on Knowledge and Data Engineering*, *25*(7), 1614-1627. doi: 10.1109/TKDE.2012.97

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). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/2076732.2076748` doi: 10.1145/2076732.2076748

Huber, M., Mulazzani, M., Weippl, E., Kitzler, G., & Goluch, S. (2011, May). Friend-in-the-middle attacks: Exploiting social networking sites for spam. *IEEE Internet Computing*, *15*(3), 28-34. doi: 10.1109/MIC.2011.24

Huitsing, G., Veenstra, R., Sainio, M., & Salmivalli, C. (2012). it must be me or it could be them?: The impact of the social network position of bullies and victims on victims adjustment. *Social Networks*, *34*(4), 379 - 386. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0378873310000377` doi: http://dx.doi.org/10.1016/j.socnet.2010.07.002

Husslage, B., Borm, P., Burg, T., Hamers, H., & Lindelauf, R. (2015). Ranking terrorists in networks: A sensitivity analysis of al qaeda's 9/11 attack. *Social Networks*, *42*, 1–7.

Ipsos. (2012, January). *one in ten (12world say their child has been cyberbullied, 24has experienced same in their community"*. The New Yorker. Retrieved from `http://www.ipsos-na.com/news-polls/pressrelease.aspx?id=5462`

Irani, D., Balduzzi, M., Balzarotti, D., Kirda, E., & Pu, C. (2011). Reverse social engineering attacks in online social networks. In T. Holz & H. Bos (Eds.), *Detection of intrusions and malware, and vulnerability assessment: 8th international conference; dimva 2011, amsterdam, the netherlands, july 7-8, 2011. proceedings* (pp. 55–74). Berlin, Heidelberg: Springer Berlin Heidelberg.

Irani, D., Webb, S., Li, K., & Pu, C. (2011, May). Modeling unintended personal-information leakage from multiple online social networks. *IEEE Internet Computing*, *15*(3), 13-19. doi: 10.1109/MIC.2011.25

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007, October). Social phishing. *Commun. ACM*, *50*(10), 94–100. doi: 10.1145/1290958.1290968

Jiang, A. X., Jain, M., & Tambe, M. (2014). Computational game theory for security and sustainability. *Journal of Information Processing*, *22*(2), 176–185.

Jin, L., Takabi, H., & Joshi, J. B. (2011). Towards active detection of identity clone attacks on online social networks. In *Proceedings of the first acm conference on data and application security and privacy* (pp. 27–38). New York, NY, USA: ACM. doi: 10.1145/1943513.1943520

Joe, M. M., & Ramakrishnan, D. B. (2014). A survey of various security issues in online social networks. *International Journal of Computer Networks and Applications*, *1*(1), 11–14.

Johansson, F., Kaati, L., & Shrestha, A. (2013). Detecting multiple aliases in social media. In *Proceedings of the 2013 ieee/acm international conference on advances in social networks analysis and mining* (pp. 1004–1011).

Jones, E. E., Gergen, K. J., & Davis, K. E. (1962). Some determinants of reactions to being approved or disapproved as a person. *Psychological Monographs: General and Applied*.

Justice.gov. (2017). *Florida man arrested for illegal distribution of information relating to explosives.* Retrieved from `https://www.justice.gov/opa/pr/florida-man-arrested-illegal-distribution-information-relating-explosives` ([Accessed: 2017-02-15])

Kang, C., & Goldman, A. (2016, December). *in washington pizzeria attack, fake news brought real guns".* The New York Times. Retrieved from `https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html?_r=0`

Kean, T. H., Hamilton, T., Ben-Veniste, B., et al. (2004). *The 9/11 commission report: Final report of the national commission on terrorist attacks upon the united states,(2004).* WW Norton & Company Ltd, NY.

King, G. (1989). Event count models for international relations: Generalizations and applications. *International Studies Quarterly*, *33*(2), 123–147.

Krebs, B. (2013). *Spy service exposes nigerian 'yahoo boys'.* Retrieved from `http://krebsonsecurity.com/2013/09/spy-service-exposes-nigerianyahoo-boys/`

Krebs, V. (2002). Uncloaking terrorist networks. *First Monday*, *7*(4).

Krombholz, K., Hobel, H., Huber, M., & Weippl, E. (2015). Advanced social engineering attacks. *Journal of*

*Information Security and applications*, *22*, 113–122.

Lee, K., Caverlee, J., & Webb, S. (2010). Uncovering social spammers. In *Proceeding of the 33rd international acm sigir conference on research and development in information retrieval-sigir* (p. 435).

Li, A., & Bagger, J. (2006). Using the bidr to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment*, *14*(2), 131–141.

Li, M., Zhu, H., Gao, Z., Chen, S., Yu, L., Hu, S., & Ren, K. (2014). All your location are belong to us: Breaking mobile social networks for automated user location tracking. In *Proceedings of the 15th acm international symposium on mobile ad hoc networking and computing* (pp. 43–52). New York, NY, USA: ACM. doi: 10.1145/2632951.2632953

Lindelauf, R. H., Hamers, H., & Husslage, B. (2011). Game theoretic centrality analysis of terrorist networks: the cases of jemaah islamiyah and al qaeda.

Luo, W., Xie, Q., & Hengartner, U. (2009, Aug). Facecloak: An architecture for user privacy on social networking sites. In *2009 international conference on computational science and engineering* (Vol. 3, p. 26-33). doi: 10.1109/CSE.2009.387

Mazumder, A., Das, A., Kim, N., Gokalp, S., Sen, A., & Davulcu, H. (2013, Sept). Spatio-temporal signal recovery from political tweets in indonesia. In *Social computing (socialcom), 2013 international conference on* (p. 280-287). doi: 10.1109/SocialCom.2013.46

McBride, M., & Caldara, M. (2013). The efficacy of tables versus graphs in disrupting dark networks: An experimental study. *Social Networks*, *35*(3), 406–422.

Misener, D. (2011, March). *rise of the socialbots: They could be influencing you online.”*. Web.

Moroney, J., Jones, T., & Palumbo, A. (2016, September). *2 found guilty in snapchat rape case sentenced.* NECN. Retrieved from `http://www.necn.com/news/new-england/Sentencing-for-2-Found-Guilty-in-Snapchat-Rape-Case-393665381.html`

Morselli, C., Gigure, C., & Petit, K. (2007). The efficiency/security trade-off in criminal networks. *Social Networks*, *29*(1), 143 - 153. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0378873306000268` doi: http://dx.doi.org/10.1016/j.socnet.2006.05.001

Murphy, C. A. (2012). The role of perception in age estimation. In *Digital forensics and cyber crime* (pp. 1–16). Springer.

Mutawa, N. A., Baggili, I., & Marrington, A. (2012). Forensic analysis of social networking applications on mobile devices. *Digital Investigation*, *9, Supplement*, S24 - S33. Retrieved from `http://www.sciencedirect.com/science/article/pii/S1742287612000321` (The Proceedings of the Twelfth Annual {DFRWS} Conference12th Annual Digital Forensics

Research Conference) doi: http://dx.doi.org/10.1016/j.diin.2012.05.007

Myers, S. A., Zhu, C., & Leskovec, J. (2012). Information diffusion and external influence in networks. In *Proceedings of the 18th acm sigkdd international conference on knowledge discovery and data mining* (pp. 33–41).

Narayanan, A., & Shmatikov, V. (2009, May). De-anonymizing social networks. In *2009 30th ieee symposium on security and privacy* (p. 173-187). doi: 10.1109/SP.2009.22

Nash, R., Bouchard, M., & Malm, A. (2013). Investing in people: The role of social networks in the diffusion of a large-scale fraud. *Social Networks*, *35*(4), 686 - 698. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0378873313000567` doi: http://dx.doi.org/10.1016/j.socnet.2013.06.005

Nelson, B., Phillips, A., & Steuart, C. (2015). *Guide to computer forensics and investigations.* Cengage Learning.

Nilizadeh, S., Kapadia, A., & Ahn, Y.-Y. (2014). Community-enhanced de-anonymization of online social networks. In *Proceedings of the 2014 acm sigsac conference on computer and communications security* (pp. 537–548). New York, NY, USA: ACM. doi: 10.1145/2660267.2660324

Norouzizadeh Dezfouli, F., Dehghantanha, A., Eterovic-Soric, B., & Choo, K.-K. R. (2016). Investigating social networking applications on smartphones detecting facebook, twitter, linkedin and google+ artefacts on android and ios platforms. *Australian journal of forensic sciences*, *48*(4), 469–488.

Orebaugh, A., & Allnutt, J. (2010). Data mining instant messaging communications to perform author identification for cybercrime investigations. In *Digital forensics and cyber crime* (pp. 99–110). Springer.

Orebaugh, A., Kinser, J., & Allnutt, J. (2014). Visualizing instant messaging author writeprints for forensic analysis. In *Proceedings of the conference on digital forensics, security and law* (p. 191).

Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of information Science*, *28*(6), 441–453.

Ozgul, F., Atzenbeck, C., & Erdem, Z. (2011). How much similar are terrorists networks of istanbul? In *Advances in social networks analysis and mining (asonam), 2011 international conference on* (pp. 468–472).

Ozgul, F., & Erdem, Z. (2012, Aug). Detecting criminal networks using social similarity. In *Advances in social networks analysis and mining (asonam), 2012 ieee/acm international conference on* (p. 581-585). doi: 10.1109/ASONAM.2012.98

Ozgul, F., & Erdem, Z. (2013, Aug). Which crime features are important for criminal network members? In *Advances in social networks analysis and mining (asonam), 2013 ieee/acm international conference on* (p. 1058-1060).

Ozgul, F., Erdem, Z., Bowerman, C., & Atzenbeck, C. (2010, Aug). Comparison of feature-based criminal network detection models with k-core and n-clique. In *Advances in social networks analysis and mining (asonam), 2010* *international conference on* (p. 400-401). doi: 10.1109/ASONAM.2010.45

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The pagerank citation ranking: bringing order to the web.

Paradesi, S., Seneviratne, O., & Kagal, L. (2012). Policy aware social miner. In *Security and privacy workshops (spw), 2012 ieee symposium on* (pp. 53–59).

Pearce, M. (2013, September). *florida girl, 12, found dead after bullies said kill yourself".* Los Angeles Times, Los Angeles, CA, USA. Retrieved from `http://articles.latimes.com/2013/sep/12/nation/la-nann-florida-cyberbullying-20130912`

Peled, O., Fire, M., Rokach, L., & Elovici, Y. (2013). Entity matching in online social networks. In *Social computing (socialcom), 2013 international conference on* (pp. 339–344).

Penna, L., Clark, A., & Mohay, G. (2010). A framework for improved adolescent and child safety in mmos. In *Advances in social networks analysis and mining (asonam), 2010 international conference on* (pp. 33–40).

Perlroth, N. (2013, January). *chinese hackers infiltrate new york times computers".* New York Times. Retrieved from `https://www.nytimes.com/2013/01/31/technology/chinese-hackers-infiltrate-new-york-times-computers.html`

Piraveenan, M., Uddin, S., & Chung, K. (2012, Aug). Measuring topological robustness of networks under sustained targeted attacks. In *Advances in social networks analysis and mining (asonam), 2012 ieee/acm international conference on* (p. 38-45). doi: 10.1109/ASONAM.2012.17

Raghavan, S. (2013). Digital forensic research: current state of the art. *CSI Transactions on ICT*, *1*(1), 91–114.

Ragin, C. C., & Rihoux, B. (2009). *Configurational comparative methods: Qualitative comparative analysis (qca) and related techniques.* Sage.

RSA. (2011). *Anatomy of an attack.* RSA FraudAction Research Labs. Retrieved from `http://blogs.rsa.com/anatomy-of-an-attack/`

Sageman, M. (2004). *Understanding terror networks.* University of Pennsylvania Press.

Sarvari, H., Abozinadah, E., Mbaziira, A., & Mccoy, D. (2014). Constructing and analyzing criminal networks. In *Security and privacy workshops (spw), 2014 ieee* (pp. 84–91).

Sayaf, R., Rule, J., & Clarke, D. (2013). Can users control their data in social software? an ethical analysis of control systems. In *Security and privacy workshops (spw), 2013 ieee* (pp. 1–4).

Schechter, S., Egelman, S., & Reeder, R. W. (2009). It's not what you know, but who you know: A social approach to last-resort authentication. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1983–1992). New York, NY, USA: ACM. doi: 10.1145/1518701.1519003

Schwartz, M. (2013). *Microsoft hacked: joins apple, facebook, twitter.* Information Week. Retrieved from `http://www.informationweek.com/security/`

Attackacks/microsoft-hacked-joins-apple-facebook-tw/240149323

Schweinberger, M., Petrescu-Prahova, M., & Vu, D. Q. (2014). Disaster response on september 11, 2001 through the lens of statistical network analysis. *Social networks*, *37*, 42–55.

Seigfried-Spellar, K. (2013). Measuring the preference of image content for self-reported consumers of child pornography. In M. Rogers & K. Seigfried-Spellar (Eds.), *Digital forensics and cyber crime* (Vol. 114, p. 81-90). Springer Berlin Heidelberg. doi: 10.1007/978-3-642-39891-9_6

Seigfried-Spellar, K., Bertoline, G., & Rogers, M. (2012). In P. Gladyshev & M. Rogers (Eds.), *Digital forensics and cyber crime* (Vol. 88, p. 17-32). Springer Berlin Heidelberg.

Semenov, A., Nikolaev, A., & Veijalainen, J. (2013). Online activity traces around a boston bomber. In *Advances in social networks analysis and mining (asonam), 2013 ieee/acm international conference on* (pp. 1050–1053).

Shapiro, J. N. (2005). Organizing terror: hierarchy and networks in covert organizations. In *annual meeting of the american political science association, washington, dc.*

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3-4), 591–611.

Short, M. B., D'ORSOGNA, M. R., Pasour, V. B., Tita, G. E., Brantingham, P. J., Bertozzi, A. L., & Chayes, L. B. (2008). A statistical model of criminal behavior. *Mathematical Models and Methods in Applied Sciences*, *18*(supp01), 1249–1267.

Sikdar, S., Kang, B., O'Donovan, J., Hollerer, T., & Adah, S. (2013). Understanding information credibility on twitter. In *Social computing (socialcom), 2013 international conference on* (pp. 19–24).

Singh, I., Butkiewicz, M., Madhyastha, H. V., Krishnamurthy, S. V., & Addepalli, S. (2012). Enabling private conversations on twitter. In *Proceedings of the 28th annual computer security applications conference* (pp. 409–418).

Singh, K., Bhola, S., & Lee, W. (2009). xbook: Redesigning privacy control in social networking platforms. In *Proceedings of the 18th conference on usenix security symposium* (pp. 249–266). Berkeley, CA, USA: USENIX Association.

Singh, M., Bansal, D., & Sofat, S. (2016). Behavioral analysis and classification of spammers distributing pornographic content in social media. *Social Network Analysis and Mining*, *6*(1), 41. Retrieved from http://dx.doi.org/10.1007/s13278-016-0350-0 doi: 10.1007/s13278-016-0350-0

SocialEngineer. (n.d.). *what is phishing paypal phishing examples.".* Social-Engineer.org. Retrieved from http://www.social-engineer.org/wiki/archives/Phishing/Phishing-PayPal.html

Spezzano, F., Subrahmanian, V., & Mannes, A. (2013). Stone: Shaping terrorist organizational network efficiency. In *Proceedings of the 2013 ieee/acm international conference on* *advances in social networks analysis and mining* (pp. 348–355).

Squicciarini, A., & Griffin, C. (2014). Why and how to deceive: game results with sociological evidence. *Social Network Analysis and Mining*, *4*(1), 1–13.

StatsDirect. (2017). *Cox (proportional hazards) regression.* `http://www.statsdirect.com/help/survival_analysis/cox_regression.htm`. ([Accessed: 2017-02-16])

Steven, B., Ewan, K., & Loper, E. (2009). *Nltk: The natural language toolkit.* OReilly Media Inc.

Subrahmanian, V., Mannes, A., Sliva, A., Shakarian, J., & Dickerson, J. P. (2012). *Computational analysis of terrorist groups: Lashkar-e-taiba: Lashkar-e-taiba.* Springer Science & Business Media.

Sundsoy, P. R., Bjelland, J., Canright, G., Engo-Monsen, K., & Ling, R. (2012). The activation of core social networks in the wake of the 22 july oslo bombing. In *Advances in social networks analysis and mining (asonam), 2012 ieee/acm international conference on* (pp. 586–590).

Sweeney, L. (2002, October). K-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, *10*(5), 557–570. doi: 10.1142/S0218488502001648

Sweeney, L. (2000). *uniqueness of simple demographics in the u.s. population* (Tech. Rep.). Technical report, Carnegie Mellon University, Laboratory for International Data Privacy.

Tayebi, M., Ester, M., Glasser, U., & Brantingham, P. (2014, Aug). Crimetracer: Activity space based crime location prediction. In *Advances in social networks analysis and mining (asonam), 2014 ieee/acm international conference on* (p. 472-480). doi: 10.1109/ASONAM.2014.6921628

Tayebi, M., & Glasser, U. (2012, Aug). Investigating organized crime groups: A social network analysis perspective. In *Advances in social networks analysis and mining (asonam), 2012 ieee/acm international conference on* (p. 565-572). doi: 10.1109/ASONAM.2012.96

Tootoonchian, A., Saroiu, S., Ganjali, Y., & Wolman, A. (2009). Lockr: Better privacy for social networks. In *Proceedings of the 5th international conference on emerging networking experiments and technologies* (pp. 169–180). New York, NY, USA: ACM. doi: 10.1145/1658939.1658959

Tucker, D. (2008). Terrorism, networks, and strategy: Why the conventional wisdom is wrong. *Homeland Security Affairs*, *4*(2).

Tuna, T., Akbas, E., Aksoy, A., Canbaz, M. A., Karabiyik, U., Gonen, B., & Aygun, R. (2016). User characterization for online social networks. *Social Network Analysis and Mining*, *6*(1), 104. Retrieved from `http://dx.doi.org/10.1007/s13278-016-0412-3` doi: 10.1007/s13278-016-0412-3

Turnbull, B., & Randhawa, S. (2015). Automated event and social network extraction from digital evidence sources with ontological mapping. *Digital Investigation*, *13*, 94–106.

Turner, R. E., Edgley, C., & Olmstead, G. (1975). Information control in conversations: Honesty is not always the best policy. *Kansas Journal of Sociology*, 69–89.

Tuttle, M. (2016). *Terrorism recruitment using internet marketing* (Unpublished doctoral dissertation). UTICA COLLEGE.

Twitter. (2017). The Streaming APIs. (`https://dev.twitter.com/streaming/overview` [Accessed: 2017-01-31])

UCINET. (2017). *Caviar.* Retrieved from `https://sites.google.com/site/ucinetsoftware/datasets/covert-networks/caviar` ([Accessed: 2017-02-07])

UNICEF Office of Research, I. (2011, December). *unicef child safety online: Global challenges and strategies.”.* https://www.unicef-irc.org/publications/pdf/ict_eng.pdf.

Utz, S. (2005). Types of deception and underlying motivation: What people think. *Social Science Computer Review*, *23*(1), 49–56.

Van Der Galien, M. (2017a, January). *3 men gang-rape young woman in sweden, broadcast it live on facebook.* PJ Media. Retrieved from `https://pjmedia.com/lifestyle/2017/01/26/3-men-gang-rape-young-woman-in-sweden-broadcast-it-live-on-facebook/`

Van Der Galien, M. (2017b, January). *'friends' live-streamed sexual assault of teen on snapchat while she slept.* PJ Media. Retrieved from `https://pjmedia.com/lifestyle/2017/01/31/friends-live-streamed-sexual-assault-of-teen-on-snapchat-while-she-slept/`

Vermande, M. M., Van den Oord, E. J., Goudena, P. P., & Rispens, J. (2000). Structural characteristics of aggressor–victim relationships in dutch school classes of 4-to 5-year-olds. *Aggressive Behavior*, *26*(1), 11–31.

Vigliotti, M. G., & Hankin, C. (2015). Discovery of anomalous behaviour in temporal networks. *Social Networks*, *41*, 18–25.

Viswanath, B., Bashir, M. A., Crovella, M., Guha, S., Gummadi, K. P., Krishnamurthy, B., & Mislove, A. (2014). Towards detecting anomalous user behavior in online social networks. In *23rd usenix security symposium (usenix security 14)* (pp. 223–238). San Diego, CA: USENIX Association.

Vosecky, J., Hong, D., & Shen, V. Y. (2009, July). User identification across multiple social networks. In *2009 first international conference on networked digital technologies* (p. 360-365). doi: 10.1109/NDT.2009.5272173

Vosecky, J., Hong, D., & Shen, V. Y. (2010). User identification across social networks using the web profile and friend network. *IJWA*, *2*(1), 23–34. Retrieved from `http://dline.info/ijwa/fulltext/v2n103.pdf`

Wagner, C., Asur, S., & Hailpern, J. (2013). Religious politicians and creative photographers: Automatic user categorization in twitter. In *Proceedings of the 2013 international conference on social computing* (pp. 303–310). Washington, DC, USA: IEEE Computer Society. Retrieved from `http://dx.doi.org/10.1109/SocialCom.2013.49` doi: 10.1109/SocialCom.2013.49

Wagner, C., Mitter, S., Körner, C., & Strohmaier, M. (2012). When social bots attack: Modeling susceptibility of users in online social networks. *Making Sense of Microposts (# MSM2012)*, *2*(4).

Wakabayashi, D., & Isaac, M. (2017, January). *in race against fake news, google and facebook stroll to the starting line".* The New York Times. Retrieved from `https://www.nytimes.com/2017/01/25/technology/google-facebook-fake-news.html`

Walther, J. B. (1996). Computer-mediated communication. *Communication Research*, *23*(1), 3-43. Retrieved from `http://dx.doi.org/10.1177/009365096023001001` doi: 10.1177/009365096023001001

Wang, A. H. (2010). Detecting spam bots in online social networking sites: A machine learning approach. In S. Foresti & S. Jajodia (Eds.), *Data and applications security and privacy xxiv: 24th annual ifip wg 11.3 working conference, rome, italy, june 21-23, 2010. proceedings* (pp. 335–342). Springer Berlin Heidelberg.

Wang, Y., & Nepali, R. K. (2013). Privacy measurement for social network actor model. In *Social computing (socialcom), 2013 international conference on* (pp. 659–664).

Wash, R., & Rader, E. (2011). Influencing mental models of security: a research agenda. In *Proceedings of the 2011 workshop on new security paradigms workshop* (pp. 57–66).

Wei, W., Xu, F., Tan, C. C., & Li, Q. (2012, March). Sybildefender: Defend against sybil attacks in large social networks. In *2012 proceedings ieee infocom* (p. 1951-1959). doi: 10.1109/INFCOM.2012.6195572

Weimann, G. (2004). *www. terror. net: How modern terrorism uses the internet* (Vol. 31). DIANE Publishing.

Weng, L., Ratkiewicz, J., Perra, N., Gonçalves, B., Castillo, C., Bonchi, F., ... Flammini, A. (2013). The role of information diffusion in the evolution of social networks. In *Proceedings of the 19th acm sigkdd international conference on knowledge discovery and data mining* (pp. 356–364).

White, J., Park, J. S., Kamhoua, C. A., & Kwiat, K. A. (2013). Game theoretic attack analysis in online social network (osn) services. In *Proceedings of the 2013 ieee/acm international conference on advances in social networks analysis and mining* (pp. 1012–1019). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/2492517.2500257` doi: 10.1145/2492517.2500257

Wiil, U. K., Gniadek, J., & Memon, N. (2010). Measuring link importance in terrorist networks. In *Advances in social networks analysis and mining (asonam), 2010 international conference on* (pp. 225–232).

Wolak, J., Finkelhor, S. D., Mitchell, K. J., & Ybarra, M. L. (2010). Online predators and their victims: Myths, realities, and implications for prevention and treatment. *Psychology of violence*, *1*(S), 13–35.

Wondracek, G., Holz, T., Kirda, E., & Kruegel, C. (2010, May). A practical attack to de-anonymize social network users. In *2010 ieee symposium on* *security and privacy* (p. 223-238). doi: 10.1109/SP.2010.21

Xu, J., & Chen, H. (2005). Criminal network analysis and visualization. *Communications of the ACM*, *48*(6), 100–107.

Xu, W., Zhang, F., & Zhu, S. (2010). Toward worm detection in online social networks. In *Proceedings of the 26th annual computer security applications conference* (pp. 11–20). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/1920261.1920264` doi: 10.1145/1920261.1920264

Yamada, A., Kim, T. H.-J., & Perrig, A. (2012, February). Exploiting privacy policy conflicts in online social networks.

Yan, G., Chen, G., Eidenbenz, S., & Li, N. (2011). Malware propagation in online social networks: Nature, dynamics, and defense implications. In *Proceedings of the 6th acm symposium on information, computer and communications security* (pp. 196–206). New York, NY, USA: ACM. doi: 10.1145/1966913.1966939

Yang, C. C., & Ng, T. D. (2007). Terrorism and crime related weblog social network: Link, content analysis and information visualization. In *Intelligence and security informatics, 2007 ieee* (pp. 55–58).

Yang, Z., Wilson, C., Wang, X., Gao, T., Zhao, B. Y., & Dai, Y. (2014, February). Uncovering social network sybils in the wild. *ACM Trans. Knowl. Discov. Data*, *8*(1), 2:1–2:29. doi: 10.1145/2556609

Yu, H., Gibbons, P. B., Kaminsky, M., & Xiao, F. (2008, May). Sybillimit: A near-optimal social network defense against sybil attacks. In *2008 ieee symposium on security and privacy (sp 2008)* (p. 3-17). doi: 10.1109/SP.2008.13

Yu, H., Kaminsky, M., Gibbons, P. B., & Flaxman, A. (2006, August). Sybilguard: Defending against sybil attacks via social networks. *SIGCOMM Comput. Commun. Rev.*, *36*(4), 267–278. doi: 10.1145/1151659.1159945

Zafarani, R., Abbasi, M. A., & Liu, H. (2014). *Social media mining: an introduction.* Cambridge University Press.

Zangerle, E., & Specht, G. (2014). "sorry, i was hacked": A classification of compromised twitter accounts. In *Proceedings of the 29th annual acm symposium on applied computing* (pp. 587–593). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/2554850.2554894` doi: 10.1145/2554850.2554894

Zetter, K. (2010). *Google hack attack was ultra sophisticated.* Wired. Retrieved from `http://www.wired.com/threatlevel/2010/01/operation-aurora/`

Zhang, C., Jiang, A. X., Short, M. B., Brantingham, P. J., & Tambe, M. (2013). Modeling crime diffusion and crime suppression on transportation networks: An initial report. In *Aaai fall symposium* (Vol. 2013).

Zhang, C., Sun, J., Zhu, X., & Fang, Y. (2010, July). Privacy and security for online social networks: challenges and opportunities. *IEEE Network*, *24*(4), 13-18. doi: 10.1109/MNET.2010.5510913

Zubiaga, A., & Ji, H. (2014). Tweet, but verify: epistemic study of information verification on twitter. *Social Network Analysis and Mining*, *4*(1), 1–12.