Key aspects of covert networks data collection

Diviák, Tomáš

Published in:
Social Networks

DOI:
10.1016/j.socnet.2019.10.002

Publication date:
2022

Social Networks 69 (2022) 160–169

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Key aspects of covert networks data collection: Problems, challenges, and opportunities

Tomáš Diviák [a,b,*]

[a] Department of Sociology, University of Groningen, and Interuniversity Center for Social Science Theory and Methodology, the Netherlands
[b] Department of Sociology, Faculty of Arts, Charles University in Prague, Czech Republic

ARTICLE INFO

Keywords:
Covert networks
Criminal networks
Data collection
Hidden populations
Framework

ABSTRACT

Data quality is considered to be among the greatest challenges in research on covert networks. This study identifies six aspects of network data collection, namely nodes, ties, attributes, levels, dynamics, and context. Addressing these aspects presents challenges, but also opens theoretical and methodological opportunities. Furthermore, specific issues arise in this research context, stemming from the use of secondary data and the problem of missing data. While each of the issues and challenges has some specific solution in the literature on organized crime and social networks, the main argument of this paper is to try and follow a more systematic and general solution to deal with these issues. To this end, three potentially synergistic and combinable techniques for data collection are proposed for each stage of data collection – biographies for data extraction, graph databases for data storage, and checklists for data reporting. The paper concludes with discussing the use of statistical models to analyse covert networks and the cultivation of relations within the research community and between researchers and practitioners.

Introduction

Since the events of 9/11, there has been a growing interest in the study of covert networks (Morselli, 2014, 2009; Gerdes, 2015a). Covert networks are defined by the aim of actors involved in them to avoid detection and to remain concealed (Oliver et al., 2014; Morselli, 2009). The fact that actors aim to avoid detection affects research on covert networks and also data collection in this area. Primary data collection is almost impossible under the assumption that actors aim to avoid detection, because reporting on fellow members of the network and activities shared with them would violate their secrecy. Thus, researchers have to rely on secondary data from sources such as phone wiretaps, police investigation documents, or even media, which bears its own issues and disadvantages.

The research on criminal networks has already brought revealing insights mainly by identifying central actors and describing network structures. As for central actors, previous research focused on their roles within the networks or on their individual attributes. Regarding covert network structures, previous research investigated their density, centralization, or segmentation into subgroups (for a comprehensive review, see Faust and Tita, 2019; Bichler et al., 2017). Our ability to generalize findings, point out contradictory results, and innovate research relies on our ability to be able to compare results across multiple studies. In order to do so, it is necessary to be able to assess to what extent results are comparable. Comparability is then dependent not only upon applied measures, but also on the data and the way it was processed prior to the analyses. However, the way data is collected, stored, and processed is frequently not treated systematically, which complicates not only the assessment of a single study, but also our ability to make cross-study comparisons and meta-analyses as a crucial step in advancing any field of inquiry (Cumming, 2012).

In this paper, I discuss the issues, decisions and complications of data collection on covert networks. I argue that being aware of these problems and being transparent about which decisions were taken during the process of data collection, coding, and analysis does not only add more clarity in the research, but may also contribute to research in this area in three important ways. First, it enables meta-analysis and comparison which is important to be able to derive more general conclusions. Second, there are various theoretical points and research questions that cannot be addressed without a clear delineation of some aspects in covert network data. For instance, it is impossible to study dynamics of covert networks without distinguishing different time periods in the

* Corresponding author at: University of Groningen/ICS, Department of Sociology, Grote Kruisstraat 2/1, 9712 TS, Groningen, the Netherlands.
E-mail address: t.diviak@rug.nl.

https://doi.org/10.1016/j.socnet.2019.10.002

Available online 31 October 2019

data. Such efforts unlock new research questions and contribute to theory formation in the field, which is considered to be underdeveloped (Carrington, 2011; van der Hulst, 2011). Third, better data allows to use more advanced methods, such as statistical models for networks, and to combine social network analysis (SNA) with qualitative approaches (Bellotti, 2014; Domínguez and Hollstein, 2014; Snijders, 2011; Robins, 2013). The goal of this paper is two-fold. The first goal is to review the main issues of data collection for covert networks together with good practices in dealing with them. The second goal is to argue for a more systematic approach towards data collection in order to increase transparency and comparability of research.

I start with identifying six key aspects of covert network data. Each of these aspects comes with a specific set of challenges and problems. Each aspect also comes with a specific set of theoretical opportunities, which may be addressed with better data. I demonstrate each of the identified problems using real data, which are all publicly available in the covert networks database maintained by the Mitchell Centre for Social Network Analysis at the University of Manchester (2019). For each aspect, I outline the problems first, then I show a fruitful approach towards it, and I also show which theoretical questions may be addressed. Furthermore, I discuss some considerations stemming from problems with secondary and missing data. I propose using biographies, checklists, and graph databases as more complex ways to systematically and transparently collect and store covert network data. Note that some problems discussed below also pertain to social network research in general. However, I will not go beyond the subdiscipline of covert network studies, as there are specifics in this area of inquiry that make the transition of tools and practices from or to the subdiscipline difficult or impossible in some cases.

Six aspects of covert networks data collection

Nodes

The problem with the definition of the node set is the problem of boundary specification (Laumann et al., 1983). The boundary specification problem refers to the fact that when conducting a network analysis, researchers need to specify which nodes to include and which to exclude from the network representation. Two broad approaches can be distinguished. In the nominalist approach, the researcher imposes some external criteria on the network (e.g., nodes are included based on shared membership or because they were mentioned in a certain document). In the realist approach, the nodes themselves define the boundaries (e.g., respondents nominate other respondents). Because covert network data are usually secondary, this puts the researcher into the nominalist approach.

The question then is how to set the boundaries or what criterion to use for the inclusion/exclusion of nodes. This has far-reaching implications for calculations and the interpretation of results. One important decision needs to be made about including only directly involved actors or actors from the broader social context as well, which depends on the research question such as when investigating recruitment, support, or acceptance of covert activities by overt actors. Additionally, in some cases of criminal networks, it may be necessary to consider the inclusion of victims, such as in the case of women trafficking (Mancuso, 2014), which shows how victims interact with offenders and thus actively contribute to the organization of crime, or in the cases of fraud, in which the fraud diffuses across victims and thus it wouldn't be possible to understand it fully without considering the victims (Nash et al., 2014). Similarly, in trafficking and illegal commodities distribution networks, this consideration needs to be made with regard to both the supply and the demand side, that is, whether both producers and consumers should be included. Lastly, especially important for terrorist groups, it needs to be clearly stated whether the studied network includes actors participating in one particular action (e.g., 9/11 hijackers) or whether the network represents the whole organization (e.g., Al-Qaeda).

Morselli (2009: 44–45) proposed what he calls criminal justice rings, which refer to different stages of criminal investigation. Criminal justice rings describe the increasing precision of information contained within criminal justice data sources. It is the least precise about actors who happen to be monitored in general criminal monitoring (the widest criminal justice ring) and it is the most detailed about those actors who are confirmed as guilty. Although not originally intended for this, the criminal justice rings can be used as a framework for boundary specification. Defining the boundary of the networks by a specific criminal justice ring provides a criterion which can be compared to other definitions of boundaries, e.g., to other criminal justice rings, and subsequently subjected to sensitivity analysis. Similar approach was taken by Ouellet and Bouchard (2018) in their study on the Toronto 18 terrorist network. They found that considering only the 18 actors charged in the case captures predominantly the operational subpart of the network, whereas if 22 complementary non-charged actors are included, it also captures the ideological component of the network. In some cases, it may not be possible to draw a clear-cut boundary based on criminal justice rings, yet varying criteria may still be used to draw boundaries. As an example, consider Krebs' (2002) analysis of the 9/11 network. Krebs showed that with the inclusion of wider sets of actors the structure changed in some aspects (depicted in Fig. 1): it shortened the distances among actors (diameter dropped from 9 to 7) and also made the network denser (average degree increased from 2.84 to 4.77), whereas transitivity and centralization do not change markedly. In general, exploratory research may inspect several different network boundaries, whereas explanatory research should consider the boundaries corresponding to the research question, both types of research with regard to limitations of the data and its sources.

The definition of network boundaries in several, more or less fine-grained ways, opens opportunities to answer theoretical questions on the embeddedness of covert networks in overt settings by comparing boundaries based on substantively different criteria. This is important for the study of recruitment patterns, as for instance Sageman (2004) showed that the involvement in terrorist networks is a gradual process facilitated by expressive ties to those, who are already involved in radical and/or terrorist activities. Another theoretical problem, which may be addressed by using a more fine-grained distinction between different types of network boundaries, is the facilitation of organized crime in legitimate settings. Previous research showed that illegal activities are facilitated by connections to actors who are not directly involved in criminal activities, but have specific skills (e.g., lawyers or accountants, Morselli and Giguere, 2006).

Ties

The problem with ties is how to define the content of ties, specifically how to treat substantively different types of relations, such as personal ties, criminal cooperation, or exchange of resources. It used to be quite common, perhaps due to paucity of available data, to aggregate different types of ties and interpret the results as if these ties represented cooperation. This potentially leads to misinterpreting ties such as kinship as if they automatically implied criminal cooperation. In the seminal study by Erikson (1981), she points out the crucial role of pre-existing ties for covert networks, which has since then been documented in many other cases (Diviák et al., 2018a, 2018b; Smith and Papachristos, 2016). Conflating these relations would make it impossible to investigate the social embeddedness of criminal ties.

The challenge for researchers is how to distinguish different types of ties substantively as well as actually in the data. Some studies proposed a more general framework for multiplex covert or criminal networks. Smith and Papachristos (2016) distinguished three types of ties relevant for criminal networks: personal, legal, and criminal relationships. Bright and colleagues (2015) specifically aimed at mapping exchange of resources and classified multiple resources as tangible and intangible. Diviák and colleagues (2018b), distinguish three types of ties based on

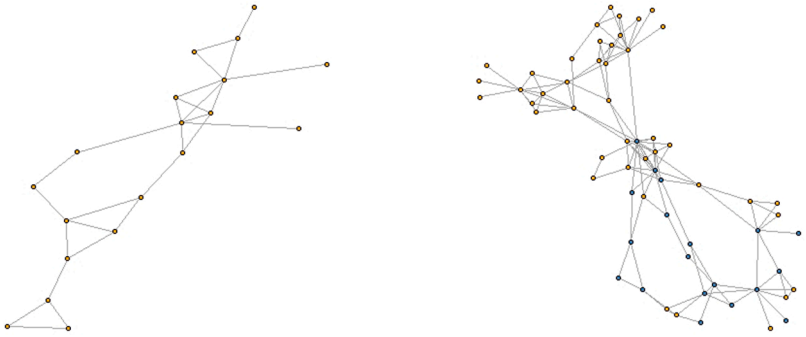

Fig. 1. 9/11 perpetrators network with only those, who hijacked the planes (left) and with other associates (right, hijackers = blue nodes). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the theoretical elements of corruption networks: collaboration, resource transfer, and pre-existing ties. The example in Fig. 2 is taken from Diviák et al. (2018)a, 2018b, which illustrates why it may be potentially misleading to aggregate different types of ties. The two depicted layers are collaboration and resource transfer. Although they overlap (in 22% of cases a tie in one layer is mirrored by a tie in the other), aggregating the two layers would yield a network in which a tie could be interpreted as transferring resources even though it might not be the case. Thus, conflating different types of ties may yield misleading results, which may further distort, for instance, centrality indices, as some actors may be narrowly focused in one type of tie, while others may have their ties spread more evenly across multiple relational dimensions. In a network with all ties aggregated, focused as well as multiplex actors may appear to have the same centrality, even though they are actually central in different ways. Given the heterogeneity in identified types of ties in the literature, it is not surprising that Gerdes (2015b) identified ten different classes in his review of different classifications of ties in covert networks. Although it is understandable that the coding will be different across studies as they will always depend on theory and available data, one generalization can be drawn from this – the choice between coding/-classification scheme for ties needs to balance specificity and generality. On the one hand, a classification scheme that is too specific yields very narrow categories which may be difficult to code reliably, as the information in the data sources may not be precise enough. On the other hand, too general classification yields codes containing heterogeneous relations/interactions, which makes it difficult to interpret validly. Sometimes, the data source may not be specific enough about the content of ties, as some scientifically interesting information may not be considered essential by courts or police investigators. If that is the case, researchers may at least try to distinguish ties reflecting some sort of activity related to the case at hand (e.g., communication or collaboration) from ties representing some antecedents to the case or relational opportunities (e.g., pre-existing ties, similarities, or distances).

Paying attention to different types of ties allows to clearly focus on a specific relation among actors in the network (e.g., focusing only on the flow of resources without confounding the results by pre-existing ties). Considering different types of ties jointly yields a great theoretical opportunity to study multiplexity in covert networks, referring to the fact that there may be multiple types of ties among the same set of actors. Treating covert networks as multiplex may help us understand some of their specific features. Some authors argue that multiplexity compensates for the lack of legitimate institutions enforcing contracts in covert settings by anchoring criminal relationships in other types of relationships (Smith and Papachristos, 2016). Acknowledging the multiplex nature of covert networks enables to study its underlying mechanisms. For instance, tie exchange, which denotes the tendency of actor to reciprocate a tie of one type with a tie of a different kind, such as in the case of exchange of different resources (Bright et al., 2015). Another mechanism worthy of attention is tie translation, that is, the tendency to create ties on the basis of already existing ties of different kind (Diviák et al., 2019), which may be one way how to operationalize the importance of pre-existing ties for creation operational criminal ties.

Attributes

Attributes come into play in covert network analysis in two ways. First, attributes capture substantively meaningful characteristics of actors, which create opportunities and constraints for individual behaviour including creation, maintenance, and dissolution of ties, or for reaching individual goals (Robins, 2009; Steglich et al., 2010). This is something which analysis of covert networks shares with the rest of SNA. However, due to specific circumstances with covert network data, the data collection may be focused on particular individuals creating what Smith and Papachristos (2016) call the 'spotlight effect'. Whereas descriptive measures (e.g., centrality measures) cannot really account for this, it is important for a correct interpretation to know who was in the spotlight. Models can include control nodal variables for each of these and thus correct for the effect of data collection which might otherwise distort the results (Bright et al., 2018; Smith and Papachristos, 2016). Thus, the second role played by attributes in the analysis of

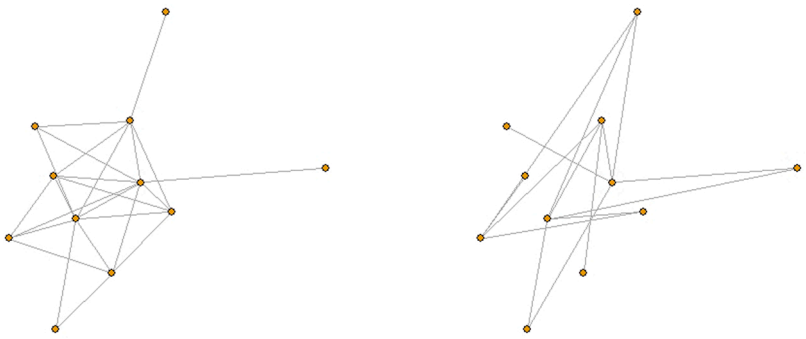

Fig. 2. A corruption network with two types of ties: collaboration (left) and resource transfer (right). The position of nodes is the same in both sociograms.

covert networks is that of variables helping to account for how the given dataset was collected.

It is therefore important to know which variables we want to measure substantively and whether we need any control variables to account for the data collection. In terms of the substantive attributes, which attributes to analyse and how to define them depends heavily on theory. One parsimonious approach which may be helpful in systematically transposing theory to data collection is script analysis (cf. Morselli and Roy, 2008). Script analysis decomposes the process of organizing illicit activities into a sequence of events. The idea is that in each part of the illicit script different types of activities need to be carried out by different actors with particular skills. For example, production and distribution of drugs requires someone first to actually create the product, then it is necessary to distribute it, and perhaps it is also necessary to protect the dealers. From this simplified script, three types of roles can be identified, which may be used as attribute(s) in the analysis – cooks, dealers, and thugs. With regard to attributes as controls, researchers may want to include an attribute referring to whether an actor was among the initial nodes under surveillance, as further observations are contingent upon being related to those under the initial surveillance. If the surveillance proceeds to further focus on those connected to the initial set of nodes, it starts to resemble a snowball or link-tracing sample and it may even be worthwhile to analyse the resulting network with appropriate methods for snowball and link-tracing samples (Heckathorn and Cameron, 2017; Pattison et al., 2013). An example of a control variable for the spotlight effect is Smith's and Papachristos' (2016) study on prohibition era Chicago criminal networks, where all the information revolved around Al Capone and so authors created a dummy variable which had the value of 1 for Al Capone and 0 for the rest of actors.

Traditional quantitative criminology has focused on identifying important predictors of individual characteristics related to important criminological concepts such as delinquency, substance abuse, or commission of different types of crime. Network research may enrich the modelling of individual level outcomes with structural network effects (e.g., positions of actors within networks). This is arguably an important area of further research, as traditional individual profiling has been criticized for having a poor explanatory power (cf. Horgan, 2008), but there are indications that structural network effects may be key to more profound explanation of phenomena such as involvement in terrorist activities (Sageman, 2004). This does not include only attributes in the role of substantively meaningful variables, but also in the role of control variables. Attributes as controls may be investigated as dependent variables providing the opportunity to reflect upon investigation and surveillance methods. On the one hand, it is possible that investigations overlook individuals with specific traits or network positions. On the other hand, they might predominantly focus on specifically positioned and predisposed actors.

Levels

Some covert network datasets have an intrinsic bipartite or even multilevel structure. For instance, Crossley and colleagues (2012) and Calderoni and colleagues (2017) studied networks of co-participation in arrests or in meetings, which are essentially bipartite networks with actors in the first mode and arrests/meetings in the second mode. Often, this is the only possibility to collect data on covert networks as exact information on interaction between actors is difficult to obtain. All network information then is derived from co-participation, co-appearance, or co-membership structures. However, it is important to note that affiliation does not necessarily mean interaction, it is only an opportunity to engage in it (Borgatti and Everett, 1997). This fundamentally limits what inferences we can draw from such data.

What researchers often do when they study co-participation structures in covert networks is that they either explicitly or implicitly work with a projection from two-mode data to one-mode. It is important to seriously consider the consequences of such data transformation, as it comes with the loss of information about the structure of the network. For example, 3-star and 6-cycle configurations in two-mode networks both yield a triangle in one-mode projection, albeit being initially very different structures. This illustrates that projection artificially introduces closure and clustering into the data. Therefore, care needs to be taken when interpreting these findings, as they may not be genuine tendencies of actors to form transitive ties, but rather a product of projection. For example, Fig. 3 captures the initial bipartite network of N'dranghetta mafiosi and their meetings. The bipartite network's density is 0.06 and its transitivity (measured by bipartite clustering coefficient) is 0.46, whereas the actor-to-actor projection (where ties represent co-attendance in events) displays density of 0.13 and clustering coefficient of 0.58. But the loss of information also applies to information about the attributes of the second mode, that is, settings, places, affiliations, or groups. These may themselves be an important part of the explanation, which is completely disregarded when focusing solely on the actor-to-actor projection. It is a matter of the specific research question whether projection is a fruitful avenue for the study of a given network, or whether the loss of information hinders crucial parts of the explanation.

What I propose is to carefully consider projecting the data, as the original bipartite structure not only contains full information, but might also be worthwhile to investigate in itself. Bipartite networks offer a way to study an important theoretical concept in criminology – convergence settings (Felson, 2006, 2009). Convergence settings denote social or spatial settings that facilitate crime or cooperation of offenders, such as clubs, bars, restaurants or parks. This concept has also been used in the literature on extremist networks as radical settings (Wikström and Bouhana, 2017) facilitating radicalisation, diffusion of norms and ideas, providing an opportunity to pool resources for extremists such as clubs, shops, extremist party secretariats or radical temples for religiously motivated offenders. These settings can be operationalized as a mode in bipartite networks. This approach may in turn draw upon recent developments in the methodology for both descriptive analysis of bipartite networks (Everett and Borgatti, 2013) and for modelling of such network structures (Wang et al., 2013; Lazega and Snijders, 2016). The extension to multilevel network opens the possibility to analyse the relationship between cooperation among criminals (first level) and its facilitation by certain convergence settings (second level) or to jointly analyse ties among actors (e.g., gangsters), their affiliations to groups (such as gangs), and ties among the groups (such as territorial disputes).

Dynamics

It has been emphasized that covert networks are flexible, adaptive, and dynamic. Yet such claims have primarily remained metaphorical assumptions rather than empirically shown properties which has already been pointed out elsewhere (Campana, 2016; Bright et al., 2018). This may be due to lack of appropriate data to study the evolution of covert networks over time. However, there are pioneering studies aiming at unravelling the process of evolution of these networks and data are becoming increasingly available. Assessing covert network dynamics is a crucial task as it allows researchers to empirically test the concepts of flexibility and adaptability, and it also enables practitioners to improve monitoring and interventions in covert networks. For instance, without longitudinal data researchers cannot distinguish between the processes of selection and influence and therefore to assess whether a particular observed pattern is an outcome or a precondition (Steglich et al., 2010). For practitioners, cross-sectional data aggregated over time may yield a picture of a network which in this form actually never existed at any given time point (e.g., one actor might have died before another one joined). This may have serious implications for designing an intervention.

The first issue related to longitudinal covert network data collection is how to define the periods or waves for coding and/or observation of

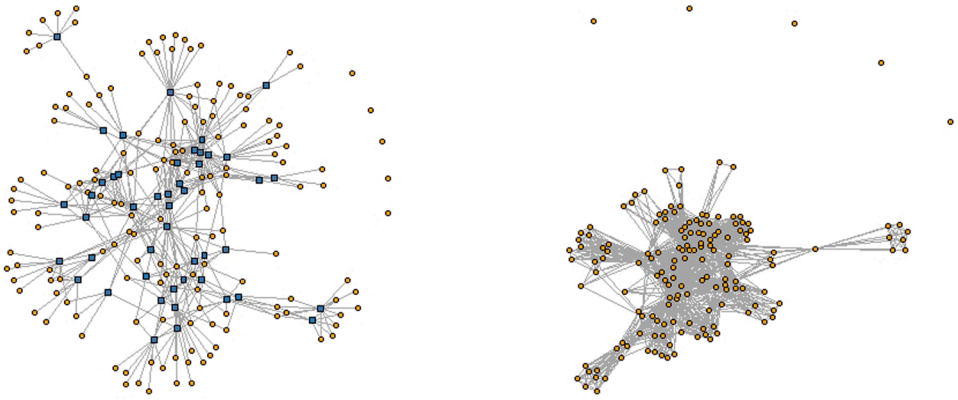

Fig. 3. A bipartite network of mafiosi and their meetings (left; Mafiosi = yellow circles) and corresponding mafiosi-to-mafiosi projection (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the network. Generally speaking, there are two possible ways to do this: time-based and event-based (Campana and Varese, 2012). A time-based definition requires to select specific time periods (e.g., weeks, months or years) and subsequently record the state of the network in each of these periods. An event-based definition demands to define specific events in the evolution of the network, which were theoretically important and/or interesting. Whereas the time-based may seem to be more based on 'objective' time periods, testing certain hypotheses about development of structures in response to particular events (e.g., disruption attempts) or environmental conditions (e.g., abundance of opportunities for organized crime) may require more theoretically founded periodization. Related to this is the question of successful and failed covert networks, as one might argue that all the studied covert networks are failed cases, as they were uncovered after all (Morselli, 2009). Hence, these cases are supposed to provide a distorted picture of reality as the successful ones elude the attention of researchers and practitioners alike. A counterargument may be that success or failure is not a fixed binary state, but rather a status changing over time. Therefore, some networks may be considered successful (such as reaching their collective goal) at some point in time, but they may be uncovered and dismantled at another time point, considering them as failed at that point. This is demonstrated with an example of a drug trafficking network originally analysed by Morselli and Petit (2007). Fig. 4 shows how the activity of actors in the network (measured by average degree) changed over time depending on how successful (for instance, at time points 4, 6, and 10) or unsuccessful (for instance, at time points 5 or 8) it was in terms of distribution of illegal drugs.

Longitudinal data opens up the opportunity to assess the recovery and adaptation of covert networks after disruption. Research has shown performance and effectiveness of different disruption strategies, such as central node removal or random node removal (Bright, 2015). While simulation studies, for instance, consistently show that central node removal is a more efficient strategy for disruption than random node removal, they do not provide further evidence about the process of recovery from disruption. This is, however, crucial, as some observational studies show that attempts to disrupt covert networks may trigger unintended consequences and actually strengthen their structural cohesion (Duijn et al., 2014). Longitudinal data provides the opportunity to combine simulation and observational research and to realistically simulate not only the impact of disruption strategies, but also recovery from disruption. Vigorous development of models for network dynamics in recent years (cf. Snijders et al., 2010; Block et al., 2018) equips researchers with tools to address these issues and thus to move from metaphors to empirical evidence.

Context

The very definition of covert networks, covertness, is contingent upon the context of the network. Why is it covert? From whom? And how? It is assumed that covertness critically modifies the structure of

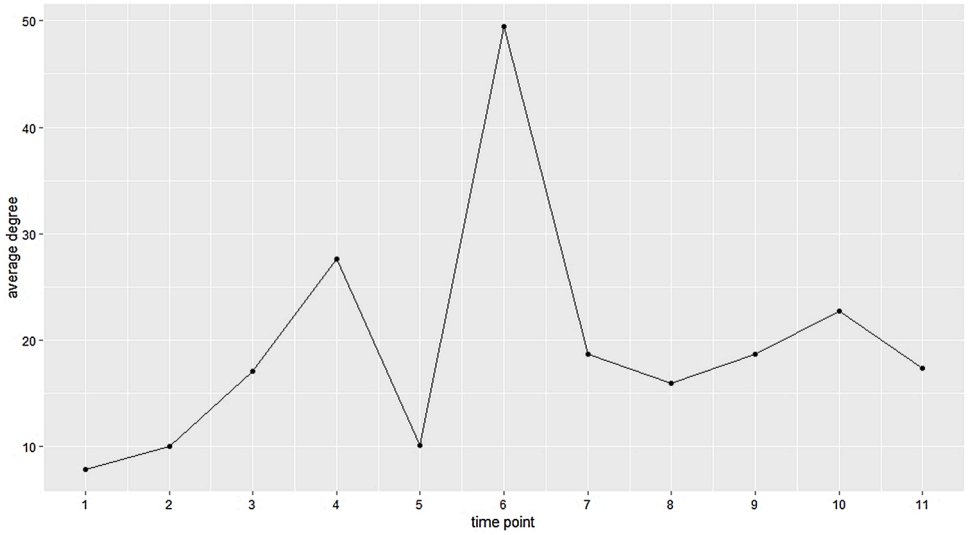

Fig. 4. Average degree of actors involved in a drug trafficking over eleven time points.

networks and thus justifies the study of covert networks as distinct from overt networks (Morselli, 2009). However, the information about context is frequently more qualitative and non-network, i.e., difficult to combine with network structure, as it goes beyond nodes and ties. At the present, vast majority of studies incorporates these non-network aspects of their studies as a brief description in the section of case or context description, and subsequently, some of the information is ad hoc evoked when interpreting results of network analysis. It is of course pivotal for a good study to situate the SNA results within the context to adequately interpret findings and draw valid conclusions from the results. However, the contextual information should be used systematically. The danger here is in confirmation bias – choosing only those bits of contextual information which confirm the theory rather than scrutinizing the network analytic results with confirming as well as rejecting contextual information.

In essence, this touches upon a broader recent methodological debate on how to combine qualitative methods with SNA (Bellotti, 2014; Domínguez and Hollstein, 2014). Almost all empirical studies of covert networks are case studies as they examine a particular network within a given context with respect to some aspects which are deemed as theoretically important. This may seem obvious and not very revealing, however the realisation that these studies are in fact case studies is crucial (Crossley and Edwards, 2016). There is now a growing body of methodological literature on systematic case study research from which the study of covert networks (or networks in general) may draw inspiration. Two promising methods are process-tracing (Beach and Pedersen, 2013) and qualitative comparative analysis (QCA; Rihoux and Ragin, 2009). Process-tracing is a way to systematically use both network and qualitative evidence with regard to a given theoretical explanation of a case at hand. It provides a method to qualitatively test whether a certain condition is necessary or sufficient to explain given outcome. QCA offers a way to rigorously compare several cases, using set theory and Boolean algebra. Both network and non-network variables can be included in such analysis. The method can then distinguish different configurations of conditions to show which conditions and how they affect the outcome of interest (Fischer, 2014). This is in principle similar to using meta-analysis, although QCA may be especially useful in studies where non-network qualitative aspects are important for explanation, which may be difficult in traditional meta-analysis of network statistical models (cf. Lubbers and Snijders, 2007), and in cases where comparison of smaller number of cases is done (e.g., five to ten). For instance, one may be interested in successful commission of terrorist attacks (an outcome). It may hypothetically be argued that centralized network structure, short distances among actors, sufficient resources, and absence of law enforcement opposition explain the success of a terrorist attack. A researcher may collect data on several networks, some of which succeeded in committing an attack. QCA may then be used to assess which combinations of network (centralization and path length) as well as non-network factors (law enforcement and resources) are related to the outcome, and how.

The treatment of qualitative contexts opens up the opportunity to put the same weight on both network and non-network information in explaining studied cases. An important research issue is the individual perception and phenomenology of network structures and positions within them (Hollstein, 2014). For instance, the concept of strategic positioning has become frequently studied in criminal networks (Bright et al., 2015; Diviök et al., 2018a, 2018b; Morselli, 2010). Strategic positioning refers to tendency of some actors in covert network seek out less visible positions (low degree) while retaining influence by being on top of many flows (high betweenness). From the point of view of the researcher, strategic positioning is usually explained as the attempt of actors to reduce their exposure while retaining some influence within the network. However, the intentions of these actors and their motivations for seeking (or avoiding) such positions may be very different, such as when actors are behaving “irrationally” in terms of their network positions. This happens, for instance, when actors proliferate their ties and thus expose themselves to detection, because they are strongly self-confident and believe they are invincible because of their elite membership status (e.g., politicians; Demiroz and Kapucu, 2012; Diviák et al., 2018b).

Further considerations

In this section, I will discuss further considerations which typically arise in the research on covert networks. Note that these considerations are not a standalone aspect of data collection, but relate to all six aspects covered above.

Secondary data

As already stated above, research on covert networks usually draws upon secondary data, limiting researchers to whatever data that is available. This data may come from offender databases, transcripts of physical and/or electronic surveillance, summaries of police interrogation, transcripts of court proceedings, and on-line and print media (Bright et al., 2012). None of these types of sources is perfect in terms of validity or reliability. In terms of validity, a critical issue is that none of these sources is primarily collected for research purposes. Those who collect and process these data do so for very specific purposes, which critically determine the type of information available in the data source. So while researchers may, for instance, be interested in communication among a group of offenders, using data on phone calls among them does not capture their face to face communication. Similarly, some important piece of information may not be recorded, yielding invalid representation of the phenomenon in question. For example, police interrogation may not uncover certain features of the investigated criminal activities, which offenders themselves may be motivated to hide from police. Or some offenders may not yet be caught and thus they do not figure in the offender databases. In extreme cases, this may yield analytical results which are merely artefacts of the data collection. In order to assure that the data does not yield artificial results, clear and mutual information exchange between researchers and practitioners is necessary so that practitioners are familiar with up-to-date research methods and findings and researchers are well aware of potential blind spots in the data.

In terms of reliability, a key issue is that the procedures used to collect data are not always consistent across different researchers, practitioners, and/or organizations. This has obvious implications for potential comparability of results based on data from different sources. Sometimes, inconsistencies may occur even within organizations or teams of practitioners as their personnel fluctuates or as their rules and regulations change. Both researchers and practitioners may benefit from guidelines for data collection. The point here is not to mentor the practitioners but rather, to make their work easier by contributing to it with scientific knowledge and best practices on how to deal with difficulties they encounter in their daily work such as how to code different relations, define temporal periods or network boundaries. This could pay off to researchers with better data eventually as well as build better relations with practitioners, which may make the data more accessible, and it can improve the work of the organization in question.

Secondary data often entail another obstacle - accessibility. All data sources outlined by Bright and colleagues (2012), except for media sources, are in the possession of law enforcement agencies and have strict rules about the conditions of their use in scientific research. At present, very little is known about how different data sources compare on different criteria such as accuracy or analytical depth of information. There are only a few studies comparing results based on different data sources about the same covert group. For instance, Rostami and Mondani (2015) analysed a network of a Swedish gang based on criminal intelligence data, co-offending records, and police surveillance. They found substantial differences in terms of centrality measures between intelligence data and the other two sources. Another study was conducted by Berlusconi et al. (2016) on a network of Italian Mafiosi with

the aim of inferring missing ties. This study used wiretap records, arrest warrants, and judicial documents, and showed that considering the same set of actors, the network of wiretap records is the densest. Media-based data are usually thought to be less valid than the remaining sources. However, such claims seem to be based solely on face validity, as no sound comparison of media-based data with other sources has been made. If there is enough evidence that media-based data consistently yield network data similar to other sources, it may encourage their more frequent usage considering that these data are also easier to access. However, this comparison may also provide substantiated evidence against using media-based, if they yield network representations incompatible with other sources. Alternatively, comparison of different data sources may point out systematic differences, which can in turn give us a hint how to use multiple data sources for triangulation.

Missing data

Missing data traditionally pose a problem for any quantitative method. In SNA, this problem maybe even more severe because of the interdependence inherent in the data. Results of some studies indicate that some network measures are quite robust even when dealing with networks containing a considerable amount of missing data (Borgatti et al., 2006; Smith and Moody, 2013; Smith et al., 2017). Yet, this robustness does not necessarily translate to the individual level and highly depends on the missing data mechanism (mechanism generating the missingness, Krause et al., 2018). Missing data present probably the most frequent objection to covert network data, which is due to the very nature of covert networks; they are covert, so it is likely that some piece of information will not be uncovered by researchers and/or practitioners.

Good practice in current research is to acknowledge this as a limitation. However, the problem with missing data should not only be acknowledged, but also tackled. In recent years, there has been a development of methods for handling missing data in networks (Krause et al., 2018; Robins et al., 2004; Huisman and Steglich, 2008). Although researchers may surely use these methods to their advantage, these methods are not automatically saving poorly collected datasets. The first thing researchers in covert networks have to realize is that there are different missing data mechanisms: missing completely at random (no relation to any observed or missing variable), missing at random (no relation to missing variable, but related to some observed variable), and missing not at random (whether some data point is missing itself depends on non-observed variables; Rubin, 1976). In covert networks, it likely that researchers will not only be dealing with data missing (completely) at random, but also with non-randomly missing data. This may stem from variety of reasons. Some highly prominent actors may have the tendency towards intentionally concealing themselves or some type of ties may be more likely to be missed due to their level of sophistication (i.e., encrypted messages). In order to at least correctly alleviate the problem of missingness, it is first necessary to identify the missing data mechanism. Then, appropriate imputation techniques can be applied.

However, before that it is important to know which information (which nodes or ties etc.) is actually missing. What researchers in this area are usually dealing with is an adjacency matrix with ones representing the existence of a tie and zeros representing the absence of a tie. The ones and zeros mask an important thing – both may be true or false. While the existence of ties is usually confirmed and thus the ones in the data are actually true ones, the absence of ties is usually not considered to require further confirmation. However, this is problematic. In order to be able to deal with missing ties, we need to be able to tell which ties are absent (i.e., it is known that there is no tie between a given pair of actors) and which are missing (i.e., we do not know whether the tie exists or not). One way to work around this problem is to use existing ways intelligence services use to classify the reliability of any given information, based on either cross validation by different sources or a measure of reliability of the original source. Sparrow (1991) mentions one such classification, where law enforcement investigators classify ties as 'strong' if their existence is confirmed from two independent sources, whereas 'weak' ties are those without an independent confirmation. A cautious analyst may want to work with weak ties as if they were missing ties and use some of the newly developed methods to impute them or they can analyse different variants of the network and see how the results differ. Of course, knowledge about which information is not confirmed may not always be available, but at least in the cases of working police investigation files or media databases (where some information is only "suspected" or "speculated"), this approach may be a way how to incorporate the uncertainty in a covert network study.

Ways forward

The points I raised above beg the question whether there is some more general and complex framework for a more systematic approach to covert networks data collection. In this section, I discuss three such frameworks – biographies, graph databases, and checklists. These three frameworks can be used in data extraction, data storage, and in reporting how the data was processed. Since these three frameworks cover different areas of data collection and they are not mutually exclusive, they can be used together in one study, in selected pairs, or just individually depending on the study at hand.

The first stage of data collection that researchers are usually confronted with is extracting the data from a source material such as court files or transcripts of police surveillance. Some sort of content analysis is typically used to code relevant information from the data source and to turn it into network data. Such coding can be done simultaneously by different coders and the reliability of the coding can be subsequently checked. However, little is usually known about how to approach coding, i.e., what type of information to look for and how to store it. Constructing so-called biographies (van Nassau et al., 2019) can be useful for this task. Such a biography is a table whose rows represent nodes and whose columns represent time points. Each individual cell (node $\times$ time point) then stores all the available information about the given node at the given time point. Specification of the node set as well as definition of time points is dependent upon selected boundaries and definition of periods. The information stored in each cell should ideally correspond as much as possible to its counterpart in the data source, which it should refer to so that the information can be easily backtracked. For example, a cell for actor A and year N may state "had repeatedly been meeting B in setting S (court file F)". Once all the available source information has been extracted into a biography, it may be coded and recoded as necessary, and so different types of ties may be distinguished, actor attributes assigned to actors, or multiple modes (such as settings) may be identified. Also, different node sets (e.g., affiliations) may be used as a starting point and periods may be recoded depending on the precision and depth of available information, as in practice, both the information about actors or time points may not necessarily be as fine-grained in some sources (typically in media or court files) as researchers would like it to be.

For storing collected data, the proposal of Gutfraind and Genkin (2017) to use graph databases may be useful. Graph databases store the information in a relational form of multimode and multilayered graphs, where pieces of information are represented as nodes and relations among them as edges instead of rows and columns. For instance, a transcript of surveillance describing a meeting between two actors in a bar can be represented in a graph database (visualized in Fig. 5) as a three-mode network where the modes represent source of the data, actors, and location, connected by edges representing mentions (solid lines from the source file A), meeting (dashed line between actors B and C) and shared location (dotted lines to location C). Different networks may then be obtained from a graph database by using suitable projection techniques. Gutfraind and Genkin (2017) argue that this approach makes processing and transformation of data more transparent and

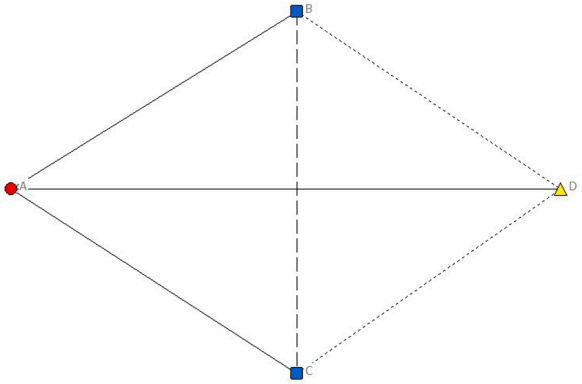

Fig. 5. An example of simple graph database depicting a source file (node A), two actors (B and C), and a location (D) connected by ties representing mentions (solid line), meeting (dashed line), and shared location (dotted line).

easier to reproduce, adding to its generalizability and comparability of the findings. From the six aspects outlined above, graph databases can readily capture five of those in a transparent and unified manner – nodes, ties, attributes, levels, and dynamics. Actors can be represented as nodes in one of the modes in the graph database, ties can be represented as ties with the capacity to distinguish different types of ties. The bipartite or multilevel structure can be similarly included in the graph database as another mode and similarly for nodal attributes. Even network dynamics can be captured, for example by creating two separate graphs for two periods. The only aspect which may not have a clear representation in a graph database framework is the qualitative context, although there may be ways to incorporate this aspect (perhaps as yet another mode of nodes in the graph). Graph databases are an efficient way to use already collected data by merging, dividing or projecting the data to obtain a dataset feasible for answering a given research question. Moreover, such a way is principled, because it is possible to backtrack what was not included in the final analysis. Graph databases may seem considerably technically complicated, but even if researchers do not want to use them, they may consider using similarly constructed edgelists for their data collection and storage as a somewhat simplified variant. Such an edgelist should contain not only the information about which node is connected to which other node, but also about the types of ties, actor attributes, and all the available information on the remaining aspects of covert network data together with a reference to the data source (e.g., a specific court file), the exact citation on which each entry is based (e.g., "A and B were reported to be together…") and a comment on some further contextual information (e.g., whether an actor is aware of being a part of a larger covert network).

There are no universal rules or algorithms prescribing exactly how to extract, store, and process covert network data. Arguably, this lack is understandable given how varied and differentiated network research is even if we consider only the subfield of covert networks research. Thus, what type of information will be coded in a biography or how a network will be derived from a graph database or a detailed edgelist depends on given research problem. In this area of research, research questions are not only delineated by theory, but also by practical limitations complicating all the supposedly ideal decisions. However, in order to facilitate comparability of findings and accumulation of knowledge, researchers need a common frame of reference. In such a frame of reference, researchers should be able to clarify both theoretical underpinnings and practical constraints of their data collection. Volk et al. (2017) propose a simple checklist for researchers studying bullying, which is supposed to enhance validity and generalizability of studies in that area. Volk and colleagues' checklist contains five items considering mainly theoretical assumptions and clarifications. I propose a checklist based on the aspects and considerations discussed above pertaining to covert network data collection that could enhance transparent and systematic reporting of the way we handle our data:

1 What are the nodes and what were the criteria for their inclusion in the network?
2 What types of relations/interactions do the ties represent?
3 What are the theoretically relevant attributes of nodes and are there any variables mitigating the effect of the way the data were collected?
4 What are the modes distinguishable within the raw data and in what way is the final network representation obtained?
5 What is the temporal span of the network and if multiple periods were distinguished, how were they defined?
6 What are the theoretically relevant pieces of contextual information and what role do they have in the explanation?
7 What was/were the data source(s) used to obtain the information and in what way was the coding of information into network data conducted?
8 What is the nature of missing data and how was the missingness handled? If it was not possible to distinguish missing data from absent data, what impact may the hidden missing data have on the results?

Conclusion

In the present paper, I discussed different issues, challenges, considerations, and opportunities researchers of covert networks face. I identified six aspects of data collection on covert networks (nodes, ties, attributes, levels, dynamics, and context), all of which contain unique problems as well as opportunities for researchers. All these six aspects are affected by the secondary nature of the data and the problem of missing information. There are fruitful approaches for data collection on each aspect. Besides, I brought up three potentially more general ways which may serve as a common frame of reference, namely biographies, graph databases, and checklists. While all these recommendations and good practices may be useful first steps towards making research more transparent, replicable, and comparable, they are by no means definite solutions to the problems arising in the study of covert networks. However, I hope that this paper will stimulate discussion about what to improve and how to push the research on covert networks further as a whole.

One matter which kept reoccurring in this study was the usefulness of statistical models for network data. There has been a rapid development of statistical models for network data (cf. Snijders, 2011; Robins, 2013), but the research on covert networks is still predominantly driven by descriptive measures (Campana, 2016). There is quite a steep learning curve from basic descriptive measures to advanced statistical models in SNA, but researchers in this field could benefit from investing time and effort into adopting statistical models, as with good data, these models provide powerful and flexible tools for testing a variety of (sometimes even mutually competing) hypotheses. Specific problems arising in the context of covert networks may in turn stimulate further development of network models.

Similarly to statistical modelling, another avenue for future development in the research of covert networks consists of link-tracing and other network sampling methods (Heckathorn and Cameron, 2017). I have touched upon this issue in relation to individual attributes as variables controlling for data collection induced effects. Due to difficulty (or even impossibility) to map the entire network in covert settings, link-tracing and network sampling methods have been considered to be particularly useful for collecting the data on hidden populations (Frank, 2009). At present, very little is known about specific procedures used by police or intelligence services to collect the data, e.g., how they build offenders databases, how they choose whom to surveil, or which phones to wiretap. Mapping these techniques may critically improve the data quality and open the way for using appropriate estimation methods.

Since science is not only a system of knowledge production but also a

matter of social relations and communication, researchers should communicate more with one another and share their knowledge, experience, and data. In short, we as a community of researchers should continue networking. Initiatives such as the Illicit Networks Workshop or organized sessions in both general network or general criminological conferences are productive platforms in this regard. However, this communication and cooperation should not be restricted to the community of covert network researchers. We critically rely on practitioners such as law enforcement agencies, courts, and media and it is necessary to further cultivate our relations with them. Researchers should keep working with practitioners, try to use their data and warn them about potential pitfalls pertaining to data collection and storage. However, this should not be a one-way street – we should reciprocate and show what SNA, and science in general, has to offer for practitioners and how we can help them understand covert phenomena or make their day-to-day routines easier with tools and methods for data collection and analysis. This is especially important given that most of the recommendations outlined above are only available if researchers have access to the data - if not, the practical and logistical constraints prevail over scientific guidelines. Helping practitioners may in turn relax some of the constraints and therefore make our research easier.

Funding

This work was supported by the Charles University Grant Agency (Grantová agentura Univerzity Karlovy) under grant number 256119 'Criminal networks dynamics'.

Acknowledgements

I am grateful to the members of the statistics and network analysis cluster at the Department of Sociology at the University of Groningen for their comments on earlier version of this paper. I am especially thankful to Jan Kornelis Dijkstra, Tom Snijders, Robert Krause, Gert Stulp, and the two reviewers for their help and suggestions.

References

Beach, Derek, Pedersen, Rasmus Brun, 2013. Process-Tracing Methods: Foundations and Guidelines. University of Michigan Press.
Bellotti, Elisa, 2014. Qualitative Networks. Routledge, London.
Berlusconi, Giulia, Calderoni, Francesco, Parolini, Nicola, Verani, Marco, Piccardi, Carlo, 2016. Link prediction in criminal networks: a tool for criminal intelligence analysis. Daniele Marinazzo, ed. PLoS One 11 (4), e0154244.
Bichler, Gisela, Malm, Aili, Cooper, Tristen, 2017. Drug supply networks: a systematic review of the organizational structure of illicit drug trade. Crime Sci. 6 (1) (Accessed 4 February 2018). http://crimesciencejournal.springeropen.com/articles/10.1186/s40163-017-0063-3.
Block, Per, Koskinen, Johan, Hollway, James, Steglich, Christian, Stadtfeld, Christoph, 2018. Change we can believe in: comparing longitudinal network models on consistency, interpretability and predictive power. Soc. Netw. 52, 180–191.
Borgatti, Stephen P., Carley, Kathleen M., Krackhardt, David, 2006. On the robustness of centrality measures under conditions of imperfect data. Soc. Netw. 28 (2), 124–136.
Borgatti, Stephen P., Everett, Martin G., 1997. Network analysis of 2-Mode data. Soc. Netw. 19 (3), 243–269.
Bright, D.A., Greenhill, C., Ritter, A., Morselli, C., 2015. Networks within networks: using multiple link types to examine network structure and identify key actors in a drug trafficking operation. Glob. Crime 16 (3), 219–237. Scopus.
Bright, David A., 2015. Disrupting and dismantling dark networks. In: Gerdes, Luke M. (Ed.), Illuminating Dark Networks: The Study of Clandestine Groups and Organizations. Cambridge University Press, Cambridge, pp. 39–52.
Bright, David, Hughes, Caitlin, Chalmers, Jenny, 2012. Illuminating dark networks: a social network analysis of an australian drug trafficking syndicate. Crime Law Soc. Change 57 (2), 151–176.
Bright, David, Koskinen, Johan, Malm, Aili, 2018. Illicit network dynamics: the formation and evolution of a drug trafficking network. J. Quant. Criminol. (Accessed 5 May 2018) http://link.springer.com/10.1007/s10940-018-9379-8.
Calderoni, Francesco, Brunetto, Domenico, Piccardi, Carlo, 2017. Communities in criminal networks: a case study. Soc. Netw. 48, 116–125.
Campana, Paolo, 2016. Explaining criminal networks: strategies and potential pitfalls. Methodol. Innov. 9, 2059799115622748.
Campana, Paolo, Varese, Federico, 2012. Listening to the wire: criteria and techniques for the quantitative analysis of phone intercepts. Trends Organ. Crime 15 (1), 13–30.
Carrington, Peter J., 2011. Crime and Social Network Analysis. The SAGE Handbook of Social Network Analysis, pp. 236–255.
Covert Networks-UCINET Software N.d. https://sites.google.com/site/ucinetsoftware/datasets/covert-networks, (Accessed 17 February 2019).
Crossley, Nick, Edwards, Gemma, 2016. Cases, mechanisms and the real: the theory and methodology of mixed-method social network analysis. Sociol. Res. Online 21 (2), 1–15.
Crossley, Nick, Edwards, Gemma, Harries, Ellen, Stevenson, Rachel, 2012. Covert social movement networks and the secrecy-efficiency trade off: the case of the UK suffragettes (1906–1914). Soc. Netw. 34 (4), 634–644.
Cumming, Geoff, 2012. Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis. Multivariate Applications Series. Routledge, Taylor & Francis Group, New York.
Demiroz, Fatih, Kapucu, Naim, 2012. Anatomy of a dark network: the case of the Turkish ergenekon terrorist organization. Trends Organ. Crime 15 (4), 271–295.
Diviák, Tomáš, Dijkstra, Jan Kornelis, Snijders, Tom A.B., 2018a. Poisonous Connections: A Case Study on a Czech Counterfeit Alcohol Distribution Network. Global Crime forthcoming.
Diviák, Tomáš, Dijkstra, Jan Kornelis, Snijders, Tom A.B., 2018b. Structure, multiplexity, and centrality in a corruption network: the Czech Rath Affair. Trends Organ. Crime 1–24.
Diviák, T., Dijkstra, J.K., Snijders, T.A.B., 2019. Poisonous connections: A case study on a Czech counterfeit alcohol distribution network. Glob. Crime 1–23. https://doi.org/10.1080/17440572.2019.1645653.
Domínguez, Silvia, Hollstein, Betina, 2014. Mixed Methods Social Networks Research. Cambridge University Press, Cambridge.
Duijn, Paul A.C., Kashirin, Victor, Sloot, Peter M.A., 2014. The Relative Ineffectiveness of Criminal Network Disruption. Scientific Reports 4 (Accessed 29 November 2015). http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3937802/.
Everett, M.G., Borgatti, S.P., 2013. The Dual-Projection Approach for Two-Mode Networks. Social Networks 35(2). Special Issue on Advances in Two-Mode Social Networks, pp. 204–210.
Faust, Katherine, Tita, George E., 2019. Social networks and crime: pitfalls and promises for advancing the field. Annu. Rev. Criminol. 2 (1), 99–122.
Felson, Marcus, 2006. The ecosystem for organized crime. HEUNI 25th Anniversary Lecture. http://www.heuni.fi.
Felson, Marcus, 2009. The natural history of extended co-offending. Trends Organ. Crime 12 (2), 159–165.
Fischer, Manuel, 2014. Coalition structures and policy change in a consensus democracy: coalition structures and policy change. Policy Stud. J. 42 (3), 344–366.
Frank, Ove, 2009. Network sampling and model fitting. *in* models and methods in social network analysis. Reprinted. In: Carrington, Peter J., Scott, John, Wasserman, Stanley (Eds.), Structural Analysis in the Social Sciences. Cambridge Univ. Press, Cambridge, pp. 31–3156, 27 [i.e. 28].
Gerdes, Luke M., 2015a. Illuminating Dark Networks: The Study of Clandestine Groups and Organizations. Cambridge University Press, Cambridge.
Gerdes, Luke M., 2015b. Dark dimensions: classifying relationships among clandestine actors. Illuminating Dark Networks: The Study of Clandestine Groups and Organizations. Cambridge University Press, Cambridge, pp. 19–38.
Gutfraind, Alexander, Genkin, Michael, 2017. A Graph Database Framework for Covert Network Analysis: An Application to the Islamic State Network in Europe. Social Networks. http://www.sciencedirect.com/science/article/pii/S0378873316302428.
Heckathorn, Douglas D., Cameron, Christopher J., 2017. Network sampling: from snowball and multiplicity to respondent-driven sampling. Annu. Rev. Sociol. 43 (1), 101–119.
Hollstein, Betina, 2014. Mixed methods for social networks research: an introduction. In: Domínguez, Silvia, Hollstein, Betina (Eds.), Mixed Methods Social Networks Research. Cambridge University Press, Cambridge, pp. 3–35.
Horgan, John, 2008. From profiles to pathways and roots to routes: perspectives from psychology on radicalization into terrorism. Ann. Am. Acad. Pol. Soc. Sci. 618 (1), 80–94.
Huisman, Mark, Steglich, Christian, 2008. Treatment of non-response in longitudinal network studies. Soc. Netw. 30 (4), 297–308.
van der Hulst, Renée C., 2011. Terrorist Networks: The Threat of Connectivity. The SAGE Handbook of Social Network Analysis, pp. 256–270, 2011.
Krause, Robert W., Huisman, Mark, Steglich, Christian, Sniiders, Tom A.B., 2018. Missing network data a comparison of different imputation methods. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). Barcelona: IEEE, pp. 159–163 (Accessed 30 December 2018). https://ieeexplore.ieee.org/document/8508716/.
Krebs, Valdis, 2002. Uncloaking terrorist networks. First Monday 7 (4) (Accessed 1 April 2016). http://firstmonday.org/ojs/index.php/fm/article/view/941.
Laumann, Edward, Marsden, Peter, Prensky, David, 1983. The boundary specification problem in network analysis. Appl. Netw. Anal.: A Methodol. Introd. 61, 18–34.
Lazega, Emmanuel, Snijders, Tom A.B. (Eds.), 2016. Multilevel Network Analysis for the Social Sciences. Springer International Publishing, Cham (Accessed 18 April 2016). http://link.springer.com/10.1007/978-3-319-24520-1.
Lubbers, Miranda J., Snijders, Tom A.B., 2007. A comparison of various approaches to the exponential random graph model: a reanalysis of 102 student networks in school classes. Soc. Netw. 29 (4), 489–507.
Mancuso, Marina, 2014. Not all madams have a central role: analysis of a nigerian sex trafficking network. Trends Organ. Crime 17 (1/2), 66–88.
Morselli, C., 2009. Inside Criminal Networks. Studies of Organized Crime, vol. 8. Springer New York, New York, NY.
Morselli, C., 2014. Crime and Networks. Routledge, New York.

Morselli, C., Roy, J., 2008. Brokerage qualifications in ringing operations*. Criminology 46 (1), 71–98.
Morselli, Carlo, Giguere, Cynthia, 2006. "Legitimate strengths in criminal networks." crime. Rev. Law Soc. Change 45 (3), 185–200.
Morselli, Carlo, Petit, Katia, 2007. Law-enforcement disruption of a drug importation network. Glob. Crime 8 (2), 109–130.
van Nassau, Casper S., Diviák, Tomáš, de Poot, Christianne J., Tubergen, Frankvan, 2019. Explaining Tie Formation in Salafi-Jihadi Networks Operating in Western Contexts. Paper presented at the EUSN 2019, Zurich.
Oliver, Kathrine, Crossley, Nick, Everett, Martin G., Edwards, Gemma, Koskinen, Johan, 2014. Covert Networks: Structures, Processes and Types. The Mitchell Center for Social Network Analysis Working Paper. http://www.socialsciences.manchester.ac.uk/medialibrary/research/mitchell/covertnetworks/wp/working_paper1.pdf.
Ouellet, Marie, Bouchard, Martin, 2018. The 40 members of the Toronto 18: group boundaries and the analysis of illicit networks. Deviant Behav. 39 (11), 1467–1482.
Pattison, Philippa E., Robins, Garry L., Snijders, Tom A.B., Wang, Peng, 2013. Conditional estimation of exponential random graph models from snowball sampling designs. J. Math. Psychol. 57 (6), 284–296.
Configurational comparative methods: qualitative comparative analysis (QCA) and related techniques. In: Rihoux, Benoît, Ragin, Charles C. (Eds.), 2009. Applied Social Research Methods Series, 51. Sage, Thousand Oaks.
Robins, Garry, 2009. Understanding individual behaviors within covert networks: the interplay of individual qualities, psychological predispositions, and network effects. Trends Organ. Crime 12 (2), 166–187.
Robins, Garry, 2013. A tutorial on methods for the modeling and analysis of social network data. J. Math. Psychol. 57 (6), 261–274. Social Networks.
Robins, Garry, Pattison, Philippa, Woolcock, Jodie, 2004. Missing data in networks: exponential random graph (P*) models for networks with non-respondents. Soc. Netw. 26 (3), 257–283.
Rostami, Amir, Mondani, Hernan, 2015. The complexity of crime network data: a case study of its consequences for crime control and the study of networks. Thomas Niederkrotenthaler, ed. PLoS One 10 (3), e0119309.
Rubin, Donald B., 1976. Inference and missing data. Biometrika 63 (3), 581–592.
Sageman, Marc, 2004. Understanding Terror Networks, 1st edition edition. University of Pennsylvania Press, Philadelphia.
Smith, Chris M., Papachristos, Andrew V., 2016. Trust thy crooked neighbor multiplexity in Chicago organized crime networks. Am. Sociol. Rev. 81 (4), 617–643.
Smith, Jeffrey A., Moody, James, 2013. Structural effects of network sampling coverage I: nodes missing at random. Soc. Netw. 35 (4), 652–668.
Smith, Jeffrey A., Moody, James, Morgan, Jonathan H., 2017. Network sampling coverage II: the effect of non-random missing data on network measurement. Soc. Netw. 48, 78–99.
Snijders, Tom A.B., van de Bunt, Gerhard G., Steglich, Christian E.G., 2010. Introduction to stochastic actor-based models for network dynamics. Soc. Netw. 32 (1), 44–60. Dynamics of Social Networks.
Snijders, Tom A.B., 2011. Statistical models for social networks. Annu. Rev. Sociol. 37 (1), 131–153.
Sparrow, Malcolm K., 1991. The application of network analysis to criminal intelligence: an assessment of the prospects. Soc. Netw. 13 (3), 251–274.
Steglich, Christian, Snijders, Tom A.B., Pearson, Michael, 2010. Dynamic networks and behavior: separating selection from influence. Sociol. Methodol. 40 (1), 329–393.
Volk, Anthony A., Veenstra, René, Espelage, Dorothy L., 2017. So you want to study bullying? Recommendations to enhance the validity, transparency, and compatibility of bullying research. Aggress. Violent Behav. 36, 34–43.
Wang, Peng, Pattison, Philippa, Robins, Garry, 2013. Exponential random graph model specifications for bipartite networks—a dependence hierarchy. Soc. Netw. 35 (2), 211–222.
Wikström, Per-Olof H., Bouhana, Noémie, 2017. Analyzing radicalization and terrorism: a situational action theory. In: LaFree, Gary, Freilich, Joshua D. (Eds.), The Handbook of the Criminology of Terrorism. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 175–186. https://doi.org/10.1002/9781118923986.ch11 (Accessed 14 September 2018).