Information exchange paths in IRC hacking chatrooms

David Décary-Hétu, University of Montreal

Introduction

Criminals who wanted to socialize with like-minded individuals used to visit the local taverns where they could drink beers and learn a trick or two about their criminal trade. In the digital age, web pages, forums and online chat rooms have replaced local taverns. Cybercriminals can now find detailed online tutorials that teach them anything they could ever need on denial of service attacks, software piracy and credentials hijacking. Forums and chat rooms provide an environment where hackers can meet and casually discuss their problems and look for mentors or co-delinquents. These web 2.0 "taverns" are where social interactions are now happening and have proven to be rich environment for academic research.

The aim of this paper will therefore be to understand how hackers in the computer underground interact with each other and build their own personal networks. Information can either be shared by posting it publicly on forums, websites and blogs or by socializing with others through online communications. This paper will focus on the latter and examine not what information is exchanged but how it is exchanged. In order to do so, we will look at the personal social networks of hackers and build two partial correlation models that explain how individuals can maximize the amount and diversity of information they gather. Mentoring and social learning theory will be used as the theoretical framework of this paper.

Computer (in)security

Movies have a funny way of presenting hackers. In many cases (see Hackers, 1995; The Net, 1995; or The Matrix 1999), it seems that rapidly punching keys on a keyboard magically grants you access to whole computer systems. This illusion has been fed by many mediated "hacks" where the hacker merely guessed the password of a Hollywood star or used commercial phone-number spoofing services to gain access to voicemails (Sullivan, 2005). The reality, unfortunately for hackers and tabloid editors, is much more complex.

There is no disputing that corporations and individuals should be concerned with computer and network security, especially since the advent of broadband Internet in the beginning of the 2000s. Two new emergent trends in the criminal underground have drawn their attention particularly (Potlapally, 2011). First, while many attacks used to be the product of fame-driven hackers with modest skills, new threats now come from "well-funded criminal organizations or [...] sophisticated organizations with access to significant resources and talent" (Potlapally, 2011: p.93). This increases the chances of success of attacks while reducing the rates of detection by law-enforcement agencies. Second, hackers are now targeting

victims at a much lower level than the software layer meaning that their malicious software can run even after an operating system is erased and restored. This is done by forcing the computer to first run the malicious software and then to launch the operating system, giving the ultimate control of the software to the malicious software. By taking over the devices on which operating systems and applications run, hackers can more much easily bypass security systems and take advantage of them.

To meet these constantly evolving threats, network administrators have adopted a number of techniques (Ahmad & Habib, 2010). The first one is to securely encrypt all confidential data. Using commercial and open-source software, administrators can hide their most treasured information in encryption that would require centuries of cracking by large amounts of supercomputers. The second is intrusion detection systems (IDS). These programs monitor networks for suspect or unusual behavior and warn the security officers when such a behavior is detected. These systems have become incredibly complex and can be configured to detect changes in established connections, login/logoff timestamps and the amount of bandwidth used by a user. Lastly, almost every system connected to the Internet is now protected by a firewall, a virtual bouncer who controls who and what comes in and out of network. The first firewalls used blacklists where all behaviors were allowed by default and where the administrators blocked suspicious activities manually. Today, modern firewalls use whitelists where all behaviors are blocked unless they are expressly authorized.

If these technologies keep on evolving and becoming more complex, it is largely due to the attackers' ability to compromise targets. To hone their skills, hackers can easily access detailed online tutorials on how to hack computers. A simple query on Google with the phrase "how to hack" returned more than 97 million documents (Google, 2012). While most of these results are probably useless, a dedicated individual will likely find all the needed information to improve his skills or even get started in the hacking field through some of these websites.

Hackers can also use resources that are geared towards the legitimate security industry. Websites such as The Academy Pro (http://www.theacademypro.com) produce high-quality tutorials on how to configure and use defensive and offensive security software. These videos are intended for an audience of penetration testers and security professionals but provide in-depth insight on how software packages work and how to exploit them.

While these online sources do and can provide helpful tips to hackers, some papers such as Morselli's et al. (2006) paper has shown that social interactions are of the upmost importance in the

underworld. Their paper describes how a mentor in the criminal underworld can reduce the cost associated with a criminal career as well as increase illegitimate gains. Rogers (2000) also addressed the question of social capital and claimed that: "the area of learning theory may have the best chance at providing an understanding of hacking" (Rogers, 2000: p17). He analyzed theoretical concepts to determine how they applied to the problem of hacking, and concluded that Aker's social learning theory, although not a perfect fit, was the best theory available. Built on top of Sutherland's differential association (1939), Aker's theory can be summarized in the following four propositions (Rogers, 2010) which stipulate that people are more likely to commit crimes when they:

1. Differentially associate with other criminals;
2. Receive more reinforcement from their illegal actions than for their legal actions;
3. Are more exposed to deviant definitions and individuals than normal ones and;
4. Learn that committing crimes is normal and accepted.

These propositions revolve around three main concepts: differential association, differential reinforcement and definitions. Differential association originates from Sutherland's (1947) work and stresses the impact of close relationships between individuals. Different social groups influence people at each stage of their life. Family is most influential at an early age followed by school and peers at adolescence, and by neighbors and mass media later in life. The differential association varies in strength depending on the frequency, the duration, the time of first contact and the significance of the relationship. Differential association is therefore a distant cousin of peer pressure; as effective but with a much lower profile. Differential reinforcement refers to the reinforcement of past behavior. The probability that an individual will commit a crime depends on the positive or negative consequences of that act in the past. Once a criminal behavior has been reinforced, it is said to be incredibly difficult to change. Finally, definitions are any and all attitudes or perceptions that tend to indicate whether a behavior is right or wrong. General definitions are largely adopted values, such as religion, that are usually socially accepted. Specific definitions refer to opinions regarding a certain type of behavior. This explains how an individual can believe in high moral values but still commit a specific type of crime. Definitions that are favorable towards criminal behavior can be categorized in two types: positive definitions and neutralizing definitions. Positive definitions are those that consider certain types of criminal behavior as acceptable, whereas neutralizing definitions find that specific crimes are acceptable but only under a certain set of circumstances (ex: stealing to feed a family).

The social learning theory has been validated through empirical work in many settings. Akers et al. (1979) conducted a study on high school students and found that a significant portion of their criminal behavior could be explained by social learning theory. Skinner and Fream (1997) and others (Denning, 1998; Parker, 1998) also tested the validity of the social learning theory in the context of computer crime and found similar, although weaker, relationships between the propositions of social learning theory and criminal behavior.

Social learning theory is at its core a theory of flows and exchanges. Individuals interact with other and through that process learn techniques, values and information. In the case of Akers' (1979) students, this was done in face-to-face meetings. In the case of modern hackers, these interactions are quite different. As Rogers (2010) notes, hackers do not meet face-to-face. They instead use computer-mediated communications like online chat rooms, online forums, emails and instant messaging to stay connected. Although many channels are available to them, it appears that hackers are particularly fond of one in particular: IRC. The Internet Relay Chat (IRC) was invented by Jarkko Oikarinen in 1988 and is a synchronous group instant messaging application that allows users (clients) to connect to chat rooms on IRC servers and to exchange messages and files with each other (Reid, 1991).

Figure 1: Example of IRC software

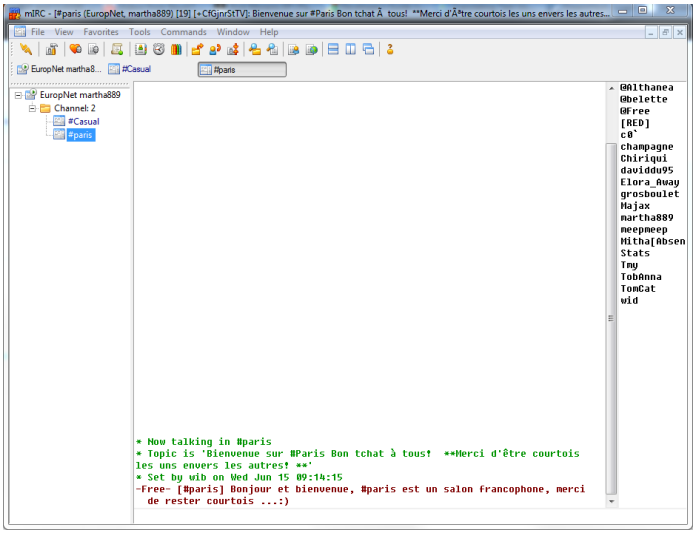


Figure 1 shows an example of an IRC client connected to the *casual* and *Paris* chat rooms on the EuropeNet server. There are hundreds of IRC servers all over the world each sporting their own list of chat rooms. As shown in Figure 1, all messages are public by default and are posted in the middle column of the client software. Users can exchange private messages and those cannot be seen by anyone else. Both

humans and computer program can coexist on IRC. Known as *bots,* computer programs can mimic the human behavior and provide certain services to other users. A bot can be programmed, for example, to ban from a chat room the users who curse too much. Such bots can be used monitor chat rooms, a feature we will be revisiting later in this paper.

It has been known for years that hackers gather and socialize in IRC chat rooms (Bratus, 2007). As a consequence, security researchers, journalists and law-enforcement agencies constantly monitor the shadier parts of IRC to gather intelligence on hackers. Although only the public messages can be accessed, the data captured in this fashion provides incredible insights in how hackers socialize and exchange information.

Hackers' social networks play an important role in their learning of new techniques and values. Morselli et al. (2006) and Rogers (2000) both highlight the importance of personal relationships in the transfer of human and cultural capital such as specialized skills and rational justifications. By understanding how knowledge and perceptions flow amongst hackers, we hope to increase the security of computer networks. Monitoring online channels could become, we believe, an early warning system for upcoming threats. By looking at the network features of hacking chat rooms in IRC, this paper aims to understand the structure of the hackers' personal networks, particularly those that are the most involved in the criminal underground. Since IRC is a nexus of interactions for hackers, much of the mentoring and information sharing should happen through its chat rooms. By looking at the flow of information inside them, it will be possible to understand how hackers develop their skills and their network of contacts. This better understanding of these deviant individuals will enable us to build better prevention programs and stop the flow of information and its negative impact on society.

Data

To understand how hackers exchange information, we focused on one of the most bustling environment for hackers, the Internet Relay Chat (IRC). A North-American law-enforcement agency who has been monitoring IRC channels over the past few years agreed to share some of the data it has gathered over the past few years with us. This paper is based on this data the police collected from a single IRC server. Table 1 provides the descriptive characteristics of this sample data. It covers a 30 month period ranging from January 2009 to June 2011.

Table 1: Characteristics of study sample

Length of monitoring	30 months (January 2009 - June 2011)

Number of chat rooms monitored (monthly)	7-15
Number of events	16,978,269
Number of relevant messages	2,232,729
Number of unique relevant messages	1,618
Number of hackers	262
Number of people in social network of hackers	356
Number of ties (reciprocal)	6,168

The number of hacking chat rooms monitored each month fluctuated over this period with a low of 7 and a peak of 15. The size of the data collected was massive as more than 16 million events were registered. These events could be public messages, users that login/logout or users that change their nickname. Making sense of such a massive dataset of interconnections proved to be very problematic at first. In order to simplify our analyses, we decided to focus only on specific messages that were explicitly related to the world of hacking. To do so, we created a list of 79 discriminating keywords to extract the hacking messages from the mass of data. The list contained 79 words in order to cover a large scope of criminality (ex: warez, botnet, ddos).

Two million messages contained at least one of the 79 aforementioned keywords. They unfortunately included a great deal of duplicates as the monitored chat rooms were often filled with bots that spammed visitors with the same message over and over again. These messages were mostly used to advertise illegal goods and services for sale. To set apart the real users from the bots, we calculated the number of messages posted by each nickname and divided it by the number of distinct messages posted by that nickname in our database. This provided us with an index of the uniqueness of their messages. Anyone who had messages that were unique less than 90% of the time was discarded and categorized as a bot. This left us with a sample which only included people who posted original messages 9 times out of 10. While some individuals may have been eliminated during this classification process, it also ensured that all bots were eliminated from the sample. The end result was a sample of 262 individuals who posted 1,618 messages containing each at least one keyword between January 2009 and June 2011. While fairly limited, the resulting dataset is still large enough to provide significant and interesting statistical analysis.

Methodology

As this paper focuses on information exchange paths between individuals, we began by looking at the relationships between these 262 individuals and their contacts in IRC chat rooms. Doing so proved challenging as our dataset only included public chat rooms records where many individuals participated simultaneously in public discussions. Identifying precisely the flow of discussions as well as the individuals

who were present when the messages were posted was unfortunately not possible with the dataset provided to us by the law-enforcement agency. We therefore had to resort to a less than perfect solution nut one that allowed us to build the social map of each hacker in our sample. We decided on a methodology which provided us a good approximation of each individual's social network that could be done automatically and without the need for costly content analysis.

In order to do so, we grouped all messages in blocks of 20 units around each message containing a keyword. We then considered that all individuals who had posted a message within that window of 20 messages were in a relation of some sort and we built an undirected relational matrix around these ties. Our decision to group messages in blocks of 20 was not random. We analyzed a sample of messages and determined that conversations were usually fairly short and that the ten messages before and after a message containing a keyword were almost always linked to the same discussion. Occasionally, there would be concurrent discussions in public chat rooms but this was not the norm by far. Using this methodology, it is possible that our relational matrix included ties that did not exist or left out ties from people who were merely watching and not participating in the conversations. In the first case, our sample analysis shows that the signal to noise ratio is very high and that the added ties did not have a significant impact on our dataset as they were far and few in between. As for the second limit, our dataset did not allow us to include into our analysis the individuals reading but not participating in the conversations. These individuals did profit from the information exchange but did so discreetly and additional research design will need to be developed to monitor these individuals as well. This paper will therefore only focus on individuals who actively participated in chat room discussions.

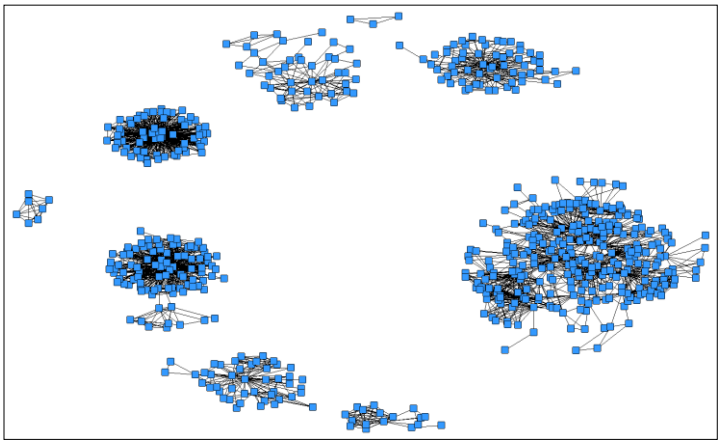
Social network analysis (SNA) can either work with directed or undirected matrices. Directed matrices are used when the direction of a tie can be determined (ex: who visited who, who called who). In the case of public discussions, it is often difficult to determine the intended recipient or recipients of a message. People tend to shout messages and wait for an answer from any of the members present at the time. Rather than guess the direction of messages and ties in this network, we opted to build an undirected matrix where all ties are reciprocal. Our analysis will take into account this research design when determining the impact and structure of the personal networks of hackers.

A relational matrix representing each of the 262 hackers' social graph was built using the aforementioned blocks of conversations. We found a total of 356 individuals who were tied to (spoke with) the 262 IRC users who had posted at least once a message containing a keyword. These 356 individuals never used one of the 79 keywords but were involved in conversations where that keyword

was spoken so to speak. The sample for this study therefore includes 262 IRC hackers and their 356 contacts.

Figure 2 displays the structure of the ties between the hackers and their social graph. It is clear by looking at this figure that most users did not visit more than one chat room. Most clusters are made from the messages of a single chat room and there is very little overlap between them even though all chat rooms were centered on the hacking underground.

Figure 2: Structure of ties in hacking chat rooms



Given the structure of the network, we decided to focus on the ego or personal networks of individuals rather than look at the structure as a whole. The IRC data at our disposal did not include one unique network but nine separate networks that were linked by their origin on the Internet Relay Chat. Furthermore, as we intended to focus on the individual features of hacker networks rather their structure, an approach focusing on the individual networks appeared as the most sensible.

Studying personal networks is known in the social network field as ego network analysis (Wasserman & Faust, 1994). The term *ego* refers to the individual at the center of a network and the term *alters* refers to the people linked to the ego. Although this method has been used in many social science studies (Kalmijn & Vermunt, 2007; Degenne & Lebeaux, 2005; Stefanone & Jang, 2008), it has yet to be adopted widely in criminology. Analysis of ego networks in the context of criminal networks are few and far in between. Malm et al. (2011) studied the co-offending of criminal organizations in Canada. In this research, they determine that certain types of organizations tend to co-offend more often with other groups while others tend to choose their co-offenders based on their membership to the same group.

Past research such as Malm et al.'s (2011) have focused mainly on four social network metrics: centrality, betweenness, composition and density. Centrality is frequently used to assess the prominence of actors within a network (Wasserman & Faust, 1994: 172). It indicates the number of incoming and outgoing contacts and account for the direction of direct ties around each node (in directed networks). The pattern of ties originating from or sent to a network member is usually a reliable indicator of this person's prestige or status (Wasserman & Faust, 1994: 174) as it helps distinguish people with sought-after expertise. Betweenness measures the extent to which a node mediates between other nodes by its position along the geodesics (the shortest paths between two nodes) within the network, thus providing useful additional information on the structure of a network. The more often a node is located along the geodesics, the higher its betweenness centrality, making it a broker within the network. The position of broker has been associated with the notion of power in networks (Prell et al., 2008; Morselli, 2009; Toral et al., 2009) since these individuals control the flow of information between the different actors. They can decide whether to allow messages to pass through, modify the information, or simply ignore it. Composition is a measure of the heterogeneity of the network (Malm et al., 2011). It measures whether individuals associate with others that share common characteristics (ex: if boys talk more to boys than girls). Finally, density measures the number of actual ties compared to the number of possible ties (Wasserman & Faust, 1994). This metric gives a sense of the implication of individuals in a network; the denser the network, the more involved its participants are.

The first part of our analyses builds on Malm et al.'s (2011) work and focuses on a limited number of social network metrics: centrality, betweenness, density and longest distance. The dataset provided by the law-enforcement agency did not contain any personal characteristics of the hackers so it was not possible unfortunately to run composition analyses on our research data. We instead focused on the longest distance, a measure of the number of steps any node needs to reach any other node in the ego network. For each of these four social networks metric, the quartiles are presented in the first part of our analyses to generate a general understanding of the structure of each hacker's social network.

In the second part of our analyses, we will present a partial correlation which is a normal correlation that takes into account the number of nodes in each personal network. Differences in sizes may have tainted our results otherwise. Six variables (centrality, density, betweenness, 2 step reach, number of keywords and number of conversations) will be included in two models which aim to

understand how hackers build their social network to maximize their flow of information. In the first model, the correlation measures the relation between the number of different keywords used by each hacker and the five other variables. This is intended a as proxy of the diversity of topics discussed by each hacker. In the second model, the correlation measures the relation between the number of conversations of each hacker and the five other variables. This is also a proxy but for the bandwidth of information directed to and from each hacker. Rather than use longest distance, we instead opted to integrate the notion of 2 step reach which measures the number of nodes which can be reached within 2 ties in a given ego network. This provides a better understanding of the importance of weak ties or friends of friends in the context of information flows.

Results

Table 2 presents a summary of the social network metrics for the ego networks of hackers who visited hacking chat rooms.

Table 2: Quartiles of social network metrics

	Centrality	Betweenness	Density	Longest Distance
Min	1	0.00	0	0
25th	4	0.00	55	1
50th	7	0.00	100	1
75th	12	9.86	100	2
Max	65	100.00	100	6

N = 262

In our sample, hackers are connected to a limited number of alters in general (Minimum = 1; Maximum = 65). 25% of egos are tied to 4 alters or less and 75% of egos are connected to 12 alters of less. Since each conversation we flagged included a potential of 21 individuals (10 messages before and after plus the ego), a median of 7 connections is fairly low. That being said, there are still much variation in the sample population.

As for betweenness, the normalized version of the metric was used to eliminate the impact of the variations in the number of nodes in each ego network. Our results show that most hackers are very poor brokers. With half of our sample displaying a betweenness of 0.00, the number of ties between alters that transit through the egos are very low. Some egos are more essentials than others. The last 25% of our sample have a betweenness score that varies between 9.86 and 100.00.

*G*iven the limited size of most ego networks, it is not surprising to see that more than half of our sample displays density scores of 100. Once again, a select few hackers have different network configurations with lower densities than the others. Such high density in general means that the ego networks are close-knit groups where every actor knows each other. The undirected nature of our matrix increases however the density of the ego networks.

Finally, with half the sample showing a longest distance of 1, most ego networks are once again homogenous entities with low centrality and high density. The maximum number of steps needed to reach any node on the largest ego network is 6, a special number in social network analysis ever since Watts' book *Six Degrees* (Watts, 2004). Our data suggests this magic number still applies even in the case of chat rooms.

Table 3: Partial correlation of diversity of interests and involvement in hacking chat rooms

	Nb of keywords	**Nb of conversations**
Centrality	-0.352**	-0.140**
Density	-0.274**	-0.113**
Betweenness	0.369**	0.169**
2 Step Reach	0.102**	n.s.
Nb of keywords	---	0.819**
Nb of conversations	0.819**	---

N = 618

Table 3 presents the partial correlation models between the number of unique keywords and the number of conversations. Correlations are all statistically significant except for the number of conversations and 2 step reach.

The number of keywords is moderately and negatively correlated with the number of ties and the density (-0.352 and -0.274 respectively). This could be the result of the structural holes paradigm (Burt, 2010). It is thought that close-knit networks have access to more redundant data since all the actors in the network know each other and talk about the same subjects. It could be the case that these smaller and denser networks are experts in a limited range of subjects and thus only mentioned a limited number of our keywords. Betweenness and 2 step reach are moderately and poorly, but positively, correlated to the number of keywords (0.369 and 0.102 respectively). As mentioned before, those with greater betweenness usually display more structural holes in their networks and can access more diverse sources of information. They can then expand their knowledge on a variety of subjects. This increased number of

alters is reflected in the 2 step reach metric which is correlated positively to the number of keywords. It shows once again that having access to a greater number of nodes (directly or indirectly) increases the available knowledge base.

The number of conversations is negatively and poorly correlated to the number of ties and density. Individuals with a greater number of ties have fewer conversations. As we have mentioned before, the conversations in hacking chat rooms seem to be concentrated amongst tight and small networks. It is therefore normal to see that larger ego networks show lower levels of conversations. It is surprising though that the density is negatively correlated with the number of conversations. This could be explained by the fact that we are only monitoring the public messages on IRC chat rooms and not the complete communications channels between the individuals. Some of the conversations amongst individuals could be conducted over private messages or instant messaging. It could also be that certain smaller units would rather move their discussions to more private settings rather than discuss publicly. The number of conversation is positively correlated to the betweenness. With a greater level of communication, the chances that a node will sit in the middle of a dyad is higher and it is thus normal to see these two figures positively correlated though we might have expected a slightly higher correlation between these two metrics.

Discussion

The aim of this paper was to understand the structure of personal networks of hackers who are involved in criminal activities. The first part of our analyses demonstrated that most individuals only had a limited number of contacts in their social graph. This enables them to build stronger relations with other IRC users from whom they may be learning the hacking skills they need. This configuration supports Rogers' (2000) supposition that social learning theory could very well apply to the world of hacking. Intimate relationships are needed for differential association and differential reinforcement to occur. With small ego network structures, the winning conditions are met for differential association and differential reinforcement to happen and for mentor relationships to form.

The betweenness metric was also very interesting as it demonstrated that direct ties were more important than indirect ties. This increases once again the bonds between the players and increases the chances of mentoring and association with other criminals. This preference for direct ties could be a consequence of the IRC setting. Each user can contact any other user when they want, lowering the barrier of contacts. This could also be the consequence of the low level of trust in the hacking community. As

individuals are unable to ascertain the true identity of their counterpart, they should prefer direct contacts to limit their exposure to law-enforcement agencies.

Density and longest distance also confirm the previous statements as density is mostly high and longest distance is mostly low. Both settings are fit for the transfer of differential association and differential reinforcement as well as for mentoring relationships to form. This network structure indicates that the hackers are more inclined to participate in small communities rather than large ones. In these small circles, every member knows everyone else. In this context, there should be a higher level of trust amongst its members. This should stimulate the involvement and differential reinforcement in the hacking culture and community. This should also generally raise the security levels of these communities. Any outsider wishing to join these small communities will have to demonstrate that they are worth other people's time and that they too belong with them. This significantly raises the difficulty that law-enforcement agencies and academic researchers will face when trying to get a deeper understanding of these hacking communities and they will need to invest time and resources to build credible online personas which could gain access to these small circles. Public discussions may be accessible to any and all but the real data will undoubtedly come from private messages and private chat rooms for which personal contacts are needed. These close-knit groups may however be at times the very weakness that police forces need. The case of Sabu, a leader of the Anonymous hacking movement, was arrested and served as an undercover agent for the Federal Bureau of Investigations for months (Cluley, 2012). This allowed direct access and monitoring of the criminal underground by the FBI who took advantage of the network structure of the Anonymous community.

The partial correlations further enhance our understanding of the criminal underground by evaluating the ego network structure of the individuals based on the bandwidth and the diversity of information they have access to. Both concepts of bandwidth and diversity are highly correlated together at 0.819**. Individuals who manage to position themselves at the nexus of information flows have access to more information and information that is more varied. These individuals would make ideal investigation targets and they would be able to inform us on what is happening in the hacking underground and to show us how one should position himself in the network. These individuals may also be good mentors as their developed network of information could enhance their recognition and power amongst hackers.

Both correlation models are very similar. The number of keywords and the number of conversations are bot correlated with centrality, density and betweenness. Only the number of keywords is however associated with the 2 step reach. In both cases, the strongest correlation is with betweenness

followed by centrality, density and 2 step reach. Our models show that access to large amounts of diversified information is correlated to a small network that is not dense where the ego is playing a role of broker. The ego also benefits from friends of friends as demonstrated by the 2 step reach. In this context, hackers who would like to improve their hacking skills should decide carefully with whom they associate (centrality). The number of contacts is not what matters; it is who these contacts are, a recurrent theme in criminological research on social networks (Coles, 2001). Hackers should also work to limit the ties between his contacts as this reduces his importance in his ego network (Burt, 2010). Structural holes or holes between clusters of friends provide opportunities for egos who can capitalize on the needs of his contacts. In this context, structural holes provide the ego with more information and one that is more diversified. This reinforces the role of brokers in criminal networks as measured by the betweenness measure. Brokers control the flow of information in networks as other people use them to transit information. In the case of IRC hackers, brokers have access to more information this increases the chances of them improving their hacking skills.

Mentors and social learning theory are both supported by these partial correlation models. The ego network structure described in the discussion is ideal for the appearance of mentors who can use differential association, reinforcement and definitions to recruit and teach new pupils. Individuals with access to more information and more diversified information would make ideal mentors as they could share their strategic positioning in the network as well as the information they have access to with their protégés. Such individuals would be at the center of information flows and would undoubtedly be recognized in the hacking community as more efficient in their craft. While most hackers have ego networks that are compatible with mentors and social learning theory, a select few have network structures that match those of mentors who can use their influence to teach the hacking tricks to the next generation of criminals. Using social network analysis, these individuals can be targeted in order to break the cycle and limit the upcoming threats coming from the computer underground. The network of hackers presented in this paper therefore represents a fertile ground for mentors and hackers to meet and thrive, ensuring them more illegitimate gains and lower risks of detection.

Conclusion

Monitoring IRC channels has proven very effective at providing an understanding of the hacking community. There are, however, many technical and ethical challenges that need to be addressed. As mentioned before, the information available in IRC chat room is limited as many conversations happen in private (and password-protected) channels or in private messages. These cannot be monitored and

skew the real nature of hacking networks. Furthermore, monitoring IRC chat means dealing with millions of messages and events which can be time-consuming to analyze. This paper offers a simple yet effective way of dealing with this problem but more complex solutions may be necessary to provide a better picture of the criminal underground. Future research may try to group messages by their timestamp to give a better representation of the network. Some chat rooms may have been silent for hours before someone posted a new message and simply grouping messages in blocks did not take this fact into account. By grouping messages in blocks of 2 or 5 minutes, one would limit the risk of creating ties where there are now. Researchers should also study carefully the login and logoff events in each chat room as they would indicate whether individuals were reading the messages even though they were not participating actively in the conversations. This would involve building a computer algorithm that would create an overview of each hacker's activity – where he was and when. Although difficult, this process is not insurmountable.

New ethical guidelines will also need to be developed for this new research environment. In this paper, all IRC server names, chat room names and nicknames were obfuscated in order to protect their anonymity and that of the law-enforcement agency that provided us with the data. Such techniques should be the norm in social research. While this limits the duplication of studies, it provides enough information to evaluate the strength of studies and to test the hypotheses in other chat rooms. It also ensures that hackers do not feel muffled by too many researchers who may all try to monitor their chat rooms at the same time. By using discrete bots to monitor the chat rooms, we ensure that we are not disturbing the ecosystem as hackers may very well move elsewhere once they realize that their private chat rooms are infested with researchers and law-enforcement agencies.

Studies on IRC individuals are by no means new. Computer professionals have been using them for years but with their own technical angle. This paper brings a practical framework that can be harnessed to better understand the social interactions of hackers online. This hybrid study which features the most up-to-date technical tools with tested criminological theories offers a new take on the problem of hackers, and one that we hope will be picked up and replicated amongst more datasets in the upcoming years.

References

Akers, R. L. & M. D. Krohn & L. Lanza-Kaduce & M. Radosevich. (1979). "Social Learning And Deviant Behavior: A Specific Test Of A General Theory." *American Sociological Review*. 44(4): 636-655.

Bratus, S. (2007). "Hacker Curriculum: How Hackers Learn Networking." *IEEE Distributed Systems Online*. 8(10).

Burt, R. S. (2010). "Structural Holes In Virtual Worlds." *Working Paper*. University Of Chicago Booth School of Business, Chicago, IL.

Cluley, G. (2012). "LulzSec Leader Sabu Betrayed Anonymous Hackers, Reports Claim." Retrieved on June 12th 2012 on: http://nakedsecurity.sophos.com/2012/03/06/sabu-lulzsec-betrayed-anonymous-hackers/.

Coles, N. (2001). "It's Not What You Know – It's Who You Know That Counts: Analyzing Serious Crime Groups As Social Networks." *British Journal Of Criminology*. 41(4): 580-594.

Décary-Hétu, D. & C. Morselli & S. Leman-Langlois (2012). "Welcome to the Scene: A Study of Social Organization and Recognition Among Warez Hackers." Journal Of Research In Crime And Delinquency. 49(3).

Degenne, A. & M. Lebeaux. (2004). "The Dynamics Of Personal Networks At The Time Of Entry Into Adult Life." *Social Networks*. 27: 337-358.

Denning, D. (1998). *Information Warfare And Security*. Addison-Wesley: Boston, USA.

Gromov, G. (2001). "Roads And Crossroads Of The Internet History." Retrieved on January 4th, 2012: http://www.netvalley.com/cgi-bin/intval/net_history.pl.

Kalmijn, M. & J. Vermunt. (2007). "Homogeneity Of Social Networks By Age And Marital Status: A Multilevel Analysis Of Ego-centered Networks." *Social Networks*. 29: 25-43.

Malm, A. & G. Bichler & R. Nash. (2011). "Co-offending Between Criminal Enterprise Groups." *Global Crime*. 12(2): 112-128.

Mitnick, K. (2011). *Ghost In The Wires: My Adventures As The World's Most Wanted Hacker*. New York, NY: Little, Brown & Company.

Morselli, C. & P. Tremblay & B. McCarthy. (2006). "Mentors And Criminal Achievement." *Criminology*. 44(1): 17-43.

Morselli, C. (2009). "Law-Enforcement Disruption Of A Drug-Importation Network." *Studies Of Organized Crime*. 8:1-17.

Parker, D. (1998). *Fighting Computer Crime: A New Framework For Protecting Information*. John Wiley & Sons: New York, USA.

Prell, C. and K. Hubacek and M. Reed. (2008). "Who's In The Network?" *Systemic Practice And Action Research*. 21: 443-458.

Reid, E. (1991). *Electropolis: Communication And Community On Internet Chat Relay*. Masters at University of Melbourne, Australia.

Rogers, M. (2000). "Psychological Theories Of Crime And Hacking." Retrieved on March 6th 2012 on: http://www.dvara.net/HK/theory_crime_hacking.pdf.

Rogers, M. (2010). "The Psyche Of Cybercriminals: A Psycho-Social Perspective." IN Ghosh, S. & E. Turrini. *Cybercrimes: A Multidisciplinary Analysis.* New York, NY: Springer.

Skinner, W. F. & A. M. Fream. (1997). "A Social Learning Theory Analysis Of Computer Crime Among College Students." *Journal Of Research In Crime And Delinquency*. 34(4): 495-518.

Stefanone, M. A. & C. Jang. (2008). "Writing For Friends And Family: The Interpersonal Nature Of Blogs." *Journal Of Computer-Mediated Communication*. 13: 123-140.

Sullivan, B. (2005). "Cell Phone Voicemail Easily Hacked." Retrieved on August 12th, 2011: http://www.msnbc.msn.com/id/7046776/ns/technology_and_science-wireless/t/cell-phone-voicemail-easily-hacked/.

Sutherland, E. H. (1939). *Principles Of Criminology*. Philadelphia, PA: Lippincott.

Toral, S. and M.R. Martinez-Torres and F. Barrero. (2010). "Analysis Of Virtual Communities Supporting OSS Projects Using Social Network Analysis." *Information & Software Technology*. 52(3).

Wasserman, S. & K. Faust. (1994). *Social Network Analysis: Methods And Applications*. Cambridge, UK: Cambridge University Press.

Watts, D. (2004). *Six Degrees: The Science Of A Connected Age*. New York, NY: W.W. Norton & Company.