Global Crime
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/fglc20

The social network of hackers

David Décary-Hétu [a] & Benoit Dupont [a]
[a] School of Criminology, Université de Montréal, Montréal, QC, Canada

Version of record first published: 09 Jul 2012

To cite this article: David Décary-Hétu & Benoit Dupont (2012): The social network of hackers, Global Crime, DOI:10.1080/17440572.2012.702523

To link to this article: http://dx.doi.org/10.1080/17440572.2012.702523

Global Crime
i*First*, 2012, 1–16

The social network of hackers

David Décary-Hétu* and Benoit Dupont

School of Criminology, Université de Montréal, Montréal, QC, Canada

Social researchers are facing more and more challenges as criminal networks are expanding in size and moving to the Internet. Many efforts are currently under way to enhance the technical capabilities of researchers working in the field of cybercrimes. Rather than focusing on the technical tools that could enhance research performance, this article focuses on a specific field that has demonstrated its use in the study of criminal networks: social network analysis (SNA). This article evaluates the effectiveness of SNA to enhance the value of information on cybercriminals. This includes both the identification of possible targets for follow-up research as well as the removal of subjects who may be wasting the researchers' time. This article shows that SNA can be useful on two levels. First, SNA provides scientific and objective measures of the structure of networks as well as the position of their key players. Second, fragmentation metrics, which measure the impact of the removal of *n* nodes in a network, help to determine the amount of resources needed to deal with specific organisations. In this case study, a tactical strike against the network could have had the same destabilising impact as a broader approach. The resources saved by limiting the investigation targets could then be used to monitor the criminal network's reaction to the arrests and to limit its ability to adapt to the post-arrest environment.

Keywords: organised crime; social network analysis; hackers; botnet

Introduction

Social researchers are facing more and more challenges as criminal networks are expanding in size and moving to the Internet. Today's research requires massive investments in human and material resources, which in turn generate a wealth of information that needs to be analysed carefully. Many efforts are currently under way to enhance the technical capabilities of social researchers working in the field of cybercrimes. Rather than focus on the technical tools that could enhance researchers' performance, this article focuses on a specific field that has demonstrated its use in the study of criminal networks: social network analysis (SNA). This framework is used to identify the structure of networks as well as the relative position of its participants. Morselli and others have used SNA to evaluate police investigations *ex post facto* and found that SNA metrics were correlated with certain roles in criminal organisations.[1] These studies, which focused on more traditional crime settings such as the drug trade, proved the effectiveness of SNA in the evaluation of police work.

*Corresponding author. Email: david.hetu@gmail.com

1. Morselli, Carlo, and Katia Petit, 'Law-Enforcement Disruption of a Drug Importation Network', *Global Crime* 8, no. 2 (2007): 109–30.

ISSN 1744-0572 print/ISSN 1744-0580 online

http://dx.doi.org/10.1080/17440572.2012.702523
http://www.tandfonline.com

The aim of this article is slightly different and tests the value of SNA in a new environment, the Internet. This article evaluates the effectiveness of SNA to enhance the value of information on cybercriminals. This includes both the identification of possible targets for follow-up research and the removal of investigation subjects who may be wasting the researchers' time.

The data used in this study come from the largest police operation ever carried out against computer hackers in Canada. This investigation ended with the arrest and conviction of 10 hackers (known as botmasters or botherders) who were running botnets. Botnets are network of compromised computers that can be remotely controlled by botmasters. These computers can be used to send spam, steal personal and financial information or launch attacks against other computers and websites. It is not uncommon to see botnets of tens if not hundreds of thousands of computers. The cooperation of a Canadian law enforcement agency allowed us to access the raw conversation logs of the arrested hackers and, using SNA, to travel back in time to look at the targets of the investigation to determine the structure of the ties between hackers as well as whether any people of interest (POI) might have been overlooked by the investigators.

The first section of this article presents an overview of the role that SNA has played so far in social research on criminal networks. The following section details the data and methodology that leverage the latest developments in SNA software such as the Keyplayer application.[2] The third section presents the social network metrics of the hacker network that include the centrality, betweenness and political independence (PLI) measures. It also features an analysis of the fragmentation delta of the network following multiple scenarios. The article then concludes on how SNA can contribute to social research on hacker networks.

Information overload and SNA

Today's criminal networks have very little in common with those that formed even 40 years ago. Of the researchers who have been studying the shape and size of criminal networks and organisations since the 1960s, Cressey's work on the Italian mafia is still today the most famous example of such an analysis.[3] His hierarchical view of organised crime has been challenged many times, however, and empirical data suggest that criminals are more likely to be involved in loosely associated illicit networks rather than formal organisations.[4] In Morselli's study of a drug-importation network, for example, the number and position of the individuals involved increased and diversified over time.[5] Players who were central at the beginning of the operation moved to the periphery while others left and new players joined. Criminal organisations (or networks) are now, according to Williams, much more resilient and robust.[6] This 'stems not only from the capacity to limit the damage that is

2. Borgatti, Stephen, P., 'Identifying Sets of Key Players in a Social Network', *Computational and Mathematical Organization Theory* 12, no. 1 (2008): 21–34.
3. Cressey, Donald, *Theft of the Nation: The Structure and Operations of Organized Crime in America* (New York: HarperCollins, 1969).
4. Morselli, Carlo, *Inside Criminal Networks* (New York: Springer, 2009); Reuteurs, Peter, *Disorganized Crime: Economics of the Visible Hand* (Boston, MA: MIT Press, 1983).
5. Morselli, *Inside Criminal Networks*.
6. Williams, Phil, 'Transnational Criminal Networks', in *Networks and Netwars: The Future of Terror, Crime, and Militancy*, eds. John Arquilla and David Ronfeldt (Santa Monica, CA: RAND Corporation, 2001).

inflicted but also from the ability to mitigate consequences'.[7] Removing a number of players from a network does not necessarily destabilise the whole network because redundant ties can maintain their connections, allowing the network to continue to operate.

In this context, drawing a complete map of an organisation and the roles of its members becomes increasingly difficult. Despite this, understanding how an organisation is structured, adds new recruits and evolves over time is a key in determining its weakest links and ultimately being able to disrupt its operations.[8] In criminal networks, chances are that the most visible or exposed individuals will be the first to draw the attention rather than the actual key players. It is thus important to be able to validate intuitions with scientific tools.

The complexity of today's criminal networks have led to an increase in the amount of information gathered through monitoring. As Coffman et al.[9] demonstrate, the problem for police and intelligence agencies is not finding more data on criminals but finding ways to meaningfully analyse the data they have. In addition to traditional data such as physical monitoring, paper trails and phone taps, police now have access to emails, computer hard drives, online chats and voice over IP calls. In a recent police investigation of a lone hacker who had attacked a university network, investigators were given over 300 GB of computer logs, the equivalent of 6 million pages of text.[10] Academic researchers face a similar problem when studying criminal networks as they often use police data as their source of information. Furthermore, some academics have also gathered data on criminal markets by their own means and these datasets can include thousands if not tens of thousands of individuals.[11]

The need for tools that can easily and rapidly analyse large datasets has been met through modern software packages such as NetDraw and Palantir, which enable users to visualise data and can present the information from the perspective of a specific actor, giving an understanding of the structure of criminal as well as 'dark' networks.[12] In addition to such tools, many researchers have also adopted SNA to uncover terrorist or insurgent networks.[13] The underlying principle of SNA is to use relational ties to derive the structure of a network as well as the position of each of its participants.[14] Different coefficients such as centrality, betweenness and power can be computerised in order to rank the actors in a network and determine their level of interdependence. Two factors explain the SNA's attractiveness. First, it enables the easy and rapid transformation of data so that it can be

7. Ibid., p. 80.
8. Ressler, Steve, 'Social Network Analysis as an Approach to Combat Terrorism: Past, Present and Future Research', *Homeland Security Affairs* 2, no. 2 (2006): 1–10.
9. Coffman, T., S. Greenblatt, and S. Marcus, 'Graph-Based Technologies for Intelligence Analysis', *Communications of the ACM* 47, no. 3 (2004): 45–7.
10. Rioux, Alain, 'Gestion D'incident: Piratage dans les Universités', *Présentation dans le cadre du Colloque FRANCOPOL sur la cybercriminalité* (Nicolet, Canada, 2011).
11. Décary-Hétu, D.,C. Morselli, and S. Leman-Langlois, 'Welcome to the Scene: A Study of Social Organization and Recognition among Warez Hackers', *Journal of Research in Crime and Delinquency* (2011).
12. Xu, Jennifer, and Hsinchun Chen, 'Criminal Network Analysis and Visualization', *Communications of the ACM* 48, no. 6 (2005): 101–7; Yang, C., N. Liu, and M. Sageman, 'Analysing the Terrorist Social Networks with Visualization Tools', *Intelligence and Security Informatics* (2006): 331–42; and Raab, Jörg, and H. Brinton Milward, 'Dark Networks as Problems', *Journal of Public Administration Research and Theory* 13, no. 4 (2003): 413–39.
13. van Meter, Karl M., 'Terrorist/Liberators: Researching and Dealing with Adversary Social Networks', *Connection* 24, no. 3 (2002): 66–78.
14. Wasserman, Stanley, and Katherine Faust, *Social Network Analysis: Methods and Applications* (Cambridge: Cambridge University Press, 1994).

used and analysed through specialised software such as Ucinet.[15] Second, SNA metrics can be used to measure the involvement and position of an individual in a network, providing an insight as to his or her relationships and leads to other acquaintances that could become targets in future articles.

As stated before, the effectiveness of SNA to identify the key players of criminal organisations has been clearly defined in earlier research.[16] These studies suggest that SNA should be integrated into the academic research and possibly in the investigative phase of police operations. These studies unfortunately only focus on a limited number of metrics, namely centrality and betweenness, and have not been tested in the context of the Internet. The aim of our study is to go beyond this simple recommendation and to evaluate the effectiveness of SNA to enhance the value of information on cybercriminals. To do so also means to analyse SNA's ability to identify possible key players and to discard potential individuals, which could waste researchers' resources. This article analyses the relational profiles of arrested offenders and some of their close associates to determine the structure of their personal network as well as the position of each of the actors. It also discusses the different social network metrics that can be used to better understand the role individual hackers play in the network.

Data and methodology

The data for this research were obtained from a Canadian police force investigation of a network of hackers involved in botnets. Grabosky defines botnets as networks of infected computers that can be remotely and surreptitiously controlled to perform a particular action.[17] Botnets are characterised by a high level of automation that lets hackers simultaneously control thousands or even millions of computers.[18] In order to realise their full criminal potential, botnets need to maintain a certain level of invisibility. In that respect, they are very different from traditional viruses, which quickly attract the attention of their victims either through the destruction of data or through friendly (and often silly) messages alerting them to their weak security practices.[19]

The hackers we studied kept in touch through a messaging technology called Internet Relay Chat (IRC), which allowed them to send private messages to one another in real time. The police managed to recover all the logs of the chats on the hard drives seized from the hackers. Some of these drives were a few years old and contained hundreds of personal

15. Borgatti, Stephen P., Martin G. Everett, and Linton C. Freeman, *Ucinet for Windows: Software for Social Network Analysis* (Cambridge, MA: Analytic Technologies, 2002).
16. Morselli, *Inside Criminal Networks*; Schwartz, Daniel M. and Tony D.A. Rouselle, 'Using Social Network Analysis to Target Criminal Networks', *Trends in Organized Crime* 12 no. 2, (2008): 188–207.
17. Grabosky, Peter, *Electronic Crime* (Upper Saddle River, NJ: Pearson Prentice Hall, 2007).
18. Rajab, Abu, M., Jay Zarfoss, Fabian Monrose, and Andreas Terzis, 'My Botnet is Bigger than Yours (maybe, better than yours): Why Size Estimates Remain Challenging', in *Proceedings of the First Conference on First Workshop on Hot Topics in Understanding Botnets* (ACM: New York, 2007), 1–8; and Stone-Gross, Brett, Marco Cova, Lorenzo Cavallaro, Bob Gilbert, Martin Szydlowski, Richard Kemmerer, Chris Kruegel, and Giovanni Vigna, 'Your Botnet is my Botnet: Analysis of a Botnet Takeover', in *Proceedings of the 16th ACM Conference on Computer and Communication Security*, eds. Somesh Jha, and Angelos D. Keromytis (New York: ACM, 2009), 635–47.
19. Taylor, Robert W., Tory J. Caeti, D. Kall Loper, Eric J. Fritsch, and John Liederbach, *Digital Crime and Digital Terrorism* (Upper Saddle River, NJ: Pearson, 2006).

conversations, whereas others were much newer and included only a limited set of chats. In total, the police managed to recover 4714 one-to-one conversations stored in individual text files. Access to such a trove of information allowed us to overcome the observation bias usually encountered in studies, when subjects often modify their behaviour to please or impress the researcher, and thus obtain a reliable picture of how hackers behave and interact 'in the wild'.

We used the seized data to create a database that contained, for each conversation, the nicknames of the two people talking (including information on who had initiated the conversation), the timestamp and the message itself. Through this process, we identified a total of 771 people, including 10 individuals who were arrested following the investigation.[20] Unlike previous research on the social organisation of hackers, this article does not focus on the individual (sociological or psychological) attributes of hackers in order to understand their group dynamics, but instead analyses the structural properties of their relationships.[21]

SNA requires that researchers work with matrices such as the one shown in Table 1. A 0 represents the absence of relation between two individuals (column and row), whereas a 1 indicates the presence of a relation between the two actors (actor and node are the SNA synonyms for person). Matrices can be read horizontally (whom a person has contacted) or vertically (whom a person was contacted by).

In this network of 771 individuals, some actors are bound to be of lesser importance to researchers. Presenting an in-depth look at 771 social networks is clearly out of the scope of this article. To keep this research as focused as possible on its goal, we divided the population into three distinct groups: *arrested hackers*, *POI* and *others*. The arrested

Table 1. Example of network matrix.

	N1	N2	N3	N4
N1	0	0	1	1
N2	0	0	0	0
N3	1	1	0	1
N4	0	0	1	0

20. It is common for people who use IRC to change their nicknames frequently, making it very difficult for researchers to keep track of all the nicknames used by hackers. We analysed the distribution of nicknames found in the conversations and merged the nicknames that were nearly identical (ex: Poison and Poison). We also identified the nicknames for bots (computer programs) used by hackers to send commands through IRC and removed them from our sample. (These bots were easily recognisable by the string of numbers appended to their nicknames.) The resulting dataset included 771 individuals. In order to protect their privacy, the online nicknames of the individuals will not be used in this article. We assigned each individual a numeric code (N1, N2, N3, etc.) that is used to identify them in the rest of this article. The police organisation that provided us the dataset also added an additional layer of privacy protection for the people involved in this study (including convicted hackers) by refusing to disclose the real identities behind the nicknames.

21. Meyer, Gordon R., 'The Social Organization of the Computer Underground' (Master's thesis in sociology, Northern Illinois University, 1989); Jordan, Tim, and Paul Taylor, 'A Sociology of Hackers', *The Sociological Review* 46, no. 4 (1998): 757–80; Schell, Bernadette H., John L. Dodge, and Steve S. Moutsatsos, *The Hacking of America: Who's Doing It, Why and How* (Westport, CT: Quorum Books, 2002); and Holt, Thomas J., 'Lone Hacks or Group Cracks: Examining the Social Organization of Computer Hackers', in *Crimes of the Internet*, eds. Frank Schmalleger, and Michael Pittaro (Upper Saddle River, NJ: Pearson, 2009), 336–55.

hackers are the 10 individuals who were arrested at the end of the police investigation. The POIs are those who were in contact with two or more of the people arrested. The decision to use two as the threshold, although subjective, seems reasonable if one considers that the arrested hackers lived in different cities and never met in person. Hence, an overlap in friendships might reflect a common interest in hacking. By discounting individuals who did not have ties to at least two arrested hackers, we avoided unnecessarily inclusiveness and restricted our line of inquiry, enabling us to delve into greater details of each person's profile. Although not perfect, this technique was the most effective and reliable, ensuring that if there was a bias, it would be in excluding more hackers than including non-hackers. The last category, labelled 'others', comprises all the people who did not fit in either of the first two groups.

Figure 1 shows the distribution of contacts between the total population and the arrested hackers. A total of 28 people (3.61% of sample) were classified as POIs, whereas 733 people (95.1% of our sample) were placed in the 'others' category and thus not included in our analysis. The final network studied included 38 actors: 10 arrested hackers and 28 POI. The high ratio of contacts that did not appear criminal at first sight is a good indicator of the level of informational noise a researcher must deal with in the course of a complex research that relies on telecommunication intercepts and traffic analysis and demonstrates the importance of developing methods to weed out useless data.

Analysis of this network was done in two stages. First, we looked at the pattern of conversations between the arrested hackers and the POI to determine their position in the network. We focused here on two social network metrics: centrality and power. Centrality is frequently used to assess the prominence of actors within a network.[22] In this study, we used two complementary approaches to measure centrality. The *indegree* and *outdegree* centrality measures indicate, respectively, the number of incoming and outgoing contacts and account for the direction of direct ties around each node. The pattern of ties originating from or sent to a network member is usually a reliable indicator of this person's prestige or status as it helps distinguish people with sought-after expertise.[23]

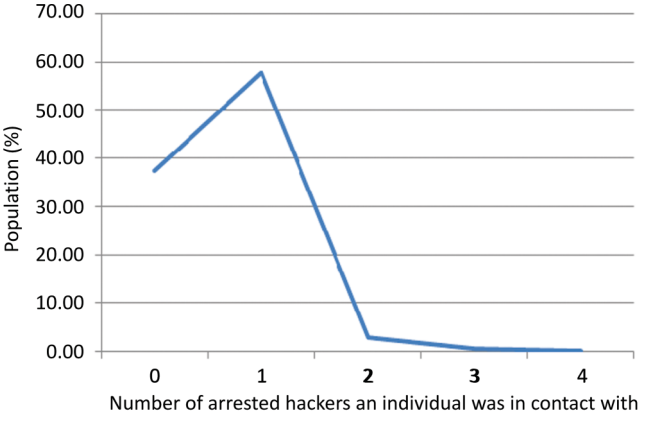

Figure 1. Distribution of the number of contacts with arrested hackers.

22. Wasserman and Faust, *Social Network Analysis*.
23. Ibid.

One of the main criticisms of degree centrality is that it underestimates the importance of actors who have fewer direct ties but are nevertheless influential because of their ability to harness these ties more strategically in order to unite network members.[24] To deal with this, we also used a third measure of centrality, flow betweenness centrality, which measures the extent to which a node is located between other nodes within the network. The more often a node is located between other actors, 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, because these individuals control the flow of information between the different actors.[25] Many alternative metrics have been created to calculate a node's power inside a network. Bonacich power, for example, calculates an individual's power by looking at the number of his or her direct and indirect ties.[26] Borgatti has improved this metric by introducing a PLI measure.[27] Instead of using a fixed constant to determine the effect of direct and indirect ties, Borgatti uses a variable constant that changes from positive to negative at every layer of connections.[28]

As shown in Figure 2, the more connected an individual is, the more power he or she gains (positive effect). The more connections his or her relations have, the more power he or she loses (the connections provide alternative ways of reaching others). If the connections of the connections are well connected, the power of the individual increases again (positive effect) because his or her connections cannot count as much on their own connections. As this approach is more fine-grained and gives a more representative view of the notion of power in social networks, we used it as an additional indication of the distribution of power in the hacker network.

The second stage of our research consisted of a series of models we created using the Keyplayer version 2 software. This software, created by Steve Borgatti, finds the most important player in a network by evaluating the impact on the network of the removal of any node.[29,30] It provides different algorithms to meet different research goals. For the purpose of this study, we used the disruption algorithm, which is designed to determine which nodes should be removed in order to break a network into as many pieces as possible and to increase as much as possible the number of connections required to link nodes in each network fragment. The Keyplayer software outputs a fragmentation delta (coefficient) that indicates, on a scale of 0 to 1, how fragmented a network becomes once *n* nodes are removed. A value of 1 indicates a totally fragmented network where communication between players is barely possible.

24. Morselli, *Inside Criminal Networks*.
25. Ibid.; Prell, C.K. Hubacek, C. Quinn, and M. Reed, 'Who's in the Network? When Stakeholders Influence Data Analysis', *Systemic Practice and Action Research* 21, no. 6 (2008): 443–58; and Toral, S.L., M.R. Martinez-Torres, F. Barrero, and F. Cortes, 'An Empirical Study of the Driving Forces Behind Online Communities', *Internet Research* 19, no. 4 (2009): 378–92.
26. Bonacich, P., 'Power and Centrality: A Family of Measures', *The American Journal of Sociology* 92, no. 5 (1987): 1170–82.
27. Borgatti is the creator of Ucinet, a software package used to study social networks that is very popular among academics. He has recently published an article in *Science* describing the current state of social networks analysis (Borgatti et al., 2009). His work on the political independence measure has yet to be published.
28. Borgatti, Stephen P., *LINKS Workshop on Social Network Analysis* (Lexington: Kentucky, 2010).
29. Ibid.
30. This software is still not feature-complete according to Borgatti (2010). It has thus never been publicized or promoted by its creator. This does not affect the core functions of the software used for this article, which is available free of charge from Analytic Technologies' website (www.analytictech.com).

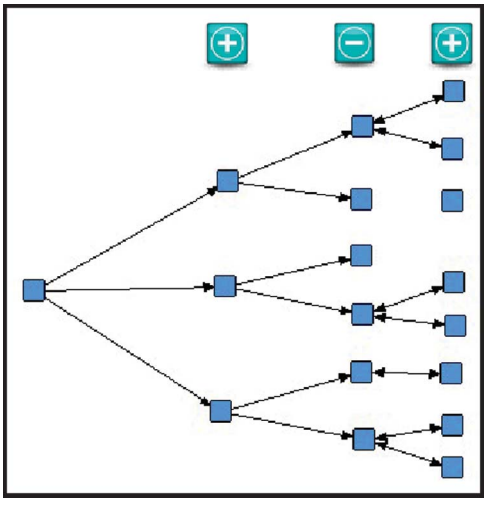

Figure 2. Demonstration of the political independence measure.

It is important to recognise that this methodology is not without limitations. First, we had access only to conversations that the hackers had among themselves or with the POI. We are thus missing the conversations that the POIs had among themselves. This limits the size and representativeness of our sample, a factor that must be taken into consideration when reading the results. Second, hackers who had installed new hard drives shortly before their arrest or were smart enough to erase their conversation logs would appear to be less-important players than those who had used the same unsecured hardware for the past few years. Although the footprint of these careful hackers would presumably be smaller, we believe that our approach can still provide important information. For instance, these ultra-cautious hackers are also likely to be the most competent, and therefore, the most sought after by their peers, even during the limited periods of time for which we have data. The fact that some of our metrics use ratios rather than raw numbers also mitigates the effect of this data deficit. Finally, we focused on the private messages these hackers exchanged, although we recognise that they also interacted through other channels, such as public chat rooms dedicated to hacking topics. This means that only a fraction of these people's online interactions are analysed in this study.

The representation of the network presented in this article is of the network at a certain point in time. It would be interesting to predict how the network would heal once some of its members were arrested, but, as this is impossible, discussion is limited to the immediate consequence of the removal of actors in the network. Although the data are incomplete and far from perfect, we believe that they yield some very insightful results.

Who is in, who is out: the fuzzy boundaries of a hacker network

The nail that sticks out gets hammered down

A graphic representation can provide useful structural insights into a network by allowing the display and visual analysis of patterns related to the distribution of ties, the grouping of actors and the positioning of particular individuals.[31]

31. Freeman, Linton C., 'Visualizing Social Networks', *Journal of Social Structure* 1, no. 1 (2000), http://www.cmu.edu/joss/content/articles/volume1/Freeman.html (accessed December 21, 2010).

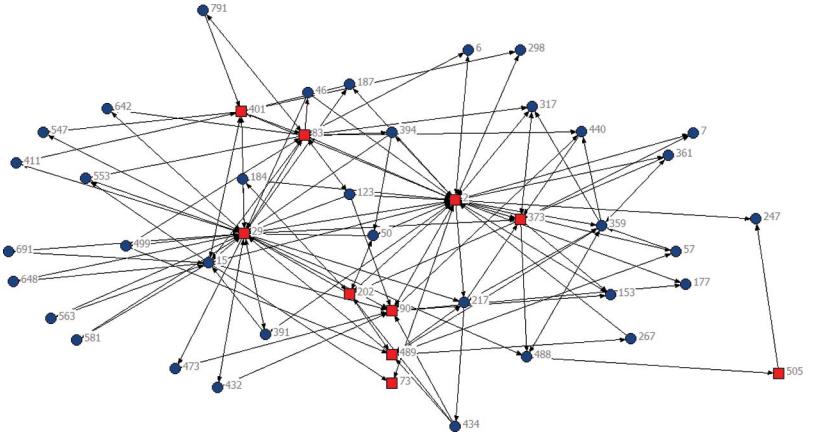

Figure 3. Graphic representation of the network.

In Figure 3, the squares represent the arrested hackers and the circles represent POI. A few nodes stand out in this graphic. N505 (bottom right) looks isolated from the network and even more so from the other arrested hackers. Although some arrested hackers seem to have many contacts (N2 and N29), others have very few ties (N73 and N505). Moreover, some POI – such as N217 – seem to have a more central position in the network than some of the arrested hackers. This suggests that some important players might have been ignored while some 'fringe' people were arrested.

Table 2 shows the patterns of conversation of the arrested hackers. The average number of conversations is relatively low, especially considering that the earliest conversation on record occurred on 13 July 2004, and the last exchange could have been conducted minutes before the arrests and equipment seizures by police investigators in February 2008. Some arrested hackers appear to have very seldom spoken with each other (N73 and N505). This could be due to the nature of the data (see the methodology section) or to the fact that these people very rarely communicated with each other through private messages. It was, unfortunately, not possible for us to confirm either of these hypotheses, which is why we used percentages instead of raw numbers. On an average, 18.89% of the arrested hackers'

Table 2. Patterns of conversations of arrested hackers.

	No. of conversations	% of conversations with other arrested hackers	% of conversations with POI
N2	1242	6.60	11.03
N29	599	15.69	13.86
N73	6	66.67	0.00
N83	255	18.43	10.98
N90	567	1.94	6.17
N202	67	17.91	8.96
N373	86	34.88	18.60
N401	88	25.00	13.64
N489	110	1.82	43.64
N505	6	0.00	33.33
Average	**303**	**18.89**	**16.02**

conversations were directed at other known hackers. This indicator shows great variability, with a minimum of 0.00% and a maximum of 66.67%. It appears that the arrested hackers can be divided into three groups based on connectivity: the connected (N73, N373 and N401), the somewhat connected (N83, N202 and N29) and the disconnected (N2, N90, N489 and N505). Only the first group shows a strong involvement with others hackers, directing a significant percentage of their communications to these individuals. At first glance, this network does not show a high level of cohesion and connectivity between its members. The total number of conversations is loosely and inversely correlated to the number of conversations with other hackers ($r = -0.373$). This would indicate that the arrested hackers with the largest social activity were less involved in the general hacking group.

Overall, the arrested hackers were more involved with each other than with the POI (average of 18.89% of conversations vs. 16.02%). However, almost half of the arrested hackers (N2, N90, N489 and N505) had more interactions with POIs than the fellow hackers. This result raises the question of whether the arrested hackers belonged to a tightly coupled network or whether some part of this network was overlooked by police investigators. Based on Table 2, the differences between the arrested hackers and the POIs are minimal. It would be difficult to determine who was arrested and who was not by looking only at the results in Table 2.

Another way to look at the network of hackers is to look at the flow of conversations (Table 3).

Both groups have a high tendency to contact others (55.84% for hackers and 70.42% for the limited sample of POI). This would indicate that they are both actively seeking each other – an even more prominent trait among the POI. Even with a very high level of conversations initiated (which could easily be irritating to others), the POIs still manage to elicit more answers from others. This suggests once again that these actors may not be peripheral players.

Moving forward, the next tables and figures use SNA indicators to enhance our understanding of this botnet operators' personal network.

Table 4 displays the degree centrality of the actors involved in this network. The normalised coefficients are provided for comparison purposes. Of the top 10 highest ranked actors on this factor, eight were arrested. Hackers seem to be more connected to other players (from both the categories) than the POIs. Two arrested hackers (N505 and N73) scored poorly, with metrics similar to those of POI. Inversely, N359 is the fourth most-connected actor in this network and was not arrested. Generally speaking, it seems that the police unit targeted the most centralised people in the network, with a few exceptions. However, although centrality is a good indicator of how ties are distributed in a network, it does not necessarily identify the people who control the flow of information. These brokers, whose influence derives from their capacity to connect people without attracting too much attention, can be better identified using the flow betweenness centrality measure shown in Table 5.

The top seven spots are occupied by arrested hackers, meaning that the people who control the flow of information were targeted during this police operation. The flow

Table 3. Flow of the conversations.

	% of conversations initiated by . . .	% of conversations w/no answer
Arrested hackers	55.84	24.12
POI	70.42	10.83

Table 4. Centrality of nodes.

	Outdegree	Indegree	nOutdegree	nIndegree	No. of spots in TOP 10
Arrested					
Minimum	1.0	1.0	2.703	2.703	8
Maximum	19.0	22.0	51.351	59.459	
Average	6.5	9.1	17.568	24.595	
POI					
Minimum	2.0	0.0	5.405	0.000	2
Maximum	7.0	6.0	18.919	16.216	
Average	2.6	1.7	7.143	4.633	

Table 5. Betweenness centrality of nodes.

	Flow betweenness	nFlow betweenness	No. of spots in TOP 10
Arrested			
Minimum	0.0	0.0	7
Maximum	451.4	33.9	
Average	121.5	9.1	
POI			
Minimum	0.0	0.0	3
Maximum	68.6	5.2	
Average	7.5	0.6	

Table 6. Political liberty independence of nodes.

	PLI measure	No. of spots in TOP 10
Arrested		
Minimum	−29.5	1
Maximum	25.4	
Average	−12.0	
POI		
Minimum	−28.4	9
Maximum	19.7	
Average	4.1	

betweenness centrality is concentrated, with N2, N29 and N83 showing much higher numbers than the rest of the players. POIs do make it into the top 10 but only in the last three spots. Two of the arrested hackers have very low scores (N505 and N73). Two actors benefit more from their indirect than direct links when compared with others (N488 and N489). Conversely, two nodes are very poor at the brokerage game (N359 and N373) compared with other players. Each of these duos is composed of an arrested hacker and a POI.

Although the control of information and the idea of power are often associated, this connection is not always found. Therefore, other measures of power were developed. Table 6 presents the PLI scores for the actors.

The POIs clearly have a dominating position when it comes to power: 19 of the top 20 spots belong to them while arrested hackers hold half of the last 18 places. This indicates that even though the hackers are generally well connected and good brokers, they are

not positioned so efficiently in the network as to be indispensable. The actors they are tied to usually have other alternatives to get the information they need. This is not the case for the POI and could explain why the POIs had such a high response rate to their communications while the hackers did not. The distribution of power in this network is thus not as concentrated as the first set of results seemed to indicate.[32]

The arrested hackers who were very involved in conversations with other hackers (N73, N373 and N401) rated higher than the others in the PLI measure. Those who were somewhat involved with other hackers scored the highest in betweenness and degree centrality. This would indicate that what the police investigators found would be better described as an association of hackers rather than a hacking organisation. Those whose ties were concentrated within the association had limited sources of information and thus less power in the network.

Network fragmentation

To model the disruption of a criminal network, many models have been proposed. Because it is seldom possible to remove every member of an organisation, the alternative is to break criminal networks in as many small pieces, or subsets, as possible and to make sure that the (metaphorical) distance between the players in each of these pieces is as great as possible. This metric is especially useful to measure the impact of future arrests on a criminal network. Table 7 presents the fragmentation delta when *n* nodes are removed.

All but two (N359 and N488) of the actors whose removal would optimally disrupt the network were arrested after the police investigation. This indicates that the police managed to target the individuals whose removal would have the most impact on the cohesion of the network. The nodes identified by the Keyplayer software as those that should be removed in order to maximise the network disruption are stable from scenario to scenario, indicating that there is no doubt as to which nodes should be removed to increase the fragmentation delta.[33] This delta is highly correlated with the betweenness scores ($r = 0.860$) and somewhat correlated with the centrality scores ($r = 0.527$), confirming past research that argues that police looking to disrupt criminal networks should focus their efforts on the brokers.[34]

Table 7. Fragmentation delta with *n* nodes removed.

No. of nodes removed	Nodes removed	Fragmentation delta
2	2, 29	0.297
3	2, 29, 83	0.585
4	2, 29, 83, 489	0.692
5	2, 29, 83, 489, 373	0.770
6	2, 29, 83, 489, 373, 359	0.816
7	2, 29, 83, 489, 373, 359, 90	0.829
8	2, 29, 83, 489, 373, 359, 90, 202	0.840
9	2, 29, 83, 489, 373, 359, 90, 202, 401	0.846
10	2, 29, 83, 489, 373, 359, 90, 202, 401, 488 or 505	0.848

32. The power in this network is held to stem from the ability to provide information and easy access to information.
33. Borgatti, 'Identifying Sets of Key Players in a Social Network'.
34. Morselli, *Inside Criminal Networks*.

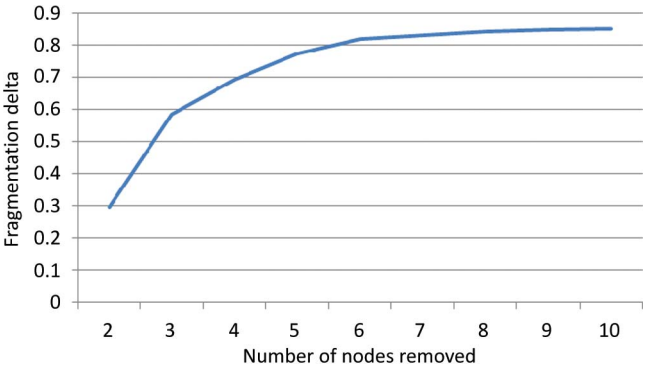

Figure 4. Fragmentation delta of the network.

Even more information on this metric can be extracted from the fragmentation delta when we take a look at the growth rate of the measure.

The curve in Figure 4 displays the fragmentation delta when *n* nodes are removed. After the sixth node is removed, the rate of growth of the fragmentation delta slows down greatly and flattens after about 10 nodes. The act of removing (arresting) a node has, at that point, almost no effect on the fragmentation of the network. Figure 4 highlights the inherent tension in criminal investigations conducted by a law enforcement organisation with limited resources: should such investigations be guided by an unyielding justice approach (go after all known criminals, without discrimination and at any cost) or switch to a utilitarian calculative approach (choose investigation targets based on their importance in a criminal network)? This question and a few others will be addressed in the following discussion section.

How can SNA contribute to research on hacker networks?

In this article, we used SNA to illustrate better ways of understanding the structure of an online criminal network. The aim of this research was to evaluate the effectiveness of SNA to enhance the value of information on cybercriminals. This information could and should be used to identify possible targets for follow-up research as well as to discard the marginal targets that waste researchers' resources.

The communication patterns show that the arrested hackers were loosely associated with one another and that their social network extended well beyond their hacking associates. Our methodology based on common relationships between a network of arrested hackers needs to be taken into account when evaluating the results and the discussion that follows. Still, the identification of 28 individuals who could be involved in the hacking world through their ties to the arrested hackers offer new insights for future research.

The investigators who led the investigation were experienced police officers who knew the hacking scene intimately and who were able to pinpoint the key members of this network, as shown by the finding that the arrested hackers scored very high on both the degree centrality and flow betweenness centrality measures. With regard to power, the arrested hackers did not fare as well, scoring only one of the top 10 spots in the PLI metrics. It is obvious that the most visible individuals were targeted in this investigation and that such a

position is associated with higher risks of arrests.[35] The POI were by no means the most central players in the network. They did, however, occupy positions of power, meaning that their strategic location gave them easier access to resources. They were also able to generate more interest from others as their messages engendered more answers than those from the arrested hackers.

These results demonstrate that individuals who looked peripheral at first sight but could arguably have played a significant role may have been overlooked. Two POIs stood out in our sample: N57 and N440. These two actors were among the top 10 POIs for each of the SNA metrics we examined in this article. Although it is not unusual for nodes to have high betweenness as well as a high degree of centrality, high measures of power and centrality tend to indicate the importance of actors in a given network. These two people would be ideal targets for future research because of their position in the network. It would also be interesting to take another look at N177 and N434, who also scored high measures of power and betweenness centrality (but not degree centrality). The fact that we were able to identify these nodes shows how SNA methodologies can help identify suspicious actors from among a large sample of individuals. Inversely, N505 and N73 appeared as marginal players based on their IRC private messages. They both ranked poorly in centrality and even worse in betweenness with a normalised score of 0. N73 was the best positioned according to the PLI measure but was not listed as one of the 10 nodes that should be removed according to the Keyplayer software.[36] N505 was selected by the software but was on par for the 10th and last place with N488, a POI.

The second part of our analysis brings another dimension to this article. As we mentioned, investigators did a good job of identifying the most visible players in the network. They also managed to arrest the people whose removal from the network was the most disruptive. The Keyplayer software clearly indicates that the criminal network was seriously damaged following the operation, with a fragmentation delta close to 0.900.[37] This achievement is, however, offset by the fact that the arrests would have had a bigger impact had a POI (N359) been arrested and that the fragmentation delta remains almost flat once six nodes are removed from the network. This leads us to raise the issue of what rationale should be used by investigators: an ideal of justice that can be achieved only by prosecuting all known offenders or the more pragmatic utilitarian calculative approach paradigm that is more concerned with the effect of outcomes than with principles.

We believe that, in the case of property crimes such as cybercrimes, a calculation-based approach to the problem would yield interesting results in terms of efficiency. The arrested hackers were botmasters; they controlled thousands if not tens of thousands of computers in order to earn money by renting their herd of computers to the highest bidder. At no point was the safety of individuals put at risk by these hackers. In a case such as this, following every lead might have involved more resources than was strictly necessary. Very often, only a small core of those in a hacking group has the technical skills needed to build botnets. The rest of the group knows barely enough to use the software created by others and often need to get numerous pointers from the skilled hackers in order to overcome technical hurdles. Removing the core group from the hacking community could become an alternative strategy in the police investigators' toolbox. Complex inquiries, such as the one studied in this article, often involve hundreds of people whose activities need

35. Ibid.
36. Borgatti, 'Identifying Sets of Key Players in a Social Network'.
37. Ibid.

to be monitored and thousands of others whose privacy must be protected. Even the most experienced investigators are constrained by what psychologists call the 'channel capacity' (or cognitive load), which is classically defined as the upper limit on the extent to which a human being can match his cognitive processes to external stimuli.[38] SNA tools can lighten this cognitive load by processing large amounts of data and producing the results in a user-friendly format. Of course, SNA methods are only useful in a particular field (such as online crime investigations) if they are supported by a relevant set of concepts and theories that provide useful interpretations of the various metrics produced.[39] More bluntly, it would be very easy, in the absence of a readily accessible body of research focusing on the structural and relational features of malicious hacker communities, for the SNA tail to wag the investigation dog. Knowledge of how hacker communities function is essential to determine more precisely how skills, trust and power accrue and are transferred among hackers, and to allow more accurate interpretation of the metrics obtained from public and private data streams. In this specific case, given the fragmentation delta of the network once the arrested hackers are removed, suggests that the threat posed by this group of individuals would be vastly reduced. Police investigators should monitor the best players of this network to detect how organic ties grow back after the arrests. Just as with hydras, once the head of a criminal network is removed, it does not take long for a new one to grow. It would be of much interest to be able to study a criminal network after multiple waves of arrests – something that has not been published so far.

Conclusion

This article has shown that SNA could be useful on two levels. First, SNA provides scientific and objective measures of the structure of networks as well as the position of their key players. As mentioned before, the problem in modern research is not finding data but finding a way to make sense of it. Using the metrics and methodology detailed in this article, researchers can confirm their suspicions and gain early insights into the constantly evolving shape of criminal organisations. Second, fragmentation metrics help to determine the amount of resources needed to deal with specific organisations. This article has shown how a tactical strike against this botmaster network could have had the same destabilising impact as a broader approach. The resources saved by limiting the investigation targets could then be used to monitor the criminal network's reaction to the arrests and to limit its ability to adapt to the post-arrest environment.

Integrating SNA software can lead to widespread abuse if basic ethical and legal guidelines are not followed. SNA raises specific ethical dilemmas that must be acknowledged in order to avoid making careless decisions. In particular, missing data can produce significant distortions that can be extremely misleading. In cases, for example, where the missing information concerns a very central actor or a key broker, minor or secondary actors might end up with a much higher than warranted 'criminal' score, leading to unnecessarily extensive scrutiny. To mitigate this problem, researchers must make use of as many data sources as possible: emails, online chat rooms, telephone calls and meetings. The quality of this type of analysis depends on the quality and quantity of relational data available. Incomplete datasets will create partial networks, which are less reliable.

38. Miller, George A., 'The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information', *The Psychological Review* 63, no. 2 (1956): 81–97.
39. Scott, John, 'Social Network Analysis', *Sociology* 22, no. 1 (1988): 109–27.

It is especially important to remember that human beings are multidimensional and have a very broad range of interests and connections beyond their main income-generating activity. A malicious hacker is also a friend, a family member or a customer involved with many individuals who play no part in their contact's criminal enterprises. As a growing number of our social interactions occur online, we need to set-up safeguards to ascertain which contacts are suspicious and which are trivial. Being unable to make this distinction will only perpetuate the 'guilt by association' syndrome that generates unacceptably high levels of false positives. The goal here is not to extract a large quantity of useless leads from the data, implicating innocent people in the process, but rather to improve the analytical capabilities of researchers. It is critical that SNA not be presented as a 'silver bullet' for researchers but rather as a promising research method that needs to be applied thoughtfully and in conjunction with more traditional practices in order to realise its potential.

Notes on contributors

David Décary-Hétu is a PhD candidate at the School of Criminology of the University of Montreal. He specialises in cybercrime issues such as intellectual property theft, botnets and carding. He has received the Daniel Éli prize of his School of Criminology and has presented in conferences in Australia, Canada, Japan, the United States and Europe.

Benoit Dupont is an associate professor at the School of Criminology of the University of Montreal. He is the director of the International Centre for Comparative Criminology and the head of the Canadian Chair of Research on Security, Identity and Technology. He specializes in cybercrime issues such as the social network of hackers and botnets.