Social Cyber Forensics: Leveraging Open Source Information and Social Network Analysis to Advance Cyber Security Informatics

Samer Al-khateeb[1] and Nitin Agarwal[2]

[1] Creighton University
2500 California Plaza, Omaha, NE 68178
sameral-khateeb1@creighton.edu
[2] University of Arkansas at Little Rock
2801 S University Ave, Little Rock, AR 72204
nxagarwal@ualr.edu

Abstract. In this paper, we introduce the concept of Social Cyber Forensics (SCF) and its usability. Then, we introduce a tool, i.e., Maltego that can be used to study the cross-media affiliation and uncover hidden relations among various online groups. We also provide three stepwise methodologies that leverage Maltego and various Open Source Information (OSINF) to uncover the hidden relationship among (1) Twitter accounts and a set of websites/blogs; (2) websites/blogs and other websites/blogs; or (3) infer the ownership of a set of websites/blogs. These methodologies have been tested during many cyber propaganda campaigns that were projected against NATO forces. A high-level view of these case studies that leveraged the concepts and methodologies provided in this paper are briefly highlighted here while the details of each case were published in various venues which are pointed out for interested readers.

Keywords: Social Cyber Forensics · Open Source Information · Propaganda campaigns · Twitter · Online Deviant Groups.

1 Introduction: Social Cyber Forensics Analysis (SCF)

Most scientific work to date has focused on the acquisition of social data from digital devices as well as applications installed on them. Over time, Online Social Networks (OSNs) have become the largest and fastest growing entities on the Internet, containing hundreds of millions of people, and now social bots (computer programs designed to carry out specific tasks). OSNs hosted on platforms like Facebook, LinkedIn, and Twitter, contain a plethora of data about its members, which is of interest to both digital forensic scientists and practitioners [1]. The forensic potential of these OSNs has been acknowledged by researchers, and there have been a number of studies on extracting this forensically relevant data from them. As OSNs continuously replace traditional means of digital storage, sharing, and communication [1], collecting this ever-growing volume of data is

becoming a challenge. Within the past decade, data collected from OSNs has already played a major role as evidence in criminal cases, either as incriminating evidence or to confirm alibis [2–4].

Online Deviant Groups (ODGs), e.g., terrorist groups and criminal organizations continue to utilize OSNs to promote, enhance, and facilitate their respective goals. It might be more efficient to take an intelligence-driven approach for identifying evidentiary trials. Harvesting forensically relevant data directly from the metadata associated with the OSN user accounts themselves, as this paper aims to discuss, would be more efficient than traditional forensic techniques of analyzing the hardware, network traffic, file systems, and other traditional scenarios in digital forensics.

The deviant practices conducted by ODGs over modern Information and Communication Technologies (ICTs), especially the social media, call for an in-depth study to better understand how these ODGs use social media to advance their strategic communication and how these ODGS evolve over time. Using Social Cyber Forensics (SCF) will help in:

- identifying the hidden relationships among various ODGs using various publicly available data and metadata associated with these groups social media footprints.
- studying and identifying ODGs cross-media affiliation, which might help in discovering new deviant groups.
- enhancing the collected raw data from Twitter, Facebook, blogs, etc to have an idea about various events or campaigns size, strategy, and/or goal.

So SCF can help authorities in developing countermeasures to mitigate the risks that might be caused by these ODGs. Hence, there is a need to conduct more research in this arena, develop methodologies to collect this ever-growing source of information, and create tools that can access/extract and analyze these data to derive insights.

For the last three and half decades, digital forensics tools have been evolved from simple tools, which were used mainly by law enforcement agencies to important tools for detecting and solving corporate fraud [5]. Cyber forensics tools are not a new type of tools but they are evolving over time to have more capabilities, more exposure to the audience (i.e., investigators or public users), and more types of data that can be extracted using each tool. Cyber forensics tools can be traced back to the early $1980's$ when these tools were mainly used by government agencies, e.g., the Royal Canadian Mounted Police (RCMP) and the U.S Internal Revenue Service (IRS) and were written in Assembly language or C language with limited capabilities and less popularity. As time passes these tools got more sophisticated and in the mid of $1980s$ these tools were able to recognize file types as well as retrieve lost or deleted files, e.g., XtreeGold and DiskEdit by Norton. In the $1990s$ these tools become more popular and also have more capabilities, e.g., they can recover deleted files and fragments of deleted files such as Expert Witness and Encase [6]. Nowadays, many tools are available for public use and have data collection and visualization capabilities that make the process of analyzing the collected data easier, e.g., Maltego tool.

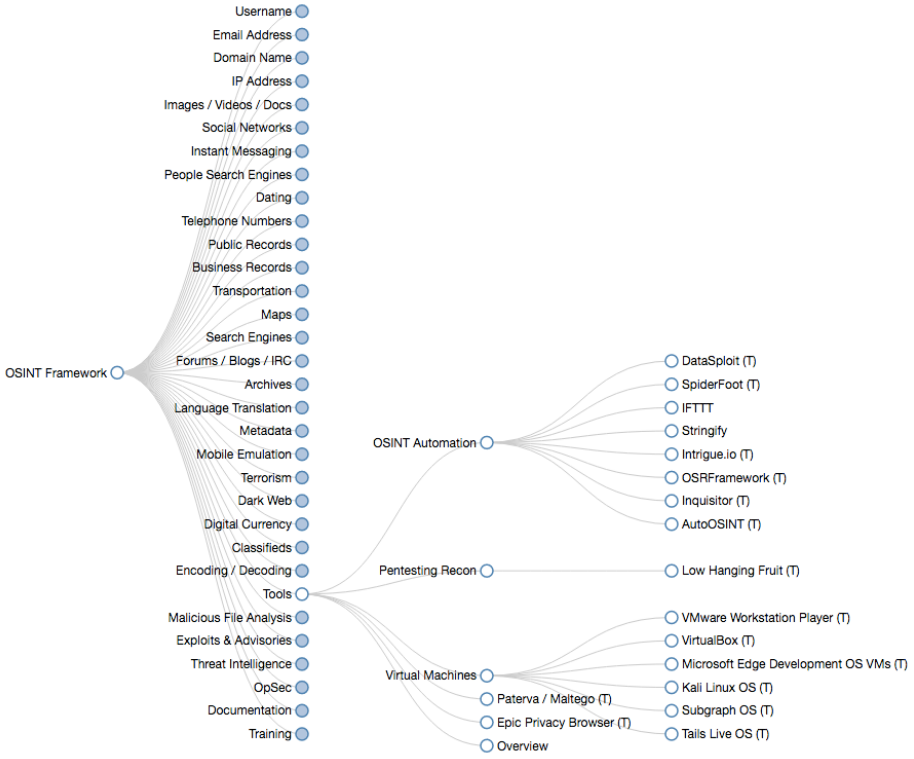

Fig. 1. The available OSINT tools. Not many tools available.

Before introducing the concept of *Social Cyber Forensics (SCF)* lets explain what *forensics* mean. The word *forensics* refers to the discipline of collecting, analyzing, and reporting of evidence to build a case (in law) or to assert a relationship among two or more entities. *Cyber Forensics* is the discipline of using digital tools to find digital evidence to support an assertion about a relationship [7]. Cyber forensic evidence can be used to detect or prevent a crime or address a dispute from information drawn from a digital investigation. We define *Social Cyber Forensics (SCF)* as a branch of Cyber Forensics and its the process of investigating the relationships among "entities" and revealing the digital connections among them in social media space by extracting/collecting metadata associated with their social media accounts, e.g., affiliations of the user, geolocation, IP address. An *"entity"* is an information actor which can be a single individual, a group, an organization, a nation-state, etc. An entity, for example, can be a single "individual" that own a single blog or multiple blogs.

To be able to use social cyber forensics (SCF) to uncover hidden relationships among different entities an analyst needs a "seed". A *seed* is an initial knowledge that can be used to investigate an entity or set of entities, e.g., a seed can be a

Web Tracker Code (WTC), a blog, a Twitter account, IP address, or any other information that the SCF tools can use to reveal hidden or unknown information about the seed entity. Web Tracker Code (WTC) is an online analytics tool that allows a website owner to gather some statistics about their website visitors such as their browser, operating system, and the country they are browsing the web from, along with other metadata. These web trackers have an ID number that is usually embedded in the website HTML code [8–10]. For example, Google provides users with a capability to track their website activities using Google Analytics service[3].

Social cyber forensic analysis depends completely on publicly available sources of data, i.e., it uses *Open Source Information (OSINF)* to get the metadata needed about an entity. There are many companies that provide services to accomplish this task, e.g., PeekYou[4], Metagoofil[5], and Shodan[6], but there are very few useful tools that can obtain data from multiple sources (e.g., Shodan data is accessible through Maltego) and report them in an easy to understand format like *Maltego*. Figure 1 shows tools that are currently available and can do SCF. An interactive version of this figure can be accessed at `http://osintframework.com/`, click on OSINT Framework → Tools, to explore the tools available.

Maltego is a tool that is developed by Paterva Ltd[7] and can be used to gather various publicly available data (for example, it can be used to provide an insight on how different social media platforms such as blogs are connected or affiliated to Twitter accounts [8]). Basically, Maltego is an open source intelligence and forensics application that can determine the relationships and real world links among people, groups of people (social networks), companies, organizations, websites, and Internet infrastructure, e.g., Domains, DNS names, Netblocks, and IP addresses. Maltego saves an analyst a lot of time in mining and gathering information as well as representing this extracted information in an easy to understand format. For the reasons mentioned below, we chose Maltego to demonstrate the benefits of SCF and combining multiple OSINF to unveil the relationship among various entities. Maltego enables the user to:

- obtain data from multiple sources.
- report the results in a graph format that is suitable for link analysis.
- export the collected results in a graph format which can be analyzed using other social network analysis software.
- collect a rich set of publicly available data for free (i.e., using Maltego CE).
- use and train other users easily by following a set of steps to obtain the desired results.

Below we are providing a set of definitions of the terminologies that will be used throughout this paper, especially in the methodologies section.

[3]Google Analytics: `https://analytics.google.com/`
[4]PeekYou: `http://www.peekyou.com/`
[5]Metagoofil: `http://www.edge-security.com/metagoofil.php`
[6]Shodan: `https://www.shodan.io/`
[7]Paterva Ltd: `https://www.paterva.com/`

- *Transforms:* tiny pieces of code that take one type of information and change it into another for interoperability.
- *Pipeline:* a set of transforms and filters that are executed in sequence, i.e., like a macro in Microsoft Excel.
- *Trigger:* a graph condition and a transform, i.e., when something happens on the graph, then run this transform.
- *Feeder:* a mechanism to feed entities into Maltego.
- *Machine:* a combination of pipelines, triggers, and feeders.

Maltego has two main types of licenses, i.e., *server* and *client* licenses. Using Maltego *server license* users can deploy their own transforms and run those transforms on their own data, then share these transforms with others (e.g., across their own enterprise). The server license has three types namely: Commercial Transform Application Server (CTAS), internal Transform Distribution Server (iTDS), and Communications Server (Comms Server), more information about Maltego servers is available at: https://www.paterva.com/web7/buy/maltego-servers.php.

Using Maltego *client license* users can interact with a GUI to collect and analyze various publicly available data. It has four types, namely: Maltego XL, Maltego Classic, Maltego CE (free version), and CaseFile. The main difference among all the client versions is *the maximum number of entities in the graph* and *the number of entities returned from a single transform*, more information about Maltego clients is available at: https://www.paterva.com/web7/buy/maltego-clients.php. All Maltego client use Java, so it runs on any modern computer with Windows, Mac, or Linux operating systems, however below are the recommended system requirements provided by Paterva Ltd to run Maltego:

- Windows 7 or above, Mac OS X or above, or the latest version of the Linux operating system.
- Java 8.0 (or the latest version).
- At least 2GB of Random-access memory (RAM), but the more the better.
- Any modern multi-core processor is adequate.
- 4GB of disk space is adequate.
- Mouse to make navigating the graphs much easier.
- Internet access is a must to fully operate. The outgoing connections will be on the following ports: 80, 443, 8081, and port 5222.

Next, in section 2 we provide a set of stepwise methodologies that can be followed using Maltego to reach the desired outcome of the analysis.

2 Methodologies to Extract Open Source Information

In this section, we provide three sets of stepwise methodologies that can be followed to answer some interesting questions like the following ones:

1. Given a set of websites or blogs (i.e., seed), can we *discover* other unknown websites that are working together with the "seed" to disseminate propaganda or misinformation?

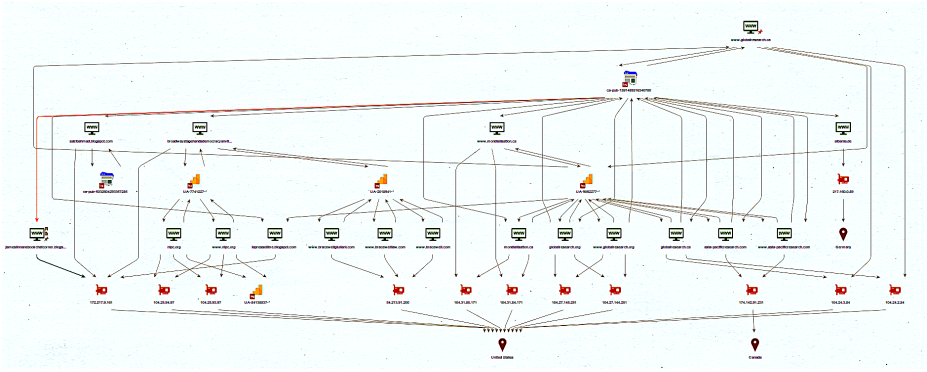

Fig. 2. Maltego results after following the steps in section 2.1.

2. Given a set of Twitter accounts or other social media accounts, can we find the footprint of the seed accounts on other social media channels, i.e., their cross-media affiliation?
   - Can we identify other accounts working with the seed to disseminate propaganda?
3. Given a set of websites or blogs (seed), can we *infer the ownership* of these websites, i.e., who they belong to individuals or organizations?
   - Are these set of websites working together to disseminate propaganda?
   - What are their strategies?

These methodologies are tested on many real-world events such as the NATO Trident Juncture 2015 exercise (TRJE2015)[8], Exercise Anakonda 2016 (AN16)[9], Brilliant Jump 2016 exercise (BRJP16)[10] among others. In all the aforementioned events these methodologies proved to be effective in uncovering hidden connections between different ODGs or show their cross-media affiliation which gave great insights to many Public Affairs Officer (PAO). The methodologies provided next are tested and used with the free community version of Maltego, i.e., *Maltego CE v4.0.11*.

2.1 Finding Related Websites From Web Tracker Code (WTC)

Here, we introduce the *first* stepwise methodology that can be followed to answer the first question mentioned above. This methodology has the following steps:

1. Insert the seed blogs (all must start with "`www.`") into Maltego by choosing "$Infrastructure \rightarrow Website$" then drag and drop the entity to the main window.

[8]Trident Juncture 2015 exercise: https://jfcbs.nato.int/trident-juncture
[9]Exercise Anakonda 2016: http://www.eur.army.mil/anakonda/
[10]Brilliant Jump 2016 exercise: http://www.jfcbs.nato.int/page5735825/brilliant-jump-2016

2. Select all the inserted blogs and right-click, then use "*To Tracking Codes*" transformation to get the Unique Identifiers for these seed blogs, e.g., Google Analytics IDs. Note: if the website points to a WTC it means that the website HTML code contains the WTC.
3. Select all the tracker codes and right-click, then use "*To Other Sites With Same Code*" transformation to get the blogs that use the same code. Note: if the WTC point to the website it does not necessarily mean that the WTC is embedded in the website HTML, but it does mean that the WTC owner added that website to their list of to be tracked websites.
4. Repeat step 2 and 3 until no new blogs nor tracking codes identified.
5. Select all the blogs and chose "*To IP Address[DNS]*" transformation to get the IP addresses of all the sites.
6. Select all the identified IP addresses and chose "*ToLocationCountry*" transformation to get the location of where these websites are geolocated.

Figure 2 shows the results we obtained from running the steps mentioned in section 2.1 using the website `www.globalresearch.ca`.

2.2 Finding Blogs From Twitter Handles

Here, we introduce the *second* stepwise methodology that can be followed to answer the second question mentioned above. This methodology has the following steps:

1. Copy and paste the Twitter username (without the @ sign) to the main window.
2. Right-click on the Twitter username, then change the type to "*People* $\rightarrow$ *UnknownSuspect*".
3. Right click on the Unknown Suspect entity and use "To Twitter Affiliation [Search Twitter]" transformation to find all possible accounts this person have.
4. Select the Twitter username and use "*To Person [Convert]*" transformation.
5. Select the person entity and use "*To Email Address [Using Search Engine]*" transformation to find possible email addresses for that person using a search engine.
6. Select the email entity and use "*To Website [using Search Engine]*" transformation to find websites associated with the email address found.
7. Select the email entity and use "*To URLs [Show search engine results]*" transformation to find any URLs associated with the email address found.
8. If step 6 and 7 did not return good or enough results then: Select the person entity and use "*To Website [using Search Engine]*".
9. Note that the **methodology in subsection 2.1** can be applied here to all the websites discovered in **steps 6 and 7** to discover more blogs if needed.

Figure 3 shows the results we obtained from running the steps mentioned in section 2.2 using the Twitter account `@recortescero`.

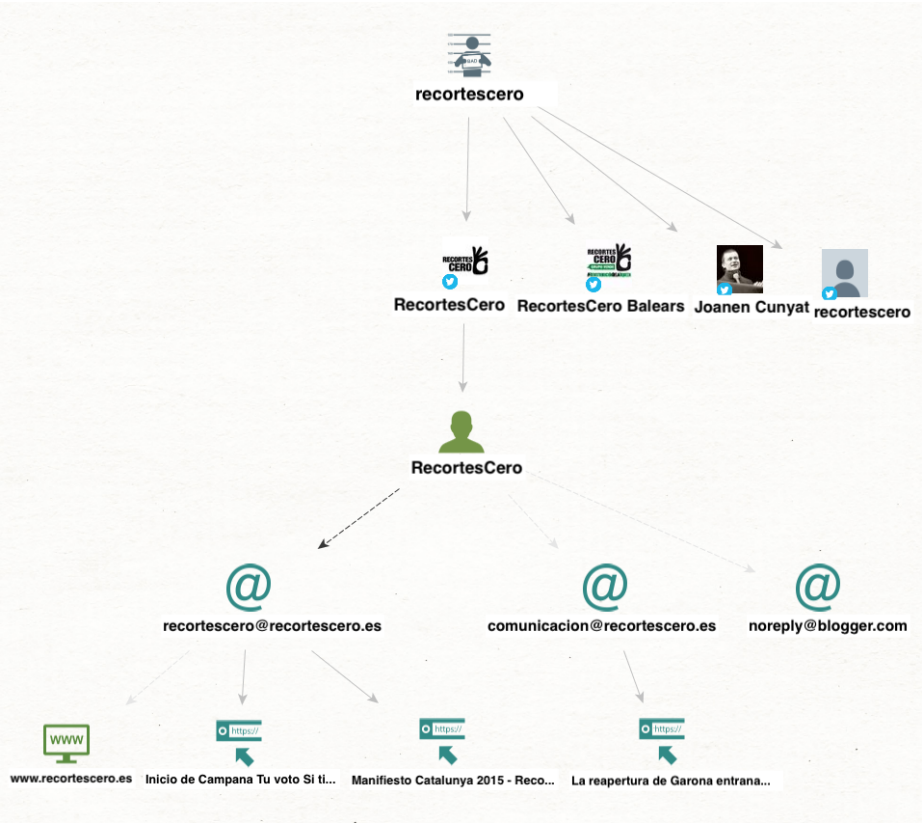

Fig. 3. Maltego results after following the steps in section 2.2.

2.3 Inferring the Ownership or Hidden Connection of Different Websites

Here, we introduce the *third* stepwise methodology that can be followed to answer the third question mentioned above. This methodology has the following steps:

1. Insert the seed websites (all must start with "`www.`") into Maltego by choosing "$Infrastructure \rightarrow Website$".
2. Select all the inserted blogs and right-click, then use "*To Tracking Codes*" transformation to get the Unique Identifiers for these seed blogs, e.g., Google Analytics IDs.
3. Select all websites and use "*To IP Address [DNS]*" transformation to get the IP addresses of all the sites.
4. Select all IP Address and use "*To Entities from Whois [Alchemy]*" transformation to get the owner name, location, phone number, email address, etc.

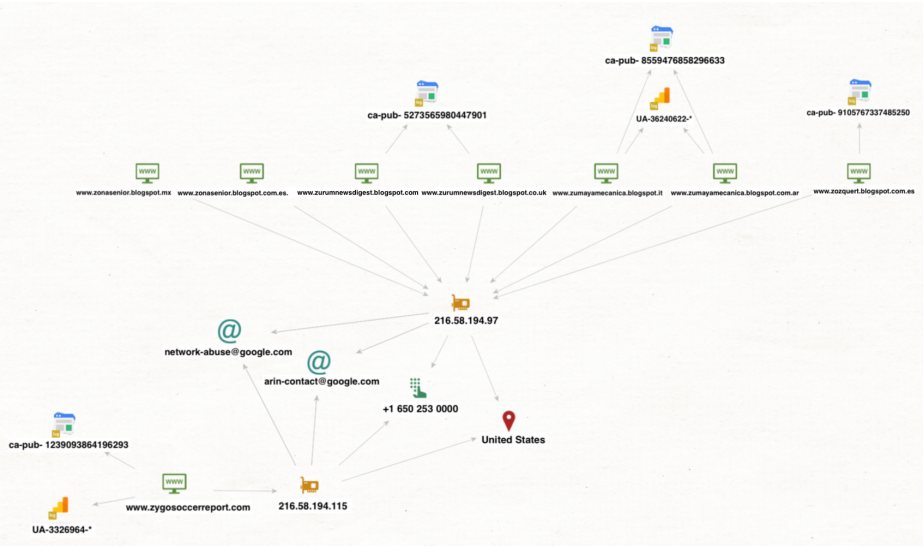

Fig. 4. Maltego results after following the steps in section 2.3.

5. Select all IP Address again and use "*to Location [country]*" transformation.
6. Select all IP Address and use "*To Telephone Number [from whois Info]*" transformation.
7. Select all IP Address and use "*To Email Address [from whois Info]*" transformation.
8. Note that the **methodology in subsection 2.1** can be applied here to the inserted blogs **step 1** to discover more blogs if needed.

Figure 4 shows the results we obtained from running the steps mentioned in section 2.3 using the set of websites:

- `www.zygosoccerreport.com`
- `www.zurumnewsdigest.blogspot.com`
- `www.zurumnewsdigest.blogspot.co.uk`
- `www.zumayamecanica.blogspot.it`
- `www.zumayamecanica.blogspot.com.ar`
- `www.zozquert.blogspot.com.es`
- `www.zonasenior.blogspot.mx`
- `www.zonasenior.blogspot.com.es`

3 Case Studies Leveraging OSINF and SCF

In this section, we introduce and highlight the main points of two case studies that leveraged the methodologies introduced in section 2. These studies were carried on by *deviant groups* who had an aim to disseminate cyber propaganda, misinformation (i.e., misleading), and disinformation (i.e., lies) during two NATO's

military exercises in Europe. The deviant events were organized using multiple social media channel, e.g., Twitter accounts tweeting propaganda about the event, or a YouTube video containing propaganda (e.g., tweets contain an image or a URL to blog post with offensive or biased memes, see figure 5), or a YouTube video that contains specific narratives such as conspiracy theory videos. The details of each case study are published and a reference to each case is provided for readers interested in more details.

3.1 Anti-NATO Propaganda During the 2015 Trident Juncture Exercise

On November 4, 2015, the US soldiers along with soldiers from more than thirty partner nations and allies moved 36,000 personnel across Europe during the Trident Juncture 2015 exercise. The exercise took place in the Netherlands, Belgium, Norway, Germany, Spain, Portugal, Italy, the Mediterranean Sea, the Atlantic Ocean, and also in Canada to prove the capability and readiness of the alliance on land, air, and sea. Also, to show that the alliance is equipped with the appropriate capabilities and capacities to face any present or future security issues. In addition to the Partner Nations and Allies, more than twelve aid agencies, International Organizations, and non-governmental organizations participated in the exercise to demonstrate "NATOs commitment and contribution to a comprehensive approach" [11].

Many opponent groups launched deviant cyber campaigns on Twitter, Blogs, Facebook, and other social media platforms that encouraged citizens to protest against the exercise or do violent/deviant acts. We identified six deviant groups by searching for their names on various social media platforms to identify their Twitter and blogging profiles. NATOs public affairs officers then verified these profiles (i.e., the domain experts). These six groups propagate their messages on social media inviting people to act against NATO and TRJE 2015 exercise.

We identified an initial set of twelve blogs along with the Twitter accounts (9 Twitter accounts) that are used to steer the audience from Twitter to blogs. We used Twitter rest API and Network Overview, Discovery, and Exploration for Excel (NodeXL) to collect a network of replies, mentions, tweets, friends, and followers for all the nine Twitter accounts and whoever is connected to them with any one of the aforementioned relationships. NodeXL is an add-in for Microsoft Excel [12] developed by the Social Media Research Foundation (SMR Foundation http://www.smrfoundation.org/research/our-network/). It can be used to collect data from various social media channels; visualize the data; and to perform social network analysis and content analysis. The Twitter handles, blogs, and names of the groups studied in this research are publicly available. However, in order to ensure their privacy, we do not disclose them here.

We did metadata extraction, e.g., Web Tracker Code (WTC) using Maltego. There are many types of WTC, for example, Google Analytics ID which is an online analytics service provided by Google that allows a website owner to gather statistics about their website visitors such as their browser, operating system, and country among other metadata. Multiple sites can be managed under a single

Fig. 5. Examples of offensive and biased memes observed on blogosphere during NATOs exercises to delegitimize exercises objectives.

Google analytics account. The account has a unique identifying "UA" number, which is usually embedded in the website's HTML code [8]. Using this identifier other blogs that are managed under the same UA number can be identified. This method was reported in [8,9].

So, using Maltego we inferred the connections among blogs and identified new sites that were previously undiscovered. We used Maltego in a snowball manner to discover other blogs. See Figure 7 that depicts these connections among various blogs. We were able to identify additional 9 blogs that are connected to the initial seed blogs by the same Google analytics IDs. These newly identified websites have the same content published on different portals and sometimes in different languages. For example, a website written in English may also have another identical version but written in another language (see Figure 6) that is native to the region. Such blogs are also known as "bridge blogs" [13]. We collected the IP addresses, website owner name, email address, phone numbers, and locations of all the websites to infer the hidden connections among these websites. Based on the websites geolocation (estimated by their IP address geolocation) we obtained three clusters of websites (see Figure 8). These clusters are helpful to know the originality of the blogs, which would help an analyst understand the propaganda that is being pushed by the specific blog. From an initial set of 12 blogs, we grew to 21 blogs, 6 locations, and 15 IP addresses. All the blogs we identified during this study were crawled and their data is stored in a database that the Blogtrackers tool can access and analyze.

We also applied social network analysis (SNA) to find the most important nodes in the network by activity type. Using NodeXL we were able to find the most used hashtags during the time of the exercise. This helps in targeting the same audience if counter-narratives were necessary to be pushed to the same audience. Additionally, we found the most tweeted URLs in the graph. This gives an idea about the public opinion concerns. Finally, we found the most used domains, which helps to know where the focus of analysis should be directed, or what other media platforms are used. Additionally, we applied Focal Structures Analysis (FSA) - on both the social network and the communication network - to identify powerful groups of individuals that are affecting the cyber propaganda

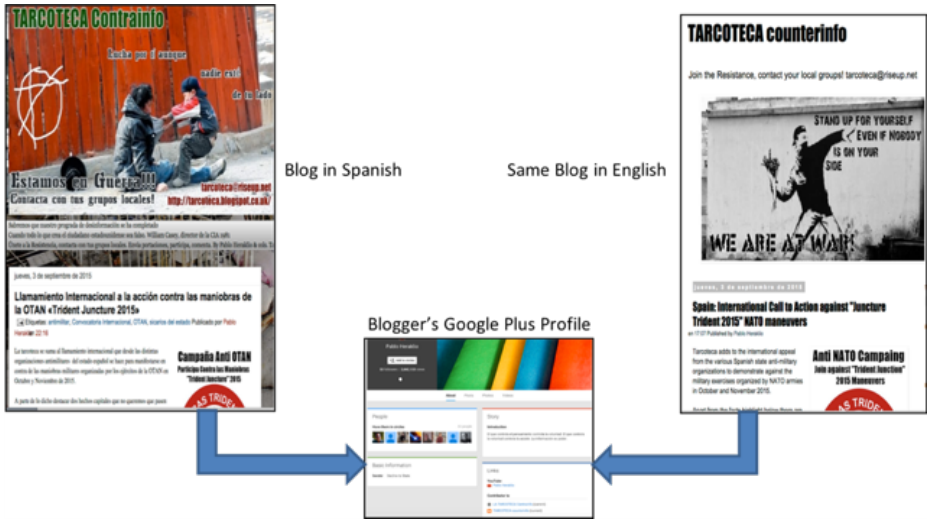

Fig. 6. A bridge blogger, Spanish and English blogs with anti-NATO narratives. The two blogs are kept unlinked. Connection between the blogs is discovered through Google Plus profile of the blogger using SCF methodologies.

campaign. Focal Structures Analysis [14](FSA) is an advanced social network analysis methodology to discover an influential set of nodes or vertices in a network or graph. The nodes do not have to be strongly connected and may not be the most influential on their own but by acting together they form a compelling power. Finally, we analyzed the blog's data that we crawled. Using SCF analysis and SNA we were able to identify a total of 21 blogs of interest. We trained web crawlers to collect data from these blogs, store the data in Blogtrackers database.

Blogtrackers is a web application that resulted from a plethora of academic research in a variety of fields such as information science, computer science, social science, and psychology. It is developed by the Collaboratorium for Social Media and Online Behavioral Studies (COSMOS) at the University of Arkansas at Little Rock. It provides a wide variety of analysis/features to analyze blogs data. Blogtrackers is a free tool that can be used by students, researchers, or any other civil or government entity. Any interested user can register and use the tool through the following URL `http://blogtrackers.host.ualr.edu`.

Using Blogtrackers a user can analyze the blogosphere at the *blogger level* and *blog level*. A user can search the blogosphere based on *keywords*, then track the blog posts/blogs and bloggers who mention these keywords by setting up a "Tracker". A user can check how sentiments change over time for a specific keyword or topic, check the blogger posting activity (e.g., daily, weekly, monthly, or yearly posting frequency), explore the entities (e.g., organizations, persons, locations) related to the searched keyword(s), analyze a network of blogs or bloggers - a very unique feature as blogs do not have a natural network - based

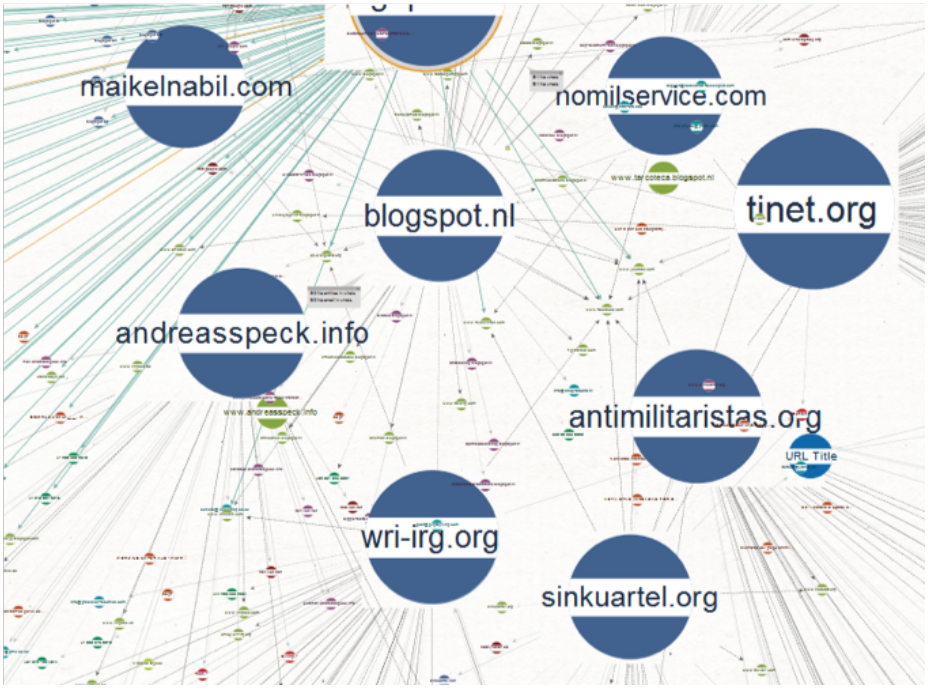

Fig. 7. Newly discovered blogs are connected via various relations identified using SCF.

on co-occurrence, metadata (such as Web Tracker Codes (WTC)), or any shared media (e.g., shared a YouTube channel, Facebook page). A user can also submit a blog of interest to be crawled by Blogtrackers crawler then once the crawling done a user gets a notification that the data of the blog are ready to be analyzed using the variety of analysis capabilities that Blogtrackers has. A user can also extract/export/download the data of interest from Blogtrackers in JSON format.

Finally, we performed the following analysis using Blogtrackers: posting frequency, influential blogs and bloggers analysis, keyword trends, sentiment analysis, etc. In this case study, the deviant groups also used deviant social bots to further disseminate their agenda to a large audience in a short period of time. For interested readers, the details of this study were published in [10] and an extended version of the study along with findings of the bots used during TRJE 2015 is published in [15].

3.2 Anti-NATO Propaganda During the 2015 Dragoon Ride Exercise

On March 21, 2015, US soldiers assigned to the 3^{rd} Squadron, 2^{nd} Cavalry Regiment in Estonia, Latvia, Lithuania, and Poland as part of Operation Atlantic Resolve began Operation Dragoon Ride. The US troops, nicknamed "Dragoons", were sent on a transfer mission crossing five international borders and covering

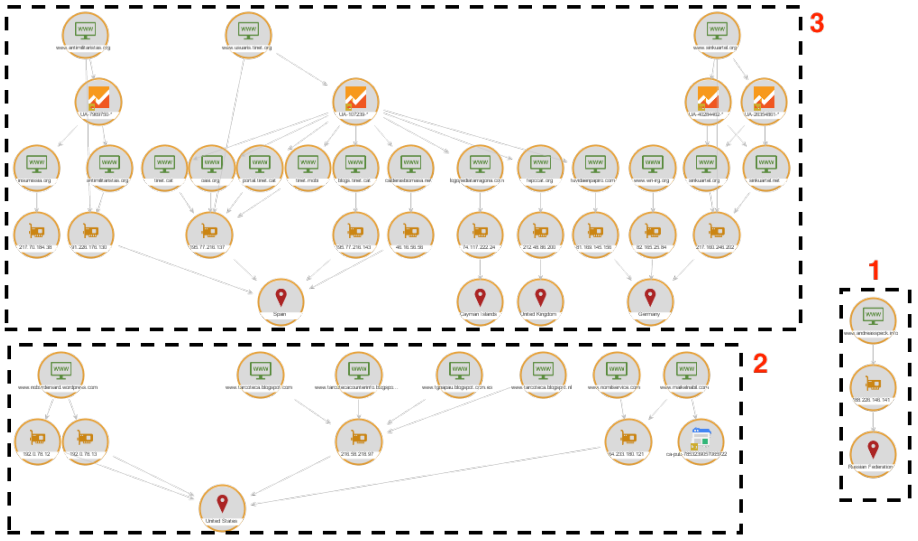

Fig. 8. The additional blogs that were identified using Maltego had three clusters labeled as 1, 2, and 3 based on their IP address geo-location.

more than 1,100 miles to exercise the units maintenance and leadership capabilities, and to demonstrate the freedom of movement that exists within NATO [16].

Many opponent groups launched campaigns to protest the exercise, e.g., "Tanks No Thanks" [17], which appeared on Facebook and other social media platforms, promising large and numerous demonstrations against the US convoy [18]. Czech President Milos Zeman expressed sympathy with Russia; his statements were echoed in the pro-Russian English language media and the Kremlin-financed media, i.e., Sputnik news [19]. The website, Russia Today (RT.com) also reported that the Czechs were not happy with the procession of the "U.S.Army hardware" [17]. However, thousands of people from the Czech Republic welcomed the US convoy as it passed through their towns, waving US and NATO flags, while the protesters were not seen.

During that time many deviant social bots or Impersonators bots were disseminating propaganda, asking people to protest and conduct violent acts against the US convoy. Bots are computer programs that can be designed and scheduled to perform various tasks on behalf of the bot's owner or creator. Impersonators bots are computer software that is designed to act like a human in social spaces, e.g., social media. These impersonators bots if they mimic human and try to do deviant behavior on social media are considered as Deviant Social Bots. Deviant social bots can influence people's opinion by disseminating propaganda or disinformation.

We identified a group of these bots (around 90 Twitter account) using Scraawl. Scraawl is an online social media analysis tool, available at www.scraawl.com.

We collected the social (friend and follower relations) and communication (tweet, mention, and reply relations) networks of these set of bots or botnets using NodeXL.

We analyzed the friends and followers network (i.e., the social network) of these bot accounts. The social network had two sub-networks, namely S1 and S2 as shown in Figure 9. The small sub-network, i.e., S2, contains only three nodes (i.e., a triad), hence it was rejected from further analysis, as it did not contribute much to the information diffusion. Since S1 is the largest sub-network - contained the majority of nodes - we examined this sub-network further.

Closer examination of the S1 sub-network revealed that the members of that network were more akin to a network we call as "syndicate" network, i.e., a network that has dense connections among their members and inter-group connections with the other groups and does not have a most central node, i.e., no hierarchy. Further examination of the within-group ties, revealed a mutually reciprocated relationship (the nodes followed each other), suggesting that the principles of "Follow Me and I Follow You" (FMIFY) and "I Follow You, Follow Me" (IFYFM) - a well-known practice used by Twitter spammers for "link farming", or quickly gaining followers [20] were in practice. This network had no central node, i.e., there was no single node feeding information to the other bots or seeder of information. This indicated the absence of a hierarchical organizational structure in the S1 network, in other words, no seeder was identified or observed. In cases where the seeder is not easily identifiable, other, more sophisticated methods are warranted to verify if this behavior truly does not exist. Although there might not be a single most influential node, a group of bots may be coordinating to make an influential group. To study this behavior further, we applied the Focal Structures Analysis (FSA) approach to find if any influential group of bots exist [21].

In this case study, deviant groups used a sophisticated tool to disseminate their propaganda and speed up the dissemination process by using botnets. These botnets were very sophisticated, i.e., botnets in the Dragoon Ride 2015 Exercise case required a more sophisticated approach to identify the organizers or seeders of information, i.e., it required applying FSA to both the social network (friends and followers network) and the communication network (tweets, replies, and mentions network). The evolution of complexity in the bots' network structures confirms the need for a systematic study of botnet behavior to develop sophisticated approaches and techniques or tools that can deal with predictive modeling of botnets. The details of this study and a caparison of these bots to the bots used during the Crimean water crises is published in [22] interested readers are encouraged to read it.

4 Conclusion

In this paper, we define social cyber forensics (SCF). We introduce a tool, i.e., Maltego that can be used to perform SCF. This tool uses open source information (OSINF) to connect various entities which can help an analyst answer various

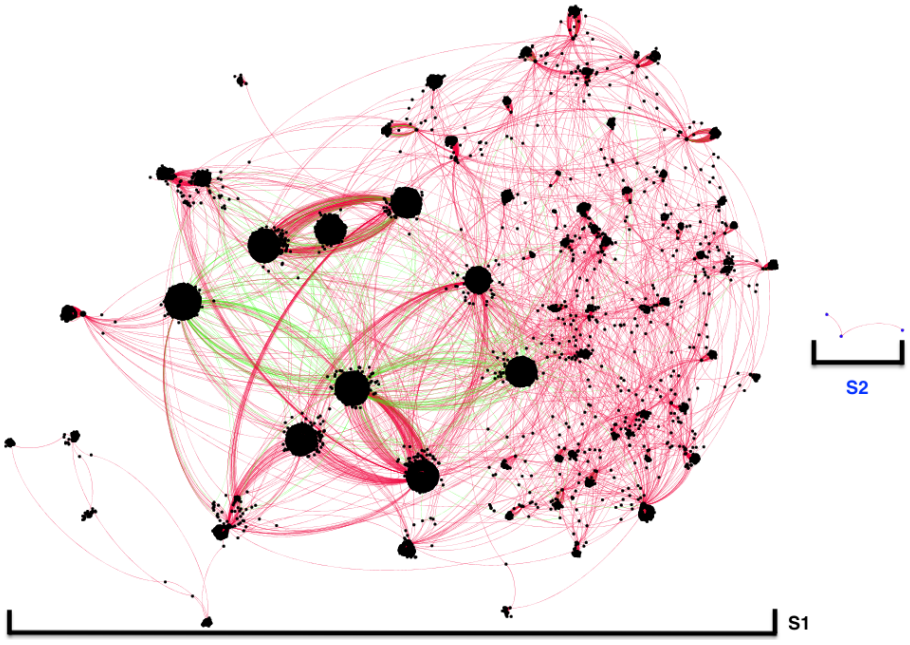

Fig. 9. Two sub-networks, S1 and S2. S1 and S2 are un-collapsed. Edges in green denote mutually reciprocal relations (bidirectional edges) while edges in red color denote non-reciprocal relations (unidirectional edges). Nodes are sized based on their in-degree centrality.

questions. We also provided three stepwise methodologies that can be performed using Maltego. Additionally, two case studies that leveraged the aforementioned methodologies were highlighted. In conclusion, open source information (OSINF) and social cyber forensics (SCF) provides a plethora of data that can be used to study various events and connect various entities, hence there is a need to develop more tools and methodologies that can serve this purpose. These tools and methodologies are helpful for various government agencies to study a deviant group, event, or behavior. It is also of interest to the industry sector as these tools and methodologies can help in marketing and advertising campaigns. Social cyber forensics bridges many disciplines such as computer science, social science, cyber forensics science, among others, hence its growth will also positively affect these disciplines.

Acknowledgements
This research is funded in part by the U.S. National Science Foundation (IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2605, N00014-17-1-2675), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-16-1-0189), U.S. Defense Advanced Research

Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Fund at the University of Arkansas at Little Rock, and Creighton University's College of Arts and Sciences. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

References

1. M. Huber, M. Mulazzani, M. Leithner, S. Schrittwieser, G. Wondracek, and E. Weippl, "Social snapshots: Digital forensics for online social networks," in *Proceedings of the 27th annual computer security applications conference*, 2011, pp. 113–122.
2. V. Juarez. Facebook status update provides alibi. [Online]. Available: http://cnn.it/2mUOo48
3. E. Grube. Assault fugitive who was found via facebook is back in NY. [Online]. Available: http://newyorkcriminallawyersblog.com/2010/03/assault-criminal-who-was-found-via-facebook-is-back-in-ny.html
4. M. Fisher. Facebook: a place to meet, gossip, share photos of stolen goods. [Online]. Available: http://www.washingtonpost.com/wp-dyn/content/article/2010/12/14/AR2010121407423.html
5. N. Alherbawi, Z. Shukur, and R. Sulaiman, "Systematic literature review on data carving in digital forensic," in *Procedia Technology*, vol. 11. Elsevier, 2013, pp. 86 – 92.
6. K. Oyeusi, "Computer forensics," phdthesis, London Metropolitan University, 2009. [Online]. Available: http://docslide.us/documents/computer-forensics-558454651e7df.html
7. D. Povar and V. Bhadran, "Forensic data carving," in *Digital Forensics and Cyber Crime*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer Berlin Heidelberg, 2011, vol. 53, pp. 137–148. [Online]. Available: http://bit.ly/2mzILFW
8. L. Alexander. Open-source information reveals pro-kremlin web campaign. [Online]. Available: https://globalvoices.org/2015/07/13/open-source-information-reveals-pro-kremlin-web-campaign/
9. M. Bazzell, *Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information*, 4th ed. CCI Publishing, 2014. [Online]. Available: https://inteltechniques.com/book1.html
10. S. Al-khateeb, M. N. Hussain, and N. Agarwal, "Social cyber forensics approach to study twitters and blogs influence on propaganda campaigns," in *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*. Springer, 2017, pp. 108–113.
11. N. OTAN. Trident juncture 2015. [Online]. Available: https://jfcbs.nato.int/trident-juncture
12. S. M. Research Foundation. Nodexl: Network overview, discovery and exploration for excel. [Online]. Available: http://nodexl.codeplex.com/wikipage?tit..
13. B. Etling, J. Kelly, R. Faris, and J. Palfrey, "Mapping the arabic blogosphere: politics, culture, and dissent," *Berkman center research publication*, vol. 6, 2009. [Online]. Available: http://www.ikhwanweb.com/uploads/lib/HNFNAB99APYNXAK.pdf

14. F. Şen, R. Wigand, N. Agarwal, S. Tokdemir, and R. Kasprzyk, "Focal structures analysis: Identifying influential sets of individuals in a social network," *Social Network Analysis and Mining*, vol. 6, no. 1, pp. 1–22, 2016. [Online]. Available: http://bit.ly/1qS8Y4D
15. S. Al-khateeb, M. N. Hussain, and N. Agarwal, "Leveraging Social Network Analysis and Cyber Forensics Approaches to Study Cyber Propaganda Campaigns," in *Social Networks and Surveillance for Society*, 1st ed., ser. Lecture Notes in Social Networks. Springer International Publishing, 2018, no. 2190-5428, p. 86. [Online]. Available: https://www.springer.com/us/book/9783319782553
16. D. M. A. DoD News. Operation atlantic resolve exercises begin in eastern europe. [Online]. Available: http://www.defense.gov/news/newsarticle.aspx?id=128441
17. RT. 'tanks? no thanks!': Czechs unhappy about us military convoy crossing country. [Online]. Available: http://www.rt.com/news/243073-czech-protest-us-tanks/
18. D. Sindelar. U.s. convoy: In czech republic, real-life supporters outnumber virtual opponents. [Online]. Available: http://www.rferl.org/content/us-convoy-czech-republic-supporters-virtual-opponents/26928346.html
19. Sputnik. Czechs plan multiple protests of US army's operation dragoon ride. [Online]. Available: http://sputniknews.com/europe/20150328/1020135278.html
20. S. Ghosh, B. Viswanath, F. Kooti, N. K. Sharma, G. Korlam, F. Benevenuto, N. Ganguly, and K. P. Gummadi, "Understanding and combating link farming in the twitter social network," in *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 61–70. [Online]. Available: http://dl.acm.org/citation.cfm?id=2187846
21. F. Sen, R. T. Wigand, N. Agarwal, M. Mete, and R. Kasprzyk, "Focal structure analysis in large biological networks," in *IPCBEE*, ser. 1, vol. 70. IACSIT Press, 2014. [Online]. Available: http://www.ipcbee.com/vol70/001-ICEEB2014-E0002.pdf
22. S. Al-khateeb, N. Agarwal, R. Galeano, and R. Goolsby, "Examining the use of botnets and their evolution in propaganda dissemination," *NATO Strategic Communication Center of Excellence (STRATCOM CoE)*, vol. 2, no. 1, pp. 87–112, 2017.