Studying organized crime networks: data sources, boundaries and the limits of structural measures

Dr Paolo Campana
Institute of Criminology
University of Cambridge

Prof Federico Varese
Department of Sociology
University of Oxford

Corresponding author:
Paolo Campana, Institute of Criminology, University of Cambridge
Sidgwick Avenue, Cambridge CB3 9DA, United Kingdom
Phone: +44 1223 767375 Email: pc524@cam.ac.uk

Acknowledgments

We are extremely grateful to Thames Valley Police for sharing their data with us. In particular, we wish to express our sincere gratitude to Chief Inspector Marc Tarbit and Data Analyst Alastair Nall for their crucial support. We are also grateful to the Editor, to three anonymous Reviewers and to Nynke Niezink for their most helpful comments on a previous version of the paper. We are happy to acknowledge the financial support received from Leverhulme Trust through the Research Grant RPG-2018-119. Authors are listed in alphabetical order.

Accepted Manuscript
Acceptance date: 19/3/2020
Network studies of organized crime (OC) normally explore two key relational issues: the internal structure of groups and the interactions among groups. The paper first discusses in depth two data sources that have been used to address these questions -- phone wiretaps and police-generated “events”-- and reviews issues of validity, reliability and sampling. Next, it discusses challenges related to OC network data in general, focusing on the ‘double boundary specification’ problem and the time span of data collection. We conclude by arguing that structural analysis cannot be divorced from a deep contextual (qualitative) knowledge of the cases. The paper refers to concrete research dilemmas and solutions faced by scholars, including ourselves.

1. Introduction: the network study of organized crime

The study of organized crime is plagued by a plethora of terms that refer to its object of study. These terms include ‘gangs’ (Thrasher 1927), ‘crews’ (Adler 1993), ‘firms’ (Reuter 1983), ‘criminal enterprises’ (Smith 1980; Haller 1990), ‘criminal networks’ (Morselli 2009) and ‘mafias’ (Gambetta 1993; Varese 2001) to name a few (in this paper, we use the term ‘organized crime group [OCG]’). Some expressions, such as ‘transnational organized crime’
and ‘serious crime’, are linked to policy agenda and vested interests, as argued by Varese (2011). Regardless of the term used, organized crime has important relational features. Since at least the 1990s, scholars have focused on two issues, the internal structure of groups and the connections between groups in a given illegal market (for reviews, see Bichler et al. 2017 and Faust and Tita 2019). As for the nature of the evidence used, social network analysis (SNA) contributions rely on two types of data: court files containing extensive transcripts of actors’ phone conversations and gangs’ own accounts, and police-generated ‘events’, such anonymised police arrest records capturing whether two gang members have been arrested together in a specific instance. The aim of this paper is first to review in depth the strength and weakness of these two sources - namely court files containing wiretaps and police-generated events - and assess how they can be best used to answers research questions on hierarchies and market activities of OC groups (Sections 2 and 3). Section 4 addresses two challenges related to all OC network data, namely the ‘double boundary specification’ problem and the time span of data collection, while section 5 discusses the role of qualitative knowledge for the study of OC networks. Section 6 presents the conclusions.

2. Data sources: Phone wiretaps

Wiretapped conversations provide a rich, relational data source. By definition, a conversation is between two or more people. Usually, such conversations can be found in court records. In several jurisdictions (e.g., the United States, Canada, Germany, France, Italy, the Netherlands and Sweden), wiretaps can be introduced as evidence in trials. They are transcribed and made available to the parties. Wiretapped conversations capture what actors talk about in their day-to-day business life, including conversations involving lower-level and upper-level actors, both within and beyond the gang. Scholars have used this source systematically since the early 1990s (Baker and Faulkner 1993; Finckenauer and Waring 1998; Natarajan 2000; 2006; Morselli 2009; Varese 2011 and 2013; Campana 2011; Calderoni 2012; Campana and Varese 2013). Below is an example of a phone transcript included in an Italian court file related to a human trafficking case (anonymised excerpt, see Campana 2016a for details of the case):

ID Conversation: 841. Type: Outbound
Date: 11.06.2007 Time: 17:38
 [...]

Trafficker 1: It is better not to show her [the madam] the girl until she has paid the money. You first have to get the money, try to get it right.

Trafficker 2: She is alone.

Trafficker 1: I know. If she wants the bingo [the girl], she has to pay first. Maybe she does not have the money with her. Tell her she has to pay first, try not to make the mistake of showing the girl first. When she sees the girl, she will try to create trouble. Try not to get there. You have to get the money first.

[...]

This transcript includes information on the two individuals involved in the call, the direction of the call, the time and date (some transcripts also provide information about the location of the phone at the time of the call). Furthermore, it provides evidence on the reason behind the call, in this case the ‘delivery’ of a trafficked victim to an exploiter (the ‘madam’). The conversation also points to a lack of trust between transporters and exploiters and – crucially – suggests that transporters and exploiters are separate and independent (Campana 2016a: 81). While even this brief example shows the richness of data, issues of validity do arise, as discussed below.

2.1. Validity and reliability

Three issues are particularly relevant when using phone wiretaps: self-censorship, group coverage and the link between conversations and behaviour.¹

*Self-censorship.* For a set of conversations to be suitable for analysis, the actors involved should have talked freely about all (or most of) the activities of a group (the use of encrypted language should not be confused with self-censorship about the topics discussed). Two different control strategies can be followed to assess the extent of self-censorship: internal and external. The internal validity control is based on the content of conversations. One can assume that, if the actors had been aware of being listened to, they would have avoided discussing self-incriminating evidence, at least most of the time. If, on the other hand, this is not the case and actors do talk on the phone about incriminating actions such as the use of violence and murders, we can be more confident about the validity of the transcripts. The external control entails double-checking the information included in the conversations with data collected from other official records related to the case and/or individuals, open source
analysis and interviews with key informants, particularly investigators (Campana and Varese 2012; Varese 2011).

*Group coverage.* Key individuals should have been put under surveillance to gain a satisfactory coverage of both the actors and activities of a group. If the police have badly missed their surveillance targets, people who might be central will not appear so in the data simply because their phones have not been monitored (Natarajan, 2000: 293; Klerks, 2001: 58; Morselli, 2009: 49). In addition, actors tend to be added to the list of suspects as the investigation proceeds, so that the group comes to look significantly different over time as a function of police decisions. The degree of coverage is notoriously difficult to assess. However, interviews with prosecutors and police officers involved in the investigation often give a good sense of potential biases and limitations and help the researcher assess the impact of time constraints, resource limitations and legal impediments on the investigation (this strategy was adopted by Varese 2011 and 2013; and Campana 2011 and 2018). Additional documentation is often available to ascertain when names and lines have been added or removed during an investigation. Furthermore, a preliminary analysis of the time stamp associated to each conversation, if available, should uncover a set of actors/phones that is relatively stable across the period under consideration (if necessary, researchers can trim the period under consideration as in Varese 2013 and disregard the initial period during which police were still calibrating the perimeter).

*Conversations and behaviour.* To what extent can conversations be considered a proxy for behaviour (actual or potential)? Schlegel (1984: 107) maintains that conversations among criminals “often involve lies, boasts and exaggerations which may not reflect the true nature of crime” (see also Smith 1975: 297). To some extent, this is true also of data collected through a standard questionnaire or less structured interviews. The advantage of wiretaps and court records is that normally outright lies can be exposed by reference to other data collected by the police or following a court case, thus producing a measure of construct validity not available to most researchers interviewing subjects through a standard questionnaire. Not only can the police validate the content of some conversations by tailing individuals and checking their physical and financial movements, but also actors themselves often try in their conversations to ascertain whether the information conveyed by a fellow conspirator is accurate. The larger the sample and the longer the period of surveillance, the more likely it is that one would be able to see whether some lies are eventually exposed. Furthermore, we
believe that a criminal group in which everybody constantly tells only lies to everybody else is doomed to fail from the start. Granted, members of OCGs do have an interest in omitting when reporting to the boss, say, how much money they made in a given deal. However, a discussion of the deal would nonetheless surface in the conversations. Further, not all the lies have the same impact on the analysis. For instance, let us assume that a member of the group deliberately gives the boss some wrong information about his actual behaviour while reporting on a task he was entrusted with. Lies and exaggerations, in this case, have a negligible impact on the analysis as the task (e.g. extortion) is nonetheless mentioned. As a general rule, the broader the topics coded, the smaller the impact of lies and exaggerations have on the results; likewise, the wider the set of conversations and actors wiretapped, the smaller the risk of misleading interpretations (Campana and Varese 2012).

2.2. Sampling

Normally, a researcher has no input in designing the collection strategy of police data and court files. The researcher needs to work within the boundaries set by the agency that has collected the data for purposes that are unrelated to his/her goals. The secondary nature of the evidence and the inability of a researcher to actively take part in the collection process make it difficult to control for errors in the data and for missing data (Malm et al. 2008). Another implication of the secondary nature of the data is that researchers have little say on the sampling strategy adopted. Yet, sampling has a potential impact on the outcome of the analysis and thus warrants some further consideration.

In Campana and Varese (2012), we discuss issues related to using samples of conversations wiretapped as opposed to the universe of the conversations or meta-data about these conversations that will not include any additional evidence on the content. Here we focus on the issue of how criminal files are built, namely the selection of individuals included. When building an investigation, law-enforcement agencies normally start with some seed individuals and then expand their reach by adding further individuals, some of whom may be connected to the seed individuals (Interview 2; Interview 3). We can define this data collection strategy as similar to a snowball sampling with elements of purposive sampling. In
other words, not all the individuals linked to the seed individuals are automatically added. Police officers do make decisions in this regard.

In principle, the data collection strategy adopted by law enforcement to build criminal cases is well suited for network analysis. As maintained by Robins (2015: 56), “snowball sampling is usually the preferred approach for investigating network structure using sampled data.” For network data, random sampling is far from being the golden rule as it can be shown to hold extremely high chances of generating distorted inferences about network structures (Robins 2015: 11). Robins (2015: 115) notes that snowball sampling designs are “often valuable for hard to reach or hidden populations.” Crime network data are often constructed by police adopting a purposive element in the sampling strategy: this is - at the same time - an extremely valuable strategy and a source of potentially serious biases. Have the police included the right individuals? Have they followed the right leads? In Frank’s (2005) terminology, we have two sets of decisions that can potentially impact our data: vertex sampling (nodes) and edge sampling. We should ask whether the right seed individuals have been targeted and the relevant additional individuals have been included as an investigation progressed. The data collection strategy adopted by the police in their investigations can be seen as a specific type of link-tracing design in which referrals are unwittingly provided by the respondents. Heckathorn and Cameron (2017: 104-106) maintain that such strategy has proved to be more robust to biases than initially thought (Frank 2005 makes a similar point). Besides quantitative metrics, however, we believe that researchers should also leverage on qualitative evidence to assess the robustness of a sample. A qualitative assessment conducted through interviews with prosecutors and police officers involved in the investigation is crucial. As part of a researcher’s data collection strategy, it is vital to conduct interviews with those involved in the primary data collection. If available, additional qualitative evidence, for instance from reports, news items and other open sources, can be utilised to probe potential biases in the sampling strategy. External sources are preferable for a fully-fledged triangulation.

When building networks from court files and police investigations, researchers are confronted with a further issue: the fact that some actors have been put under surveillance while others have not. Thus, the likelihood of appearing in the network is not the same for all actors (Campana and Varese 2012; Bright et al. 2019: 243; Diviak 2019 in this issue). This is similar to the ‘spotlight effect’ discussed by Smith and Papachristos (2016); in their case, the
spotlight effect was due to different levels of media coverage warranted to different individuals, which in turn resulted in missing ties potentially not being randomly distributed. This limitation should be made transparent and openly discussed and researchers can adapt the metrics used in order to minimise in-built biases. For instance, instead of relying on valued networks, researchers might calculate only binary measures to minimise inclusion biases. This was the strategy adopted by Campana (2018) in extracting contacts among smugglers from court files. He created two separate networks: a smaller valued network of 28 actors for whom the full set of phone conversations was available and a broader binary network of 292 actors who have appeared in the wider court file but were not all directly targeted by the police. The latter was then analysed vis-à-vis the former as a strategy to check the robustness of the findings (all the regression coefficients pointed in the same direction). Keeping the broader network in the analysis had a number of benefits, most importantly the fact that it covered a much larger part of the smuggling process than the smaller network of 28 actors. Varese (2013: 906) presents a double analysis of, respectively, the set of 16 actors under direct surveillance from the Italian police and the broader set of 164 actors with whom the 16 seed individuals have been in contact. The structure of the network - when seen through a faction analysis - remains unchanged and so is the relative central position of top actors. In both cases, the analyses were robust to the impact of potential missing data present in the broader data sets. Smith and Papachristos (2016: 660-661) contrasted the results for the overall network with those calculated for four sub-networks extracted based on the degree distribution of the actors involved in the full network (N = 1,030): the top 1% of actors (N = 10), the top 10% (N = 108), the bottom 90% (N = 922) and the bottom 99% (N = 1,020). They found that, broadly speaking, the results held across all sub-networks, i.e. the analysis, was robust to the potential spotlight effect.

3. Data sources: Police-Recorded events

Besides wiretaps, researchers can rely on police records to build their networks (Bright et al. 2012: 154; Hashimi et al. 2016: 3; Ouellet et al. 2019: 13; Charette and Papachristos 2017). While generated by the same law enforcement agencies, these records take a different form from wiretaps. They are normally recorded as a list of events logged by the police. For each event, police forces normally keep a record of the type of crime associated (‘activity’ in our terminology) as well as the place where the crime took place and the time the crime occurred (if known). Furthermore, for each event both offender(s), if known, and victim(s) are
identified. Normally, police forces also keep a record of certain characteristics of offenders and victims, such as gender, date of birth, age, nationality, ethnicity, address, sometime occupation, as well as their involvement in an OCG/gang. Affiliation to a specific OC group is normally based on intelligence collected by the police and in some cases is shared with researchers in anonymised form. It is normally very difficult for academic researchers to go beyond such a level of access to intelligence data.

While the structure of the data is similar across jurisdictions (a two-mode network, event by actor), the additional information recorded for each event/actor is context specific. In some cases, the events included in a data set are limited to instances where the OCG member was arrested. Other data sets go further and include instances in which an offender was a suspect, has been subject to a stop-and-search, has been the driver of a vehicle that has been stop-checked, has been issued with caution or has been located in relation to an investigation (see, e.g., Baika and Campana 2020). Police data might also include events that resulted in no further action. The list of events can vary depending on the jurisdiction and/or the operational procedure of a given police force. Researchers should always fully describe the type of events included in their analysis and carefully reflect on which social mechanisms these events can shed light on – and which ones they cannot (see also Faust and Tita 2019: 107).

As mentioned above, studies of gangs have used anonymised police arrest records. The formal network analysis of co-arrest data builds on an earlier intuition by Reiss (1988) and Reiss and Farrington (1991), and further developed by Carrington (2002; more recently, see also McGloin et al. 2008; Grund and Morselli 2017). While such data have allowed scholars to shift their focus away from the single individual by including – implicitly or explicitly – an element of coordination, they are limited by the fact that not every offence leads to arrest (Hughes 2005) or other forms of police action. A handful of studies have gone beyond arrest data by relying on police field intelligence’s observation cards recording non-criminal encounters with the police (Papachristos et al. 2012). Observation data typically record situations in which two or more individuals are observed in each other’s presence, i.e. being in the same place at the same time (Papachristos et al. 2012: 994-995). They are akin to field observations in ethnography, although much less rich in detail (and with no control from the researcher). Compared to police records, intelligence observations are normally richer, more granular and potentially able to cover a broader spectrum of interactions, both criminal and non-criminal. However, they typically come with two major drawbacks. Firstly, they generate
a large volume of unstructured (textual) data that requires time-consuming manual coding. Secondly, researchers might not be in a position to carry out such coding, as intelligence records are usually classified to a higher level of secrecy than (simple) police records due the likely presence of sensitive information used in on-going investigations, as well as personal non-criminal information. Data sharing can be further hampered by the difficulties in anonymizing such logs. Our experience suggests that police forces are more willing to share anonymized police-recorded events rather than unstructured intelligence logs.

While police records normally do not possess the same level of detail as wiretaps and thus are less suited to capture the internal structure of the organization, they are valuable in placing groups and actors within their criminal market(s). Police records can help researchers study the criminal interaction between a given OCG member and other offenders in relation to a specific activity, for example drug dealing. In this paper, we consider only the case in which a researcher wishes to project a two-mode offenders-by-events network into a 1-mode offender-to-offender network and work with the latter. It should be noted here that police-recorded events normally produce an undirected and weighted 1-mode projection.

To illustrate this source of data, we draw on the Thames Valley Police Data Set, which we have recently acquired and are in the process of analysing. Thames Valley Police is one of the largest territorial police forces in England and Wales, covering a population of over 2 million people located in the South East of England and serving cities such as Oxford, Reading, Milton Keynes and Slough, as well as rural areas. The data set includes police records for a five-year period. Appendix A contains a fuller description of the data set. Figure 1 presents a depiction of the interactions between OC members and non-OC members in relation to drug supply. By interaction we refer to any type of co-participation in a crime event as recorded by the police (see also Appendix A).

Figure 1. Drug-supply activities, Thames Valley, 2011-2016
In Figure 1 the grey dots refer to OCG members, while the white dots represent non-members (isolates are excluded from the picture). The network is generated using OCG members as seed individuals (see section on sampling below for details). The picture is a 1-mode undirected projection of a 2-mode offenders-by-events data set. The Figure shows that OCG members collaborate significantly with non-members when operating in the supply side of a drug market. It also shows a tendency towards a clusterisation of the market (190 components; the largest includes 45 actors). Below, we shall return to this dataset to highlight some key issues related to validity and reliability.

3.1. Validity and reliability

Data sets of police-recorded events also raise issues of validity and reliability. For instance, they can be influenced by the level of enforcement and changing policing priorities, as well as resource constraints (Morselli 2009: ch. 2; Malm and Bichler 2011; Campana and Varese 2012; Calderoni 2014; Broadfield and Marshall 2017; Baika and Campana 2020). Furthermore, police enforcement can be selective: some individuals might be targeted more than others depending on race and/or place of residence (Black 1970). As all official data, they are also affected by changes in recording practices. It is crucial for a researcher to fully understand the context and the recording practices of the agency that has generated the data.
Shifts in police priorities, changes in recoding practices and other potential biases can be ascertained through interviews with those who have generated the data. In many instances, we have found these interviews extremely valuable; they helped us understand which research questions can be reliably answered using a specific batch of data, and which cannot.

Furthermore, researchers should pay special attention to the risk of double counting. In some cases, the same event is entered multiple times, for instance when the status of an offender changes from, e.g., ‘suspect’ to ‘arrest’. If a unique event ID is used across the entire data set, this issue can be easily detected and fixed. If a different event ID is generated each time there is a change in status, or the event ID has been anonymised in such a way that its ability to identify unique events has been lost, things become difficult, and researchers should carry out specific checks, such as using offender IDs and the information on date/place, before building the network. Material errors that might have occurred when inputting the data, including different spelling of the same name or errors in recording the date of birth or addresses, are an additional challenge. Police records can be compiled late at night, early in the morning or under a heavy workload. While there is no magic wand to fix these issues, a careful scrutiny of the data through descriptive statistics can go a long way to detect material errors.

As already mentioned, some police records that are shared with researchers are limited to arrests and charges while others include a broader set of encounters with the police, including instances in which an offender was just a suspect or encounters that have led to no further action. The Thames Valley Police Data Set allow us to evaluate the changes in the network characteristics when considering arrests and charges only, as opposed to all events (Table 1).

Table 1. All encounters vs. arrests and charges only, Thames Valley Data Set, 2011-2016

<table>
<thead>
<tr>
<th></th>
<th>1-year</th>
<th></th>
<th>5-year</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Events</td>
<td>Arrested + Charged only</td>
<td>All Events</td>
<td>Arrested + Charged only</td>
</tr>
<tr>
<td>Nodes</td>
<td>243</td>
<td>95</td>
<td>745</td>
<td>472</td>
</tr>
<tr>
<td>Ties</td>
<td>154</td>
<td>12</td>
<td>11,446</td>
<td>6,774</td>
</tr>
<tr>
<td>Density</td>
<td>0.003</td>
<td>0.001</td>
<td>0.021</td>
<td>0.030</td>
</tr>
<tr>
<td>Average degree</td>
<td>0.412</td>
<td>0.105</td>
<td>2.40</td>
<td>1.151</td>
</tr>
<tr>
<td>Diameter</td>
<td>2</td>
<td>1</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>Components</td>
<td>209</td>
<td>90</td>
<td>351</td>
<td>368</td>
</tr>
<tr>
<td>Connectedness</td>
<td>0.002</td>
<td>0.001</td>
<td>0.025</td>
<td>0.003</td>
</tr>
<tr>
<td>N events</td>
<td>661</td>
<td>173</td>
<td>7,355</td>
<td>1,687</td>
</tr>
</tbody>
</table>

*Source:* Thames Valley Police Data Set.

Notes: measures calculated on the 1-mode projection of the 2-mode actor-by-event matrix. N events in the original 2-mode matrix as indicated in the table. The analysis excludes sexual offences, domestic incidents, shoplifting and traffic offences.

Focusing on events that have led to an arrest and/or a charge against an individual yields a loss of structure. If we compare the networks calculated over a one-year period, the number of OCG members drops from 243 to 95 and the number of ties (co-offending) drops from 154 to 12. The diameters halve and the average degree decreases from 0.412 to 0.105. An already disconnected network becomes hardly discernible. Over a five-year period, we observe a similar trend: a 37% drop in the number of OCG members included in the network, a 41% drop in the number of ties and a considerable reduction in diameter of the network (from 18 to 3). The average degree more than halves and the connectedness almost disappears (-88%).

The above issue is related to whom we should consider an offender. Formally, only those convicted of an offence should be considered as such. However, police data do not include such information; at the same time, they are most likely to include also suspects and people who have been arrested but subsequently not convicted. The empirical trade-off of such a decision emerges from Table 1. Morselli (2009) also conducts a similar comparison. The largest network analysed by Morselli (2009: 45) includes 318 individuals monitored, 110 targeted, 25 arrested, 22 accused and 14 convicted. Morselli (2009: 47) suggests that the “targeting stage” is the best in terms of trade-off. We believe that, as long as the names of the individuals are concealed and no other information can lead to the identification of suspects, whom to include in the analysis is ultimately the researchers’ decision, which must be made explicit and reflect the research question (and possibly go beyond co-arrest data).

3.2. *Sampling*

When extracting data from police databases in order to identify relevant events, researchers can follow two procedures: full extraction and targeted extraction. The first method generates a complete network in so far as it is captured by police records while the second generates a set of (most likely) overlapping ego-networks. Therefore, different statistical techniques can be reliably applied to each method and different research questions can be answered.
Researchers thus need to have a full understanding of how data were extracted and prepared for them. Again, interviews with the officer(s) who prepared the data are invaluable. We now discuss both strategies.

**Full extraction.** This strategy consists in setting a time window and then extract all the events that have been recorded between time $t$ and $t+1$. Besides a time frame, full extraction methods can also specify an area, such as a neighbourhood or a city, and then extract all the incidents based on the location. This was a strategy adopted by Bastomski et al. (2017) and Charette and Papachristos (2017). The advantage of this strategy is the production of a universe of contacts, as seen through the lenses of the police, without any need for seed individuals. The drawback is that, if the activities of a group span across multiple neighbourhood or cities, this will not be captured in the data. Furthermore, given the sheer amount of data, a researcher is often bound to either choose small locations or a sub-set of offences, e.g. gun-related offences or homicide (for a discussion of the issues related to the choice of time spans, see Section 4). As noted by Charette and Papachristos (2017: 5), “attempts to analyze the entire network led to a computational failure; even on high-power computing clusters, our analysis requiring too much computer memory”. Their solution was to randomly select subsets of dyads. However, sampling from network data poses two important risks: key actors can be missed out (Robins 2015: 11); network dependencies can be lost.6 Further developments in computer technology might alleviate this problem in the future.7

**Targeted extraction.** This strategy consists in selecting a set of actors (seeds), identify the events they were connected to and then extract all the alters connected to such events.8 This is akin to a multi-stage sampling design (Robins 2015: 72-72 and 114-115). Such approach has been followed, for instance, by Ouellet et al. (2019). Their starting point has been a convenience sample of 261 “seed” individuals who have been identified by the police as street gang members (Ouellet et al. 2019: 12). Next, they added the people in direct contact with the “seeds” as recorded in the police database and then a further layer of contacts. Our Thames Valley Police Data Set was also constructed starting from a list of individuals identified by the police as OC members (our “seed” individuals) and then added the first-degree contracts. When dealing with two-stage sampling designs, it is crucial to specify whether the contacts between first-degree contacts were directly extracted or not as this has a direct implication for the analysis (Robins 2015: 70-72). If such contacts were not extracted,
we inevitably have missing data and the structure of the data is better described as a collection of (overlapping) ego-networks (equal to the number of seed individuals). If all contacts between first-degree actors are also included, then we have a fully-fledged complete network (as captured by police records). Furthermore, it is important to note that, in a two-mode design, some ties between first-degree actors will be included even if they were not directly extracted. To illustrate this, let us imagine a set of three offenders: A, B and C. A is an OC member while B and C are not. A is our “seed” individual. Let us say that A, B and C have committed an offence together: in the one-mode projection of the network, this will result in a direct link between B and C (by virtue of them co-offending with A). Let us now assume that B and C have also committed a second offence, this time without A. This additional link will not be included in the network unless it was directly extracted during the preparation of the data.

4. Challenges with OC Network Data

We have discussed issues of validity and reliability that pertain to two specific data sources, police wiretaps and police-generated events. We now focus on issues that are common to all SNA OC data, namely (a) the identification of groups’ and markets’ boundaries and (b) time span of data collection. In the context of OC, the former gives rise to what we call the “double boundary specification problem” (for a general discussion on defining boundaries in SNA, see Laumann et al. 1989; Klerks, 2001: 58; Borgatti and Nagin 2011: 1169-1170; Robins 2015).

4.1. The ‘double boundary specification’ problem

In the context of OC, by definition researchers need to engage with the (challenging) task of assigning actors to organizations (OCGs). How can we attribute a specific individual to a given OCG? Scholars can adopt different strategies: 1) membership list provided by the gang itself; 2) reputational approach; 3) offence-based membership; 4) event-based attribution and 5) community detection algorithms. Below we discuss them in turn.

List provided by gang. In some rare cases, groups produce a list of members and the police acquire the document, as in the case of the Hells Angels studied by Morselli et al. (2017). Such instances are rare and also raise questions on how complete the records are. For those
OCGs that perform affiliation ceremonies, the participation in such ceremonies might provide a criteria to establish membership. However, not all groups perform such ceremonies and those that do keep participation a closely guarded secret.

Reputational approach. A second strategy relies on attributing membership based on a reputational approach (Laumann et al. 1989), namely on the basis of the judgment of key informants. The informants are normally police officers whose task is to assign people to OCGs based on confidential intelligence and, whenever available, groups’ self-identification signs (Interview 1; also Papachristos et al. 2015a). Police forces may routinely create – and update – such lists.9

Offence-based membership. In some cases, membership information is crystallised in the form of a specific offence. For instance, the Italian penal code includes a separate offence for mafia activities and it can be used as a proxy for actual membership. Yet usual caveats apply in reference to distortions built in convictions’ data. In addition, prosecutors operate under a specific incentive structure: sometimes, it might be difficult to prove a mafia-type crime, so prosecutors prefer to charge defendants with a serious crime, carrying a heavy prison sentence nonetheless, but dropping the additional charge of mafia activity. In other cases, prosecutors might charge individuals with a mafia-affiliation offence even when they are not fully-fledged members. Furthermore, courts could accept or reject prosecutors’ arguments on the basis of what can be reasonably proved in court, which might well differ from actual membership. Finally, legal reasoning changes over time and across places.

Event-based attribution. A fourth strategy is to rely on an event-based attribution, for instance by identifying a list of events distinguishing OCGs from non-OCGs. This list is necessarily theory-driven and highly dependent on the OC framework adopted by the researcher (see Campana and Varese 2018 for a discussion and some illustrative examples). However, such a list might be empirically difficult to compile. Hashimi et al. (2016) shows some of the challenges that researchers face when trying to classify incidents as OC-related. The proportion of incidents classified as OC-related ranges from 22.8% to 0.3%, based on how the criteria for inclusion of the offences are specified (Hashimi et al. 2016: 8; these authors relied on police data from the city of Montreal, 2005-2009). Furthermore, not all participants to a given event are fully-fledged members of an OCG. In other words, OCG members can
cooperate with non-members in relation to specific events or activities (for examples, see Campana 2011, Varese 2013, Morselli et al. 2017, Figure 1 above).

**Community detection algorithms.** In some cases, researchers have utilised a statistical solution to the problem, usually by relying on community detection algorithms. Yet, in the absence of qualitative information on OCG membership, such algorithms might generate artificial results that are unrealistic, such as OCGs larger than 100 individuals operating in rather small settings. Community detection algorithms, combined with a list of fully pre-identified OC members, like the “seed” individuals in the case of Ouellet et al. (2019: 14), may mitigate some of these issues; however, the extent of erroneous identifications may still be non-negligible. Errors might come in two forms: the creation of wholly artificial OCGs that do not exist in reality and the misclassification of individual actors as OCG members. Using the Thames Valley Police Data Set, we now compare the number and size of OC groups based on police intelligence with those generated by two widely used community detection algorithms, the Louvain algorithm and the Girvan-Newman algorithm (Table 3).

<table>
<thead>
<tr>
<th></th>
<th>Police intelligence (human attribution)</th>
<th>Louvain algorithm</th>
<th>Girvan-Newman algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. groups (communities)</td>
<td>153</td>
<td>362</td>
<td>352</td>
</tr>
<tr>
<td>Average size</td>
<td>4.9</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Size of the largest group</td>
<td>21</td>
<td>23</td>
<td>65</td>
</tr>
<tr>
<td>N. groups =&gt; 2</td>
<td>135</td>
<td>103</td>
<td>92</td>
</tr>
</tbody>
</table>
According to police intelligence, there are a total of 153 identified OCGs, of which 135 have two or more members and 107 have three or more.\(^{10}\) Both algorithms are rather consistent in their classification and tend to overestimate the overall number of groups and underestimate the number of groups with three or more people, i.e. the more “mature” groups. Additionally, the Girvan-Newman algorithm gives a rather unrealistic size for the largest group (65) compared to the Police intelligence (21) and Louvain algorithm (23). The average size of groups is 4.9 members (police-identified), 2.1 members (Louvain) and 2.1 (Girvan-Newman). Algorithms seem to suggest higher fragmentation and, on average, a smaller scale of OCG groups when compared to intelligence-based attribution. While we acknowledge that it might not always be possible to access police-based membership attribution, we believe that a reputational approach based on the judgement of key informants remains a superior strategy as it often taps into a wealth of information from sources of different nature that can hardly be modelled formally. Faust and Tita (2019: 113) offer indirect support to this conclusion when stating that, in relation to the collection of intergang network data to study violence, “expert’s perceptions are generally sound”. Following Decker and Pyrooz (2010), they also conclude that, “it is likely that experts in the field have a good perception of ties among gangs” (Faust and Tita 2019: 114). We believe this can be extended to group identification.

The second challenge relates to drawing the boundaries of the market(s) in which criminal organizations operate. Let us take, for example, the cocaine drug market. One can have OCGs based, say, in Oxford or Cambridge, involved with local non-affiliated dealers. At the same time, such groups would buy their cocaine from another group based in London, which, in turn, deals with a wholesale dealer based in Amsterdam. Such a dealer may be connected to a broker in Spain who, in turn, is linked to a Colombian-based wholesale trafficker, and so on until we reach the producers located in a remote village in the Andes. At each stage, actors are linked with a myriad of other actors up and down the supply chain, as well as orthogonally to the chain. The London-based group might deal drugs with groups A, B and C while carrying out arson attacks with individual X and Y and buying weapons from Z. Links can also be established with facilitators, such as corrupt officers. This simple example illustrates the complexity of the intertwined markets and the impossibility of mapping the universe of contacts. This issue relates to a broader critique of SNA, articulated, for instance,
by Tom Naylor at a seminar given by Carlo Morselli that he attended: “a social network […] could include everyone,” Naylor noted (Morselli 2009: 3; see also Fisher 1977: 33-34). We need a stopping rule that is consistent with specific research questions (Morselli 2009: 4).

Court file and police investigations can, to a varying degree, map a portion of a much broader market. Inevitably, the boundaries of such portions lie where the investigation ends. Such a pragmatic approach was followed, among others, by Varese (2013) to study the operation of a Russian mafia group in Rome; by Calderoni (2012) in his study of the involvement of the ‘Ndrangheta, in the trafficking of cocaine; and by Bright et al. (2019) in their study of a methamphetamine trafficking ring. Court files and police investigation are better at mapping interactions along a chain. This was the strategy adopted by Campana (2018) when studying smuggling activities from the Horn of Africa to Northern Europe via Libya and Italy. The boundaries of the network were, again, those set by the investigation. As a general rule, one can accept the boundaries of an investigation as a stopping rule provided that such investigation has met the requirement of adequate coverage as set out in Section 2. In an ideal world, the investigation would have covered both the entire set of relevant actors and the entire set of relevant activities. However, investigations can be stopped earlier due to funding constraints or unexpected developments (e.g., police forces have a duty to prevent a murder even if this would put the investigation in jeopardy). Besides a close reading of the investigation materials, interviews with police officers or prosecutors involved in such investigation would help detect potential coverage issues. If such issues become apparent, researchers should state them clearly and rework their research questions as to minimise biases.

Police databases (generating events) are best suited to carry out jurisdiction-based studies (see Figure 1 for an illustration). The stopping rule in this case also needs to be pragmatic. Inevitably, the network ends where the jurisdiction of the agency that has collected the data ends. While this offers a valuable picture of how OCGs operate in one or more markets, we might miss actors and ties that fall outside the jurisdiction of the law enforcement agency. For instance, in the study of local drug markets in Newport, Wales, Baika and Campana (2020) were able to model interactions among suppliers within the jurisdiction of the local police force. On the other hand, interactions between the local suppliers and higher-level importers of cocaine and heroin could not be modelled since such events took place outside the
jurisdiction of the local police force. In order to mitigate the problem, the authors
reconstructed stages of the supply chain through qualitative interviews with police officers.

Both sources of data - wiretaps and police-generated events - imply decisions about
boundaries and come with their own drawbacks and opportunities. There is no one-size-fits-
all criterion, but only strategies that are better suited to specific research questions. A clear
understanding of how the evidence was collected and what set of questions that evidence can
answer should guide scholars.

4.2. Time span of data collection

An issue that has so far received very little empirical consideration is the time span of data
collection and its impact on the analysis. Campana and Varese (2012: 25) first reflected on
the treatment of OC longitudinal data, suggesting two approaches: a time-based approach and
an event-based approach. A time-based approach would require a researcher to select a time
frame, e.g. months, quarters or years, and then extract the data accordingly. The second
approach would require identifying key events and then set the time frame accordingly. These
events are normally selected based on specific research questions and/or with the aim to
minimise biases (e.g., changes in police practices). There is no superior strategy a priori, as
noted also by Diviak (2019 in this same issue).

A time-based approach raises the issue of selecting the ‘right’ time-window. Among the few
studies to discuss this problem, Ouellet et al. (2019: 12) opt for a three-year window to
segment police data. They maintain that such a period is of “sufficient length […] to provide
the opportunity for the ‘full’ scope of the group (and their relationships) to come to the
attention of police” (Ouellet et al. 2019: 12). To further explore the impact of selecting a time
span on the structural properties of a network, we turn to the Thames Valley Police Data Set.
As shown in Table 2, very short time frames of data collection can lead to misleading
pictures when relying on police databases. A one-month time frame, for instance, is certainly
inadequate as it would suggest the absence of any structure in the OC landscape of the force
(30 actors, six ties) while we know from other sources, including interviews with officers and
intelligence data, that such structure does exist.

Table 2. Varying time frames for data collection, Thames Valley Data Set, 2011-2016
<table>
<thead>
<tr>
<th></th>
<th>1 month</th>
<th>3 months</th>
<th>6 months</th>
<th>1 years</th>
<th>2 years</th>
<th>5 years</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nodes</strong></td>
<td>30</td>
<td>83</td>
<td>141</td>
<td>243</td>
<td>428</td>
<td>745</td>
</tr>
<tr>
<td><strong>Ties</strong></td>
<td>6</td>
<td>16</td>
<td>44</td>
<td>154</td>
<td>7052</td>
<td>11446</td>
</tr>
<tr>
<td><strong>Average degree</strong></td>
<td>0.200</td>
<td>0.145</td>
<td>0.213</td>
<td>0.412</td>
<td>0.897</td>
<td>2.400</td>
</tr>
<tr>
<td><strong>Diameter</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td><strong>Components</strong></td>
<td>28</td>
<td>78</td>
<td>127</td>
<td>209</td>
<td>314</td>
<td>351</td>
</tr>
<tr>
<td><strong>Events in the original 2-mode matrix</strong></td>
<td>39</td>
<td>142</td>
<td>285</td>
<td>661</td>
<td>1879</td>
<td>7355</td>
</tr>
</tbody>
</table>

*Source:* Thames Valley Police Data Set.

*Notes:* Measures calculated on the 1-mode projection of the 2-mode actor-by-event matrix. It excludes sexual offences, domestic incidents, shoplifting and traffic offences.

Researchers normally have little say over the time range for which secondary data are available as changes in recording practices, specific IT systems or legal restrictions might all have a decisive impact. Furthermore, the suitability of a time frame is also a function of the frequency of the events. Our experience suggests that a longer time span is necessary when using police recorded events than when dealing with wiretaps as the latter are can capture a lower threshold of interaction than event data. For example, the 28 smugglers wiretapped by the Italian police discussed in Campana (2018) have exchanged an average of just above 50 phone calls per day over a period of about a month; the same individuals would have entered the police records as “arrested” or “wanted” only once.

Event-based framing requires a deeper knowledge about the list of potential events that might be relevant in that specific context. This knowledge should be acquired prior to data extraction through interviews with police officers or an open-source analysis. Ultimately, the choice of events is based on the research questions we are seeking to answer (special attention should be given not to select on the dependent variable). A combination of the two approaches is also possible, and often inevitable. If we want to understand changes in the structure of the network induced by a police operation, we would first need to identify the event (the police operation), and then extract the data based on a chosen time interval before and after the event. Another example is a situation in which police systems have been subject to a major overhaul. Far-reaching changes in recording practices need to be taken into
consideration when extracting the data: for example, the date of the overhaul may give us the starting point of our time span and we can then set the end point based on a chosen time interval (e.g. five years).

Finally, one might ask what is the time frame for a meaningful analysis using police records. While it is next to impossible to provide a general rule about setting such frame, we can offer some guidelines here. When two sources of evidence are available, one can triangulate the information to gauge the robustness of the time frame chosen. For instance, if from intelligence records we know that 100 OCG members operate in an area, we can check that the same number of actors appears in the event-based network. If no other information is available, or the extraction has been conducted using OCG members as seeds, one should consider that for police-recorded events meaningful time frames are in the region of years, not months. A rule of thumb could be to aim for the maximum length allowed by the system since, for instance, the last radical change in recording rules or police systems. However, one should also keep in mind that, if the extraction ends up covering events spanning very long periods, then collapsing the data into a single cross-sectional data set might create inaccurate pictures. In such a case, a longitudinal approach might need to be adopted when conducting the analysis (a point also raised by Diviak 2019 in this same issue). In all cases, researchers should be transparent in reporting what they included, as well as what they dropped from the analysis, and why.

5. The role of qualitative knowledge for organized crime research

Data sources differ with regard to the depth and richness of the evidence available to a researcher. So far, we have discussed how we can generate a network from wiretaps or police-recorded events. Building networks allows us to carry out structural analyses of OCG groups and markets. However, we believe that structural measures alone are not sufficient to fully grasp a criminal phenomenon. Worse, they might lead to misunderstandings. To illustrate this point, we now turn to two separate examples from the Russian mafia and a Nigerian human trafficking ring.

Varese (2013) reconstructs the network of criminal connections among a set of 164 actors based on an extensive police investigation into a Russian mafia group based in Rome. The degree centrality for the network indicates the presence of three top actors: Yakovlev, Pepe
and Sergeyev (these are fake names). The degree distribution steeply drops after these three actors are considered (Varese 2013: 904). Are we to conclude that all three actors are structurally equivalent? Based on structural measures only, the answer would be in the affirmative. However, thanks to the richness of the evidence available - and a coding strategy that captured both structure and content - Varese (2013) was able to match the structural position with substantive information about the activities of each actor. When evidence on both structure and activities is taken into account, it becomes evident that only one of the “key” actors (Yakovlev) is an integral member of the Russian mafia, engaged in a protection racket in Russia and overseeing investment decisions in Italy. Russian-speaking Mr Pepe is an Italian fixer working for Yakovlev. Sergeyev is a Russian businessman whose services Yakovlev relies upon to launder the money generated by various Russian groups. In other words, while Yakovlev and Pepe have roughly the same number of contacts, the nature of such contacts is radically different. This example suggests that structural measures taken alone can be misleading. We can improve our understanding by adding the content of a contact, for instance in the form of an attribute of a tie based on the information about the activities carried out by a dyad. It is vital for researchers to be fully aware of the nature of the ties included in a network as different sources of data generate different types of ties. The latter may differ on the nature and amount of information carried: compare, for example, a tie capturing a direct order from A to B with being co-targeted in a police operation. Arguably, the former offers more granular – and thus richer – evidence on the dyadic interaction. This, in turn, has an impact on the type of mechanisms or processes that can be legitimately studied (see, e.g., Faust and Tita 2019: 107-108 for additional examples).

Furthermore, a faction analysis of the same Russian mafia group shows that each of the three key actors has exclusive access to a number of other actors, generating three clearly distinct subgroups. The network is not fully integrated, but rather split into at least three significant subgroups. Pepe’s subgroup of contacts is very different in nature from those of Yakovlev. Pepe has contacts with the local Rome underworld and the Russians buy into that subnetwork. The fact that it is not integrated with the rest is actually the reason why Pepe is on the Russian’s payroll. Within Yakovlev’s subgroup, there are individuals who have access to violence and, when a dispute arises between Yakovlev and Sergeyev, threats are made and eventually Sergeyev, during a trip to Moscow, is killed. Only qualitative knowledge of the case can help us predict violence. The problematic split for the organization was the drifting apart between the two Russians, one of them trying to cheat the other and wanting to keep
some contacts secret, rather than the emergence of Pepe’s own subgroup. Finally, time has an effect on the structure. Yakovlev should be seen as the creator of the Russian mafia outpost in Italy. In the early phases, the network structure was more cohesive and integrated, yet also more ineffective because the bulk of Italian contacts grew only with time. Networks change over time and researchers must be aware of the point at which they observe the group.

Campana (2016a) reconstructed a network involved in human trafficking between Nigeria and Italy. Based on a court file, he built the network around 16 trafficking events that occurred during two months and saw the involvement of 25 offenders and 33 victims. In Figure 3, we present some additional analysis of the original data set.

Figure 3. Human trafficking network: degree centrality, roles and place of the offenders

Source: Nigerian Trafficking Network data set (in Campana 2016a)
Figure 3a presents the degree centrality of the offenders. As a 1-mode projection of the original 2-mode event-by-actor data set, degree centrality indicates the level of participation in trafficking activities of each individual based on the number of shared trafficking events. A core group of three actors stands out as the most central players.11 Based on this figure, one would be tempted to conclude that these three actors are the “key” players in the trafficking ring. The reality, however, is more complex. Based on the content of the phone conversations wiretapped, Campana (2016a) was able to reconstruct the role played by each individual. It then became clear that there was a full separation between those actors involved in transporting the victims and those involved in exploiting them (see also Section 2 above). Furthermore, it also emerged that it was the exploiters who were hiring the transporters in order to move their victims from Nigeria to Italy. In other words, transporters were responding to a demand generated by exploiters and were working as contractors for the exploiters on an ad hoc basis. Exploiters in the network (blue dots in Figure 3b) are at the margin of the network and score very low in centrality. This is because the exploiters were largely independent of each other and normally linked to the transporters only in one trafficking event (i.e., when the victim “belonging” to them was being transported). Further, Figure 3c places each actor in either a transit or a destination country. Exploiters are located in the destination country, while the high centrality actors at the centre of the graph are all in transit countries outside the European Union. This has crucial implications for disruption. Targeting the high-centrality actors might prove, in this case, both ineffective as other transporters will respond to the demand from the exploiters and inefficient as they operate outside the jurisdiction of this specific police force and, more generally, of the European Union. This example illustrates that structural measures are sensitive to the data collection strategy. What would have happened if, instead of an event-based network, we had been in a position to reconstruct the direct network of orders exchanged among offenders? Or a direct network based on the requests for illegal services related to human trafficking? It is plausible that, in such networks, exploiters would have shown a much more central position, particularly when measured on out-degree centrality.

Being in a position to add qualitative information in the form of actors’ attributes, as well as information about the mechanisms underpinning a given network, plays a crucial role in avoiding potential misinterpretations of the structure. We follow the plea by Robins (2009) not to ignore details about the individuals within a network. He highlighted a series of individual level factors, such as demographics, capacities and skills, psychological factors
and access to resources as well as dyadic factors (multiplexity, positive/negative ties, etc.). Node-level attributes, for instance those capturing the ability of an actor to access criminal resources, have been used in conjunction with centrality measures to identify different types of key roles, as in the case of the study of a methamphetamine trafficking ring by Bright et al. (2015). Hollstein (2014) aptly points out that “we cannot make valid statements about networks based on qualitative data alone without linking them to with data on network structure”; we suggest that the same holds true also for structural data if taken alone. Mixing qualitative and quantitative information is a two-way process that gives the best results when it takes place at various points during the research process (for a discussion of the different types of mixed methods research design, see Hollstein 2014: 11-18). In this paper, we focus on the use of qualitative evidence as a strategy to improve our understanding of a specific case study and avoid pitfalls. A different issue is how to integrate qualitative and qualitative evidence to systematically compare and contrast different case studies, as in the qualitative comparative analysis discussed by Hollstein and Wageman (2014; on this point, see also Diviak 2019 in this issue).

6. Conclusions

The paper has pointed to a number of theoretical and data-related challenges for the study of OC networks. In this field, that there are two key research questions, namely the analysis of the internal hierarchies of groups and of the group’s position within criminal markets, and they are normally best addressed using two different types of data, phone wiretaps and police-recorded events respectively. These two sources of data generate a number of validity and reliability issues. As for wiretaps, we have discussed self-censorship, group coverage and the connection between conversations and behaviour. In police-held databases, a number of “events” related to individuals are recorded, such as co-arrest information, surveillance reports and previous criminal history. Individuals differ on their degree of culpability, ranging from suspects to those cautioned, arrested and convicted. We have shown that the choice of whom to include in the analysis has significant effects on the results and discussed the ethical implications of including suspects. While there is no universal recipe on whom to include, we suggest to consider all individuals targeted by the police, while ensuring their anonymity and excluding information that might lead to identification. The choice over how many years of police-recorded events a researcher needs to include in the study also affects the analysis. Clearly, shorter periods yield less reliable results. We suggest that normally a
longer time span is needed when using police-recorded events as opposed to wiretaps; although we do not wish to insist on a universal recipe (some scholars have suggested a minimum of two years’ worth of data).

As for any SNA study, the analysis of OC networks faces the “network boundary specification” issue. In discussing this problem, we have presented studies that have pragmatically decided to accept that the network ends where the police data end, a stopping rule that must be consistent with specific research questions. We noted that these studies face an additional boundary problem, namely that of assigning individuals to OC groups. We reviewed five criteria for assigning actors to groups: membership list provided by the gang itself, reputational approach, offence-based membership, event-based attribution and community detection algorithms. We also discussed how police data sets are built, and whether sampling is feasible. Wiretaps are constructed through the equivalent of snowball sampling by the police. While normally only a limited number of actors are directly targeted, many more appear in the conversations, but we cannot know if two people talking to a person directly wiretapped also talk to each other. Scholars need to be aware of the problem and might opt for the construction of two separate networks, one including those under surveillance only and one including all actors caught up in the investigation. We have also cautioned that randomly sampling conversations or police-recorded events run the risk of missing important players such as bosses.

Structural analysis needs to be supplemented by qualitative information on how the data are constructed and on the groups (and markets) themselves. Thus, we call for an alliance between SNA and qualitative methods, stressing the importance of conducting interviews with selected key informants. More generally, we urge SNA scholars to bear in mind the lessons of the early practitioners, such as Jacob L. Moreno and Clyde Mitchell, who combined ethnographic insights with SNA methods. While some physics-inspired scholars set themselves the task of uncovering the unique structure of human networks and have suggested different models universally applicable, we advocate a mixed-method SNA. Ultimately, we believe that SNA is uniquely able to capture the relational nature of OC, both within and across groups. Yet structural analysis alone is not sufficient to capture the complexity and diversity of OC groups’ organizations and markets. As writer John le Carré said, “a desk is a dangerous place from which to view the world”.
Appendix: The Thames Valley Police Data Set

The area covered by Thames Valley Police, in the South East of England, had just below average number of OCGs per one million population in 2016 (HMIC 2017). The data we obtained include fully anonymised information on all OCGs active in Thames Valley between 2010 and 2016, as well as on individual members. For the latter, we obtained their group membership, country of birth, nationality, ethnicity and age. We were also given information on the ‘activities’ OCG members engaged in. The ‘activities’ are based on events recorded by the force (N = 14,495 between January 2010 and October 2016). Such events include type of crime, first three digits of the postcode of where the crime took place, date of the crime and their role in the event (offender/victim). The ‘events’ included in our data set are not limited to instances where the OCG member was arrested. Rather, they include instances where the member was simply a ‘Suspect’ (28.1%), when ‘No further action’ was taken (21.2%), when the suspect was simply ‘Detected’ (20.7%) or ‘Charged’ (12.0%) or ‘Arrested’ (10.7%). We also have events labelled as ‘Postal requisition’ (a type of summon, 3.2%) and ‘Other’ (4.1%).
Interviews cited in the text

Interview 3. Police officer, Head of an OC investigation team, Palermo, April 2016.

References


ENDNOTES:

1 The remaining of this sub-section is broadly based on the discussion included in Campana and Varese (2012).
2 Specific procedures have been developed to analyse data collected through snowball sampling, such as the conditional estimation procedure for exponential random graph models (Pattison et al. 2013). In this paper, we simply alert the reader to the implications of the sampling strategy on the data analysis.
3 We refer to Heckathorn and Cameron (2017) for a detailed discussion on the current debate around respondent-driven sampling. Frank (2005: 39) offers a statistical treatment of the problem and suggests that “likelihood-based inference could still be possible if the data available make the design ignorable”, or better to rely on probabilistic network models with dyadic dependence in cases when it is difficult to determine the inclusion probability of the snowball or link-tracking sampling.
4 While this has been the most common strategy in the studies on organized crime so far, this is not the only one. To avoid a loss of information on the underlying structure, a researcher can keep the bipartite structure and rely on specific measures for 2-mode networks (Borgatti and Everett 1997; also Everett and Borgatti 2010). Further, another way of conceptualising police records is as relational events with a time dimension associated, and then rely on specific models such as relational event models (Butts 2008) or dynamic network models (Stadtfeld et. al. 2017).
5 See also Rostami and Mondani (2015) who examine the same Swedish street gang using three different data sets, based respectively on intelligence, surveillance and co-offending data.
6 We are grateful to an anonymous Referee for suggesting this point to us.
7 Faust and Tita (2019: 107) further note that arrest data can be seen as “repeated samples of the population of offenders that are then used to create a co-offending network”.
8 This procedure can be extended to include the alters of alters in a way similar to the process described by Robins (2015: 70-72).
9 DellaPosta (2017) combines a reputational approach with a “label propagation” computational technique to assign known Mafia members to known Families – in the rather specific situation in which, for a subset of actors, their membership of the Mafia is known based on a reputational approach but not their affiliation to a specific Family. This is an extension of the reputational approach with a potential solution for missing information.
Police forces in the UK can record a single-member group if only one person has been identified but “professional judgement may indicate the existence of a group” (OCGM Manual 2010: 15).

A similar analysis based on the betweenness centrality points to the same result. Further consideration on this case are included in Campana (2016b).