From data to disruption

P.A.C. Duijn, P.M.A. Sloot

Research and Analysis, Dutch Police, The Netherlands
University of Amsterdam, The Netherlands
Complexity Institute, NTU, Singapore
ITMO University, St. Petersburg, Russian Federation

Organized crime groups and law enforcement agencies are caught in complex systems similar to a continuous game of cat-and-mouse, in which the latter frequently remains two or more steps behind. Law enforcement agencies are therefore seeking for more proactive strategies in targeting these criminal network structures more effectively. This starts with a better understanding of the way they operate and adapt over time. A key element to developing this understanding remained largely unexploited: big data and big data analytics. This provides novel insight into how criminal cooperations on a micro- and meso-level are embedded in small-world criminal macro-networks and how this fosters its resilience against disruption. This paper discusses the opportunities and the limitations of this data-driven approach and its implications for both law enforcement practice and scientific research.

© 2015 Elsevier Ltd. All rights reserved.

Introduction

Organized crime groups impose a continuous threat to global society, by causing harm to our economic, social, technological, political and environmental infrastructures (Europol, 2013, 2015). Their existence depends on maximizing efficiency and profit from their illegal activities, while remaining undetected by the government at the same time (Raab and Milward, 2003, Erickson, 1981, Morselli et al., 2006, Duijn et al., 2014). Law enforcement agencies on the other hand are struggling with important questions: How can we detect these criminal groups and their activities? What are the best strategies to disrupt them effectively? And how do they develop resilience against interventions? Within the law enforcement organization a key element to answering these questions has remained largely unexploited: big data and big data analytics. Since data are becoming more and more available from a plethora of new sources, they will provide opportunities for data-driven analysis towards understanding organized crime in terms of criminal network structures, dynamics, and resilience against law enforcement interventions (Duijn and Klerks, 2014a, 2014b). This paper discusses the opportunities and the limitations of this data-driven approach and its implications for both law enforcement practice and scientific research.

Understanding organized crime

Theories about organized crime have changed over time. Early perceptions of organized crime focused on hierarchical pyramid structures with kingpin leaders controlling their criminal enterprise from the top. Criminal organizations were perceived and analyzed as separate entities on a micro-level, leading to an oversimplified perspective of criminal reality (Duijn and Klerks, 2014a). In the early nineties organized crime received serious scientific attention for the first time, which led to empirical studies of organized crime. A selection of court files and
case studies from multiple criminal investigations were analyzed manually. These studies uncovered mechanisms of trust, expertise, reputation and social opportunity structures, which shape the way organized crime groups and individual actors become connected and adapted to fast changing illegal markets (Fijnaut et al., 1991; Kleemans and Van de Bunt, 1999, Klerks, 2001). It was also revealed that organized crime is an integral part of the global networked society and it was emphasized to study organized crime from a network perspective (Klerks, 2001; Kleemans and De Poot, 2008).

Due to practical limitations of manual analysis techniques, a macro-level understanding of the structure of organized crime currently consists of theoretical assumptions instead of empirical observations. On the other hand, datasets about terrorist- and criminal actors and their mutual connections are growing, due to more deliberate strategies for collecting, storing and sharing information within day-to-day law enforcement practice since 2001 (The terrorist attacks). Moreover, the influx of embedded academics within law enforcement has made these datasets more easily accessible for scientific purposes (Duijn and Klerks, 2014a). At the same time scientific disciplines such as social science and complexity science have started to exchange ideas and methodologies, leading to advanced network analysis methods being introduced into social science. This opens the door towards a data-driven approach to create an empirical understanding of organized crime.

Complex adaptive systems

From a macro-perspective criminal network structures can best be understood as complex adaptive systems. The concept of complex adaptive systems (CAS) derived from systems theory and was first introduced by Holland (1999). A complex adaptive system is a self-organizing network, which constantly adapts its structure and behavior according to change in the behavior of its individual components (agents). These agents constantly act and react to each others behaviors and the environment, meaning nothing is fixed. This makes the behavior and structure of CAS highly unpredictable, but effective in adapting to changed environments (Chan, 2001).

While CAS is an established theoretical concept for understanding complex networks in biology and economy, CAS theory also applies to criminal networks that constantly adapt to changing law enforcement strategies and government regulations (Kenney, 2007). Criminal network structures continuously balance between efficiency in the collaboration of its parts and security in staying undetected by law enforcement organizations. Shifts in law enforcement strategies may destabilize this balance and trigger shifts in the way the independent criminals in the network interact and adapt. How this affects the structure of the overall criminal network depends on how the independent actors interact with each other to adapt to these external factors from the bottom up. There are no explicit rules about how a criminal network is formed or changed. Criminal networks are emergent self-organizing systems, which changes structure at any point in time due to how its parts react to external pressures.

Criminal networks have for instance quickly adapted to the opportunities created by the Darknet that provides anonymity and access to worldwide online marketplaces for selling illegal commodities in large quantities. Law enforcement agencies responded by successfully taking down some of the most active online marketplaces (Soska and Christin, 2015). The criminal cyber networks active on these online marketplaces adapted to these interventions by increasing the number of online marketplaces and servers, outweighing the limited capacity of law enforcement to target them all effectively. The interactions of the individual actors changed the shape of the Darknet towards a more dispersed network structure. Such a continuous evolution driven by non-linear feedback mechanisms is an important feature of complex adaptive systems. The practical reality is that law enforcement will always be one or more steps behind (Soska and Christin, 2015).

To narrow this gap, researchers should focus more on capturing criminal network dynamics instead of focusing on static network representations. Understanding these dynamics can have implications for uncovering mechanisms of competitive adaptation, criminal network resilience and the effectiveness of law enforcement interventions (Duijn et al., 2014; Duijn and Klerks, 2014a). Social network analysis and computational modeling can help to uncover these dynamics, but before we can understand the output of these methodologies we need to obtain a better understanding of the sources: the data.

Law enforcement data

Criminal networks actively try to avoid detection from law enforcement. As compared with legitimate social systems, they are particularly hard to detect leading to inevitable missing data in the final network representation. The completeness of a criminal network representation is therefore highly dependent on the strengths and weaknesses of the data sources from which it is obtained. Many criminal network studies necessarily rely directly or indirectly on law enforcement data, which is not primarily collected for scientific purposes (Morselli, 2009). Two important factors should therefore be taken into account to retrieve a reliable and valid network representation from these data.

First, the accuracy of the data source is a critical consideration (Morselli, 2009). Every piece of law enforcement data is collected in the context of a specific policing task, for instance collecting evidence, preparing investigations or monitoring a situation which shapes the information collection process and the bias embedded within. To overcome these biases, researchers need to understand the background of the data collection and take into account the policing priorities while drawing conclusions (Duijn and Klerks, 2014a).

1 The terrorist attacks in New York, London, and Madrid prompted the introduction of intelligence-led policing within many national police departments. It involves a process for which every decision within law enforcement should be preceded by a structural analysis of the situation based on deliberately collected information in the frontline of police practice (see Ratcliffe, 2012).
Before criminal networks are monitored and targeted (investigative phase), law enforcement tries to detect their structures and activities (intelligence phase). Fig. 1 demonstrates how the different phases of the intelligence and investigative processes in law enforcement relates to the accuracy of the final network representation. It is obvious that a more accurate and confirmed network representation can be retrieved from actors that are found guilty based on evidence, rather than about actors that are monitored but not targeted in the intelligence collection phase preceding a criminal investigation. Many researchers therefore exclude intelligence data from their data collection process. So why even consider using intelligence data in the first place?

The answer to this question lies in the second factor: the scope. The scope of the network representation can be limited by investigative priorities (Morselli, 2009). If the aim of an investigation is to target three main suspects, data collection will be focused on their activities including their first line contacts. Second line contacts, potentially important for connecting the micro-level network with the embedding meso- and macro criminal network structures are often not included. Moreover, evidence will be structured in the final court file in a way that fits into the legal framework of a criminal organization, strongly suggesting a certain (artificial) network structure.

Intelligence data on the other hand is collected with the aim of improving the overall intelligence picture and preparing criminal investigations, without setting preliminary targeting specific suspects. In many jurisdictions across the globe, data from different sources can be more easily combined in the intelligence phase as compared to the investigative phase. In Dutch law for instance, investigative data can be combined with street cop data and data from criminal informants (human intelligence) to complete the intelligence picture. In certain cases criminal informants can even be directed or asked specifically in order to fill in intelligence gaps about the missing dots in criminal network representations (Duijn and Klerks, 2014a, 2014b). Figs. 2 and 3 demonstrate how the scope of the data-collection can provide a varied and even misleading perspective of the structure of a criminal network, in this case the a criminal micro-network involved in cannabis cultivation (the Blackbird Network).

In the trade-off between accuracy and scope the best solution is to merge as many data sources as possible into one relational database in which data are merged based on unique identifiers (e.g. Names, ID-numbers). For instance, arrest records could be combined with unconfirmed human intelligence data, to check the accuracy and reliability of relationships found in each database (Duijn and Klerks, 2014a, 2014b). However, there will always be missing data. A further step towards creating accurate network representations would be to add weight to relationships based on the reliability of the underlying sources or the duration and frequency by which a connection is observed in the data (Schwartz and Rouselle, 2009, Perer and Schneiderman, 2009).

Open source data

Important pieces of data about criminal cooperation can also be found outside of the law enforcement context. Social media provides unique observations of the social context embedding criminal networks outside the law enforcement context. However, a lot of information posted in social media remains unconfirmed or simply doesn’t resemble real-life social relations. By using it in addition to law enforcement data, it can fill in holes within the final network representation. In some cases social media data can also be mined for text indicating the presence of a criminal network structure. Dijkstra et al. (2013) identified a drug-users network within a large social media forum located in the Russian Federation by using web crawler techniques and text mining developed around specific drug-use vocabulary (Dijkstra et al., 2013). Similar
techniques could be utilized for mining communications within online marketplaces on the darknet, which inhibits a unique observation of dark network structures outside the scope of law enforcement (Décary-Hétu et al., 2014).

**Fig. 2.** A graph showing the Blackbird network gravitating around the main suspects based on data collected during three targeted criminal investigations. It provides a micro-level perspective of the network structure (Duijn and Klerks, 2014a, 2014b).

**Fig. 3.** A graph of the actors from the Blackbird network highlighted in red within the embedding criminal network structure, based on data collection in the intelligence phase. It shows that the Blackbird network is connected to other criminal micro-networks within a larger embedding meso/macro network structure (Duijn et al., 2014).

**(Big) data analysis**

If the final network representation is constructed out of different data sources combined, it can easily contain thousands of entities and connections (see Fig. 3). Manually analyzing these data would require a long-term commitment. Three methods can be utilized together to analyze such data in a way that combines automated analysis together with a manual assessment.
Social network analysis

Social network analysis (SNA) consists of a combination of network theory and mathematical techniques to uncover patterns hidden in social networks consisting of actors and their mutual connections. Raw network data is processed in matrices in a binary format that subsequently can be visualized in a sociogram (e.g. Fig. 2) (Sparrow, 1991, Van der Hulst, 2009). To analyze the data, different algorithms can be calculated uncovering network features on a micro-level (e.g. centrality, key players, and brokerage roles), meso-level (e.g. cliques, k-cores) and the macro-level (e.g. density, geodesic distance). By combining these different levels of analysis with visualizations and adding qualitative attribute data to the entities, hidden patterns within criminal networks can be uncovered.

Fig. 4 shows for instance the central positions of females within the Blackbird network. SNA contributed to the understanding that these women were important for consolidating the criminal activities (cannabis cultivation, synthetic drugs production) after the two central actors where arrested (Duijn and Klerks, 2014a, 2014b). These females came into the network via romances with core members and became part of the core members of the criminal network themselves, even after these romances ended. This case example demonstrates how SNA can contribute in uncovering relationships such as the role of females in criminal networks and network resilience.

The utilization of SNA for understanding criminal network structures is expanding fast in law enforcement and for scientific applications (Duijn and Klerks, 2014a, 2014b). Recent studies have successfully applied SNA to shed new light on drug trafficking (Natarajan, 2006, Malm et al., 2010), criminal network resilience (Bouchard, 2007), computer hacking networks (Décary-Hétu and Dupont, 2012), meetings of mafia bosses (Calderoni, 2014), child exploitation networks (Joffres et al., 2011), and terrorist networks (Davies et al., 2015).

Scripting

Scripting is another method of uncovering the complexity within criminal network structures. Within a criminal network a chain of events needs to be executed in order to commit a criminal offence. Although the complexity of these chains varies across different crime types, every criminal conspiracy needs a division of tasks. In a criminal network the responsibilities following from these tasks are often divided between the actors according to experience, skills or knowledge. By mapping the roles of these actors within the network structure different crime-scripts of the criminal activities can be identified (Fig. 5) (Cornish, 1994, Bruinsma and Bernasco, 2004). A combination between SNA and scripting is useful for understanding the deeper operational structures within criminal networks and the interdependences between actors within their illegal activities. More operational applications of scripting help to identify actors who are more difficult to replace because of the skills and expertise they bring into the crime script, in combination with their network position (Duijn and Klerks, 2014a, 2014b). Scripting is mainly used as an attribute to perform more in-depth manual analysis on a micro-level. However, practitioners have developed metrics that combine the crime-script data and network data for more automated network analysis on a meso- and macro level as well (Toth et al., 2013).

Of course the reliability and validity of the output of these matrices relies strongly on the quality of the data.

Fig. 4. Sociogram of the Blackbird network. The triangles represent females in the network while the circles represent males.
Computational modeling

Although SNA and scripting can be used to unravel mechanisms within criminal networks that contribute to resilience, adaptation and operational activity, they are mainly based on static observations of dynamic phenomena. Some scholars have used SNA to capture the dynamics of criminal networks by monitoring change within different snapshots in time, for instance before and after law enforcement or military interventions (Morselli and Petit, 2007, Bush and Bichler, 2015). This provides interesting perspectives on network resilience and the effects of interventions on a micro-level. However, to understand the effects of interventions on a meso- and macro level more advanced methods are needed. Computational modeling is a method to simulate complex network behavior with the help of mathematical models that are used as input for computer simulations. The parameters of these models are developed from previous research or by collecting knowledge from field experts. Real-life data are used as input for the computer simulations to experiment with different scenarios in a virtual environment.

Duijn et al. (2014) used this method to unravel the effects of different law enforcement intervention strategies, while also taking into account network recovery (Duijn et al., 2014). Data were collected over a five year period and contained a combination of human intelligence data, criminal investigations data, street cop data, social media data, and arrest record data. All actors in this criminal network representation were manually labeled with the roles they performed within the different crime scripts. The aim of the simulations was to analyze how a specific part of the criminal network involved in cannabis cultivation could be disrupted, taken into account that recovery could also involve actors from the embedding network. Mathematical models were created for four different intervention scenarios (degree-, betweenness-, crimescript degree- and specific role attack) and three different replacement scenarios (shortest distance, most visible and random). By simulating interventions and replacement at the same time, results showed that network disruption made its internal flow of information even more efficient. However, this also caused its network structure to ‘light-up’ since efficient communication exposed the key players that tried to limit their exposure. These simulations contributed to a better understanding of the effects of law enforcement interventions and how disrupting a criminal network takes a long term effort. Bright and Delaney (2013) used a similar technique to simulate the disruption of a methamphetamine production network. They found that a shift of intervention strategies at a certain threshold could be effective in disrupting the criminal network more effectively (Bright and Delaney, 2013).

Although these methods bring us closer to understanding the complex criminal reality, it is important to realize this is a form of applied science. Such methods help us create effective scenarios for law enforcement practice, but do not provide a target list for the upcoming months. It does point law enforcement agencies in the right direction however. Moreover, criminal reality can be oversimplified by using these methods. For instance, the removal of criminals from a criminal network in practice does not necessarily mean that they are incapable of controlling the criminal activities. Examination of Dutch police intelligence on 13 cases revealed that there are numerous possibilities for criminal leaders to manage their criminal network structures from prison (van der Laan, 2012). Such factors cannot be easily included in computational models.

Discussion

The aim of this paper is to provide a short overview of the current developments in the field of big data analysis in the study of organized crime. It shows that for studying organized crime it is important to seek rather than assume structure (Morselli, 2009). The different examples in this paper demonstrate that criminal networks should be understood as complex adaptive systems, which are unpredictable in their structures, behaviors, and criminal activities. By combining data from different law enforcement sources, the different limitations inherent to each individual network data-source can partially be dissolved. Although there will always be missing data. An understanding of the law enforcement environment, the criminal environment, as well as complexity science is therefore needed to draw meaningful conclusions about criminal network structures, -resilience, and -dynamics, without
losing sight of the limitations of the data or the methodologies on which they are based.

Another open issue remains the fundamental question on the controllability of complex networks. Ashby's law of requisite variety states that a controller must have at least as much variety (complexity) as the controlled (Ashby, 1958). Within the bureaucratically organized law enforcement environment, variety in strategies and different kinds of expertise to disrupt criminal networks remains limited. Moreover, recent research indicates that novel information metrics need to be developed to understand the (side-) effects of manipulating nodes in complex networks (Quax et al., 2013a, 2013b). More research is therefore needed for a deeper understanding in the effects of different disruption strategies.

We conclude that law enforcement agencies and researchers in this field should focus more on monitoring and uncovering the mechanisms of criminal network dynamics, instead of aiming at static observations of criminal reality. This is not an easy endeavor, however a deeper integration of different scientific disciplines could bring together the proper knowledge and tools to uncover this dynamic complexity. We conjecture that a symbiosis between scientific and law enforcement embedded academicians could open doors to the exponentially growing amount of empirical data, which is needed to develop a common understanding (Duijn and Klerks, 2014a, 2014b). Current developments are hopeful, but not without limitations. Models need to be validated and data needs to be analyzed over time. Accurate longitudinal criminal network data remains an exception. Keeping a long-term intelligence picture of a criminal network and its actors remains an issue for law enforcement agencies. Network studies aiming at identifying durable information positions (e.g. criminal informants) within criminal network structures could help law enforcement agencies towards a data-driven approach in the right direction.

Acknowledgments

PMAS acknowledges partial support from the Russian Scientific Foundation under project #14-21-0137.

References

Ashby WR. Requisite variety and its implications for the control of complex systems. Cybernetica (Namur) 1958;1(2).

Soska L, Christin N. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In: 24th USENIX security Sym- posium 33; August 2015.