21D009 Networks Project
Crime Networks & Law Enforcement

Clinton Leung, 200870
Gian-Piero Lovicu, 200752
David Ampudia Vicente, 82929

Introduction

Organised networks conducting criminal activity (also referred to as criminal networks) have been problem for much of mankind. In modern society, they remain ubiquitous and underpin much of the illegal drug trade, trafficking of various sorts and terrorism cells.

Law enforcement face two major problems in dealing with criminal networks: discovering the structure of the network and figuring out the most effective way to disrupt or dissolve it (Oliver et al. (2014), Martins (2021)).

Using graph analysis tools to tackle these problems is appealing. Without some sort of observed organisational structure, constructing a criminal network can instead leverage relational data to draw a graph that connects members. With this data, modelling can help to inform strategies to best target individuals in the network based on their structural role in the graph (Oliver et al. (2014)).

We construct a 'crime investigation and capture' algorithm to model how a criminal network could be disrupted. The algorithm is a a modified version of the Susceptible-Infected (SI) model. The algorithm 'investigates' suspected criminals (susceptible), which are then 'captured' with some probability following an investigation (infected). The algorithm can incorporate different investigation strategies, different probability of capture models and different reaction functions of nodes following an (unsuccessful) investigation. The success of a strategy is measured by how many investigations a strategy needs to capture a certain proportion of criminals in the network.

Crucially the algorithm does not know the structure of the network prior to initialisation, which is a key difference to other work in this area. Arguably, this is a more realistic representation of how criminals and law enforcement interact, where there is sometimes little information available to law enforcement for decision making. We apply the algorithm to some real-world criminal network data, obtained from the Mitchell Centre for Social Network Analysis.

Overall, our algorithm is reasonably successful at disrupting the criminal networks in our sample, though it struggles in capturing less important nodes around the fringe of the network. In terms of performance, we find that designing a investigation strategy using the structure of the network universally outperforms a random baseline. In general, networks where it was hard to reach the most important criminals quickly were able to evade the algorithm more effectively. A particular case was the Al Qaeda network, whose cell structure proved very effective at avoiding capture. Here a strategy that balanced between breadth and depth first search performed best. It was not only able to capture groups of criminals, but also recognise the weak ties it needed to capture to move between them. This particular result highlights the importance of designing an investigation strategy that takes into account the expected structure of the network.

I Literature

The literature investigating criminal networks is broad, with only a small subset focused on disruption. We follow Oliver et al. (2014) to inform our review.

One part of the literature discusses how to define connections in a criminal network, which is challenging given many criminal networks are covert. Indeed, Oliver et al. (2014) suggest that the boundary for membership into criminal networks is fuzzy and observing it is often restricted to the data available for collection. In our data we have networks where edges represent various kinds of relationships: phone calls, co-attendance at meetings, family relationships, etc. Resilience in connections can vary based on their type. For example, Lauchs et al. (2011) discuss that connections based on pre-existing social ties are often more resilient than those that are not. We treat connections as homogenous within and between networks, regardless of their type. This may affect how realistic our results are, but the criminology theory required to interpret types of connections is out of scope for us.

Another strand of the literature is aimed at categorising covert networks. Broadly networks are either classified into groups with terrorist (and/)or criminal aims (Harris-Hogan (2012), Varese (2010)). The networks in our data set fit within this classification framework (see Data section). The activities of a particular network are likely to affect its structure. For instance, Morselli et al. (2007) notes that terrorist networks tend to be less centralised than criminal ones (we also observe this in our sample). Other aspects of network structure, such as density or hierarchy are also important (Lauchs et al. (2011)). This may affect the success of particular strategies taken to disrupt the criminal network and is something we consider in interpreting our results.

Control and leadership within a network is also relevant. For instance, identifying leaders in the network is important for designing a disruption strategy. However, identifying leaders is not always straightforward. For instance, Lauchs et al. (2011) note that the most central members of a network are the most vulnerable to capture and therefore are unlikely to be the leaders of the network. However, other papers note that centrality is a good predictor of control, particularly in highly centralised networks (Carley et al. (2002), Koschade (2006)). Most studies use some measure of centrality to identify the most important nodes in a network (Varese (2010)). Consistent with this, we use eigenvector centrality to measure the importance of a criminal in the network.

A small strand of literature is devoted to the disruption of criminal networks and our work fits in here. Most disruption strategies focus on removing central nodes in the network. For example, Carley et al. (2002) argue that removing central nodes disrupts the flow of information in the network, preventing the group from taking decisions or reaching a consensus. As central nodes are also often identified as leaders, removing these nodes can also disrupt the network through encouraging the formation of destabilising factions (Crenshaw (2010)). Disruption strategies based on central nodes may be less effective in networks that are highly decentralised or have multiple central actors (Carley et al. (2002)). Some papers have also argued that the best disruption strategy is to first fully identify the network and then dismantle it all at once (Helfstein and Wright (2011)). All of these disruption strategies require information about the full network before choosing a node to target.

However, in reality one may not possess enough information to inform a strategy based on some sort of 'precision strike' (i.e. targeting a central node) or, alternatively, may wish to disrupt the network as soon as a node is discovered (for instance, in the case of some forms of trafficking). Our work reflects this reality and diverges from the literature by assuming the network is initially unknown. We then reveal information to direct a disruption strategy only as we interact with the network.

Finally, a growing strand of literature is focused on edge prediction and other strategies at imputing missing data from the network. In recent years, this has involved using machine learning algorithms to make predictions (Martins (2021)). We will not consider edge prediction in this work.

II Algorithm

Our algorithm is a modified version of the Susceptible-Infected (SI) model from the epidemiology field.

In the algorithm, each node has three 'states': undiscovered, suspected, caught. Initially, all nodes are undiscovered. The algorithm begins by randomly catching a node for law enforcement. Once a node is caught, all of its neighbours are revealed as suspects. Law enforcement then choose a suspect to investigate (defined by a pre-determined strategy). When law enforcement investigate a suspect, they may catch it with some intrinsic probability that is unknown to them (and which is possibly endogenous to the current state of the network). If the investigation is successful, the node is caught and all of its immediate neighbours are revealed as new suspects, which become candidates for future investigation. If the investigation is unsuccessful, the node remains a suspect and law enforcement will reapply the strategy (which may result in it investigating the same or another node). Investigation rounds keep occurring until all the nodes are caught or until the algorithm terminates (reaches a maximum number of investigations).

There are two aspects of the algorithm that the user defines.

The **strategy** is a set of rules for choosing which node to investigate. The strategy can be anything, though we have chosen strategies based on the structure of the network. The one restriction on the strategy is that it can only access the information so far revealed by the network. So for example, if the strategy is based on investigating the suspect with the highest degree, this must be calculated based only on the edges that are currently visible. This is intuitively appealing, because in reality law enforcement can only react to the information set available to them.

Our strategies include:

- **Naive random**. Randomly choose a suspect to investigate
- **Maximum diameter**. Investigate the suspect that would maximise the diameter of the graph of caught criminals if that suspect were caught. This is a depth first strategy that attempts to traverse as much of the network as possible.
- **Greedy**. Investigate the suspect with the highest degree: the suspect that the police has the most information about and has the highest chance of catching. This localised strategy is breadth-first and frequently results in ties. We try multiple strategies to break ties: randomly, in favour of node with highest eigenvector centrality, in favour of node that maximises the diameter of the graph. If the tiebreaker results in another tie, break randomly.
- **Balanced**. Investigate the suspect that achieves the best score between a weighted sum of the greedy and diameter strategies. This strategy attempts to combine aspects of depth- and breadth-first search. A hyperparameter α is the weight assigned to the maximum diameter strategy, so lower values mean more weight to breadth-first.
- **Least central node**. This investigates the node who is connected to already caught criminals with the lowest degrees.

The **model** refers to the way we assign probabilities of capture for each node. The probability model contains two parts. The first is an intrinsic probability of catching a node which is unknown to law enforcement and does not influence the strategy. This probability could be constant, a function of network structure and/or node attributes such as the inverse of the eigenvector centrality. It could even be dynamic, such that the probability of capture updates as investigations occur. For instance, an unsuccessful investigation could reduce the probability of capturing a node (i.e. the criminal goes 'underground'), while an increasing share of a node's neighbours revealed as suspects or captured could increase the probability of capture (i.e. if your buddies are caught, you are more likely to be snitched on). We only report results of the constant probability model, where each suspect is assigned a 5 per cent probability of being caught if they are investigated.

The second part of the probability model models the way in which the police gain information once they have caught suspects. Here we draw from ideas in the diffusion literature and assign some probability c as the increased chances of catching a suspect for each incident informed edge weight. So the total probability of catching a suspect is given by equation 1 where P(i) denotes the total probability of capture of node i, $\pi(i)$ is the intrinsic probability which we have assigned 0.05 for all i, c is a constant which we have used 0.05, $\mathcal{C}$ is the set of currently captured criminals, and $A_{i,j}$ is the i, j^{th} element of the weighted adjacency matrix.

$$P(i) = \pi(i) + c \sum_{j \in \mathcal{C}} A_{i,j} \tag{1}$$

There could also be justifications of more complex models such as exponentially decaying weight instead of a constant c if you believed, for example, that there are decreasing marginal information gains when learning about a suspect until some fully saturated level. On the contrary you could also define models that have exponentially increasing weights justified by a belief that information on a suspect builds upon itself and the total information is more than just the sum of each individual piece. We do not have enough empirical justification for these beliefs and have stuck with the simple constant model.

We also define a dynamic model in the code, but leave the results from this as an extension (simulations ended up being too expensive to consider two models). In this model, an unsuccessful investigation decreases the probability of capture by 50 per cent, but increases it proportionately with the number of a suspect's neighbours that have been caught. This modified model moves from simple diffusion toward a linear thresholding approach to spreading.

Given a particular strategy and the constant probability model, we run 1000 simulations of the algorithm on each network and averaged the results.

We measure the success of a strategy by looking at the number of investigations taken to achieve each quantile for:

- Criminals captured
- Cumulative eigenvector centrality of criminals (over the whole network) captured

The number of criminal captured is straightforward to interpret. The cumulative eigenvector centrality is supposed to identify the success of a strategy in and capturing the leaders in the network (assuming these are the most central nodes).

III Data

Most of the crime network data comes from the Mitchell Centre for Social Network Analysis at the University of Manchester.[1] In total, we have 18 networks, equally distributed between gangs, drug smuggling (drugs) and terrorist cells (terrorism). Full details on the networks can be found on the Mitchell Centre website.

Since the data provided are not consistent across networks, we make some simplifying assumptions to assist with comparability. We force each of the networks to be undirected. Where there are multiple connection types, we sum the adjacency matrices across each type (so there is an edge between two nodes whenever they have at least one type of connection). We also discard node attributes, so the information available to the algorithm is consistent. Despite these simplifying assumptions, we recognise that these data may still be important. Incorporating them could form an extension to this work, for example by being embedded in the probability model passed to the algorithm.

For each network, the algorithm is only run on the largest component. This ensures the convergence of the algorithm doesn't depend too much on the initialisation point (i.e. if the first criminal caught happens to be outside the largest component of the network, the algorithm will never reach any of the nodes in it). It also helps with computational feasibility.

Summary statistics are reported in Table 1 for the largest component of each network. The calculation of some statistics (such as degree, eigenvector centrality and clustering) utilise the edge weights (which in some instances are large). This explains some of the anomalous values in the summary statistics.

[1]Except for the Montagna crime network, which was obtained directly from Ficara et al. (2021)

Table 1: Descriptive Statistics of Criminal Networks

Network	Criminal Type	Nodes	Edges	Degree Mean	Degree Max	Diameter	Triangles	Clustering Mean	EC Mean	EC Max	Density	EC Range	EC log-ratio
17N	Terrorism	18	46	5.1	17	2	940	0.7	0.2	0.4	0.3	0.4	2.0
9/11 Hijackers	Terrorism	61	131	4.3	15	7	171	0.5	0.1	0.4	0.1	0.4	5.6
ISIL	Terrorism	25	27	2.2	9	4	3	0.1	0.2	0.5	0.1	0.5	2.5
Al Qaeda	Terrorism	125	312	5.0	17	10	361	0.6	0.0	0.3	0.0	0.3	11.6
Caviar	Drugs	110	205	3.7	60	5	646230	0.3	0.1	0.5	0.0	0.5	5.4
Operation Acero	Drugs	25	37	3.0	13	5	12	0.3	0.2	0.5	0.1	0.5	3.2
Operation Jake	Drugs	38	50	2.6	12	4	15	0.1	0.1	0.5	0.1	0.4	2.8
Operation Juanes	Drugs	51	93	3.6	16	7	46	0.4	0.1	0.5	0.1	0.5	5.9
Operation Mambo	Drugs	31	58	3.7	15	4	35	0.4	0.1	0.5	0.1	0.4	3.2
Heroin Natarjan	Drugs	38	87	4.6	20	4	46	0.4	0.1	0.4	0.1	0.4	2.6
Italian Gangs	Gangs	65	113	3.5	21	6	57	0.4	0.1	0.4	0.1	0.4	4.2
London Gangs	Gangs	54	315	11.7	25	4	4321	0.6	0.1	0.2	0.2	0.2	3.3
Mali Terrorists	Terrorism	36	67	3.7	11	7	43	0.4	0.1	0.4	0.1	0.4	4.0
Montagna Operation	Gangs	135	320	4.7	40	6	16745	0.5	0.1	0.3	0.0	0.3	5.3
Ndrangheta	Gangs	151	1619	21.4	75	5	89103	0.8	0.1	0.2	0.1	0.2	5.7
Malasyian Extremists	Terrorism	78	623	16.0	55	4	15961	0.7	0.1	0.3	0.2	0.3	5.2
Project Togo	Gangs	33	47	2.8	19	4	14	0.5	0.1	0.6	0.1	0.6	2.9

In addition, grouping the networks by their type (drugs, gang or terrorism) and and reporting the median of the statistics is helpful to summarise the networks (Table 2). Some prima-facie observations by the type of network include:

- Gangs and terrorists tend to have a higher average degree than drugs.
- The diameter of terrorists appears larger than for gangs, which is larger than for drugs.
- Consistent with the lowest average degree, drugs also tend to be less clustered. Gangs and terrorists have a higher clustering coefficient.
- The log ratio of eigenvector centrality is lower for drugs than for gangs and terrorists, consistent with a lower average degree and clustering coefficient.

Table 2: Descriptive Statistics of Network Groups

Criminal Type	Nodes	Edges	Degree Mean	Degree Max	Diameter	Triangles	Clustering Mean	EC Mean	EC Max	Density	EC Range	EC log-ratio
Drugs	38.0	72.5	3.69	15.5	4.5	40.5	0.35	0.13	0.47	0.10	0.46	3.2
Terrorism	48.5	99.0	4.64	16.0	5.5	266.0	0.50	0.11	0.41	0.10	0.38	4.6
Gangs	65.0	315.0	4.74	25.0	5.0	4321.0	0.48	0.09	0.32	0.09	0.32	4.1

Combined, these statistics imply that drug networks are more sparsely connected and less centralised than the other types, though at the same time they have a smaller diameter. Together these statistics do not give much of a prior on which type of network may be more susceptible to investigation.

IV Results

This section summarises the results across strategies and networks. First, we analyse the overall performance of different strategies. Then, we consider how network characteristics may affect the performance of the algorithm, regardless of strategy. Finally, we discuss the performance of different strategies for networks with specific characteristics, as well as some particular networks of interest.

IV.A Strategies

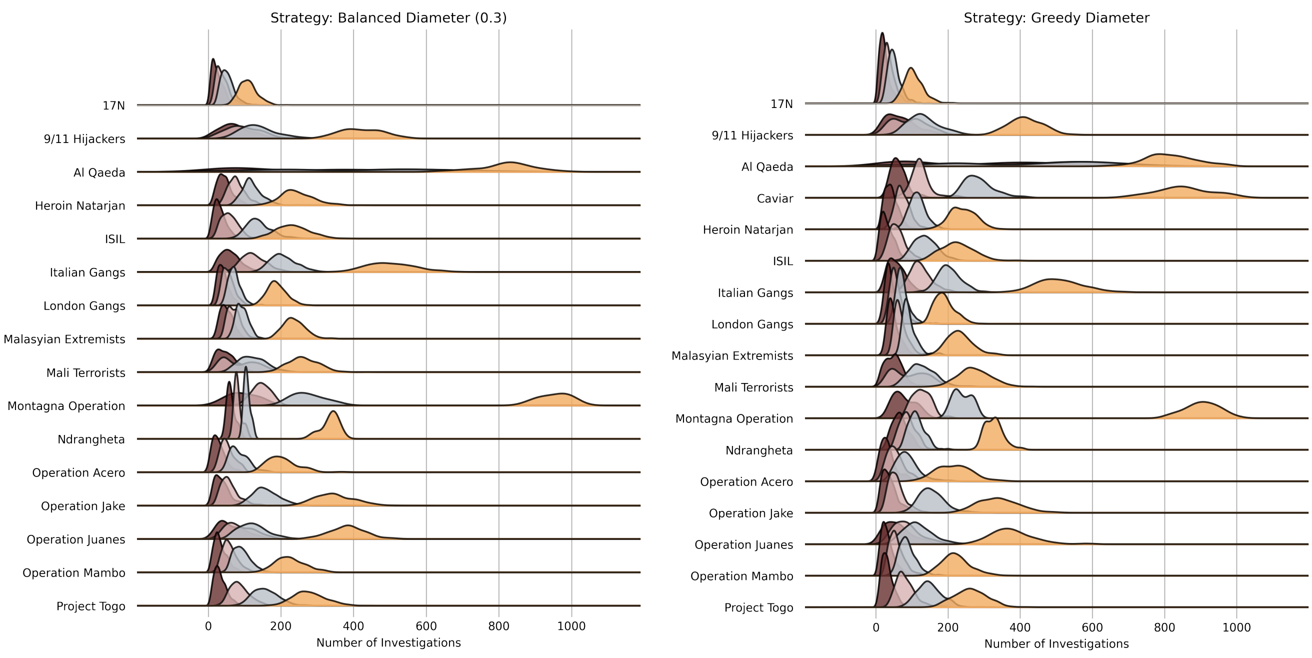
Table 3 shows the average results across all networks for each of our chosen strategies. The incomplete column captures the share of simulations where the network was still not fully captured after 1000 investigations.[2] Observe that the performance across all strategies is non-linear: it takes fewer investigations to capture the second and third quantiles, compared to the first. It also takes by far the most investigations to capture the last quantile, especially for the metric looking at the proportion of caught eigenvector centrality. This implies that later investigations are probably 'mopping up' nodes that are sparsely connected at the edges of the network and not particularly important.

[2] Results for the individual networks across each strategy are included in the appendix.

Table 3: Average Strategy Results

	Number of Investigations								
Metrics	Proportion of Caught Criminals				Proportion of Caught Centrality				*% Incomplete*
Quantile	0.25	0.50	0.75	1.00	0.25	0.50	0.75	1.00	
Balanced Diameter (0.1)	65	126	202	292	48	82	127	292	0.05
Balanced Diameter (0.3)	62	121	193	281	45	79	129	281	0.04
Balanced Diameter (0.5)	65	123	194	280	48	82	132	280	0.06
Balanced Diameter (0.7)	74	133	202	287	61	97	146	287	0.02
Balanced Diameter (0.9)	78	143	214	295	62	109	170	295	0.09
Greedy Diameter	68	133	212	310	56	95	143	310	0.17
Greedy Eigenvector	81	159	261	375	63	101	149	375	1.35
Greedy Random	80	157	260	374	61	101	154	374	1.41
Least Central	100	194	292	382	91	156	243	382	2.89
Max Diameter	80	148	228	311	65	109	174	311	0.24
Naive Random	113	210	307	390	92	160	241	390	3.56

The best performing strategies utilise a balanced diameter approach (mixing between the greedy diameter and maximum diameter strategies). The optimal value for the balanced strategy is $\alpha = 0.3$, though results for $\alpha = 0.1, 0.5$ require only slightly more investigations. This implies that a mixing of depth and breadth first strategies is most effective. Comparing the greedy diameter (only) and balanced strategies across the quantiles, observe that the performance gap (in investigations) widens as the algorithm moves to later quantiles. This implies that the maximum diameter aspect of the balanced strategy is most helpful during these later investigations. By choosing suspects to maximise the diameter, it can 'mop up' the nodes around the edge of the network most effectively. The performances of the optimal balanced and greedy diameter strategies across networks are shown in Figure 1.



(a) Balanced diameter ($\alpha = 0.3$) (b) Greedy diameter

Figure 1: Number of Investigations and captured eigenvector centrality

This figure plots the number of investigations required to capture each quantile of criminals. The distribution is based on 1000 simulations for each network-strategy pairing.

More broadly, observe that all of the strategies outperform the baseline random strategy (naive random). Paying attention to structural features of the network can therefore provide useful information over the baseline, regardless of what that information is. This applies even with extremely limited information, as in the earlier investigations when very little of the network has revealed itself.

IV.B Broad Network Characteristics

Although particular strategies may perform well in aggregate, the performance of the algorithm in general may be influenced by the underlying characteristics of a network. To explore this, we run linear regressions on the results of each individual simulation, for each network-strategy pair (Table 4). We specify four models: across each of the performance metrics and for the investigations required to capture the first and third quantiles of criminals. Note that investigations for each quantile are scaled by the total investigations required to capture the network. The scaling allows for comparison between networks of different sizes.

Table 4: The effect of network attribute on investigation effectiveness

Dependent Variable:	First Quantile		Third Quantile	
	Criminals	Centrality	Criminals	Centrality
Mean Degree	0.15***	0.06**	-0.05	0.19***
	(0.03)	(0.02)	(0.04)	(0.06)
Logratio Eigenvector Cent.	0.01	0.45***	0.14***	0.25***
	(0.02)	(0.05)	(0.02)	(0.07)
Eigenvector Cent. Mean	-0.04	0.03	0.06	0.45***
	(0.05)	(0.05)	(0.04)	(0.09)
Density	0.20***	0.44***	-0.09*	-0.25***
	(0.05)	(0.04)	(0.04)	(0.07)
Diameter	-0.05*	0.06	-0.17***	-0.12*
	(0.03)	(0.04)	(0.03)	(0.05)
Triangles per Node	-0.06***	-0.10***	-0.04***	-0.04**
	(0.01)	(0.01)	(0.01)	(0.02)
Mean Clustering	-0.01	-0.09***	-0.04	-0.02
	(0.02)	(0.01)	(0.03)	(0.03)
Number of observations	68,898	68,898	68,898	68,898
Strategy FE	Yes	Yes	Yes	Yes

Note: The table reports results for the number of scaled simulations it takes to capture a fourth and three fourths of a criminal network. Simulations are scaled using the mean number of total investigations needed to capture a network for each strategy-network pair, which ensures comparability. These dependent variables as well as the network attributes are standardized, so that the effects are in standard deviation units. Strategies are included as fixed effects. Each observation corresponds to a simulation.
***, **, and * indicate significance at the 0.01, 0.05 and 0.10 levels, respectively, using two-tailed tests.

The covariates in the regression are selected characteristics for each network and are standardised, so their magnitude is comparable.[3] We also include strategy dummy variables as fixed effects to capture variation that is strategy specific and linearly related to algorithm performance.[4] This way, we can interpret the regression coefficients as only capturing the effect that network characteristics have on the performance of the algorithm (note that a positive value for the coefficient indicates the algorithm performs worse).

Noteworthy results in Table 4 include a large and positive effect from having a large log-ratio of eigenvector centralities. This is expected: when highly central nodes are scarce relative to unimportant nodes in the network, this ought to slow down the progress of the algorithm by revealing fewer suspects for each criminal caught. Denser networks tend to slow down the initial progress of the algorithm, but is useful once a sufficient number of criminals have been caught. Observing a positive correlation between density and the clustering coefficient of a network, this could reflect the algorithm having trouble early on in 'escaping' from one cluster to explore suspects in another. It could also reflect why the coefficient on the clustering coefficient is insignificant - the variation is being captured via the density of the network instead.

The diameter of a network is not important for the early performance of algorithm, but a larger diameter seems to elicit faster progress once a majority of criminals have been caught. More triangles facilitates

[3]The log-ratio of eigenvector centralities $= log(\frac{max eigenvector centrality}{min eigenvector centrality})$
[4]Including non-linear interactions does not change results significantly.

algorithm performance, which likely reflects better connected networks being easier to explore, since on average more suspects are revealed for a criminal caught. Other coefficients offer less intuitive or insignificant results. As we will see below when exploring the results for individual networks, these may capture heterogeneity that is only explained on a case-by-case basis.

IV.C Specific Network Characteristics

Figure 2 plots the distribution of investigations required to capture each quantile of criminals (by eigenvector centrality) for the balanced diameter and maximum diameter strategy. Observe that in general the distributions are tighter for earlier quantiles, meaning results are more consistent earlier in the simulation. So even though the initialisation point for each simulation is random, the strategy helps the algorithm to quickly revert to capturing nodes with a similar eigenvector centrality.

Notice that also for the last quantile the distributions have (in some cases, extremely) fat tails, which (again) reflects the algorithm mopping up nodes on the edge of the network. With a low probability of capture and few other suspects available to investigate, this may take several investigations to achieve for each remaining suspect. In addition, for each additional suspect captured it is likely that no new suspects are revealed.

It is unclear whether the algorithm reflects the reality of investigating a criminal network. It might be realistic if one assumed 'fringe members' with only few/tenuous connections to the network may be most resource intensive for law enforcement to capture (say if there is less information available about them). Based on the results from Table 3 and Figure 2, from a resourcing perspective law enforcement may even find it optimal to abandon mopping up the edges of the network. However, this would of course depend on the type of the network and its ability to regenerate based only on the fringe nodes.

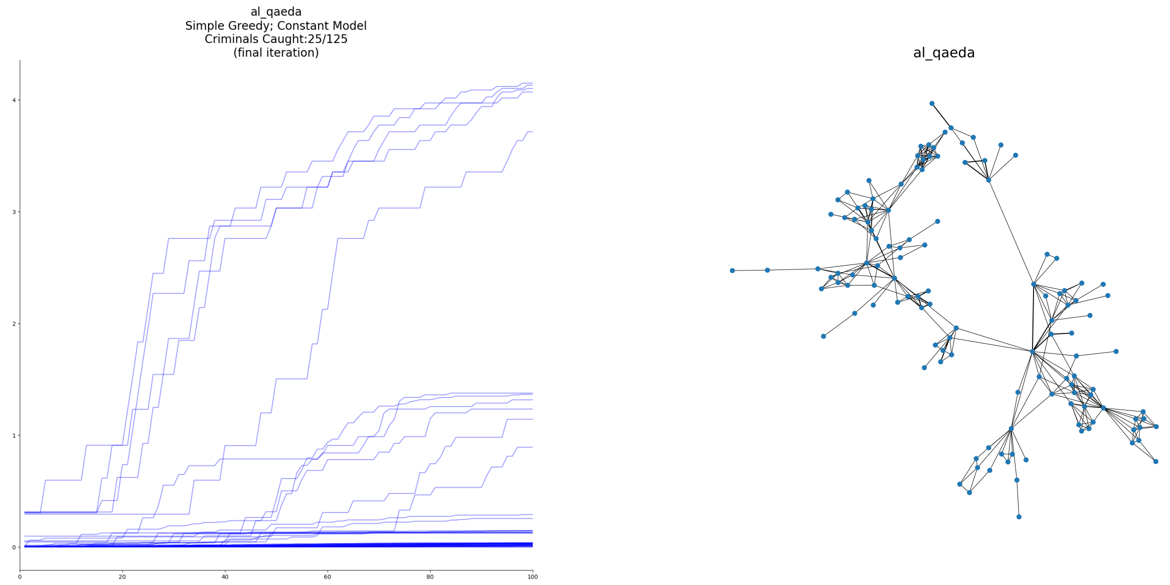
Figure 2 explores the correlations between the six network characteristics from our regression and the proportion of investigations required to capture the first quantile of eigenvector centrality. Results are displayed for the balanced diameter strategy (best overall). If the performance of the algorithm is linear, we would expect that 25% of the total investigations are needed to capture 25% of the eigenvector centrality of the network. This linear threshold is a rough benchmark for performance, though we observe that almost all networks outperform this. This is consistent with our observations in 3, where a disproportionate number of investigations were required to capture the last quantile of eigenvector centrality.

Firstly, we can observe that Al Qaeda is an extreme outlier; its network structure is uniquely able to avoid capture. Figure 2 demonstrates that Al Qaeda stands out in four measures. It has the highest diameter by a large margin, a low density, the lowest mean eigenvector centrality and an astronomical ratio between the maximum and minimum eigenvector centrality nodes (the log ratio is 11.6 compared to the next highest of 5.9, the same comparison in absolute terms would have been a ratio of 114,171 versus 359!). The combination of large diameter, low density and few important members seems to be a recipe for success in evading disruption by law enforcement.

Leaving Al Qaeda aside as an outlier, we can observe some trends across the other networks. Mean degree, average clustering and network density are correlated with poorer algorithm performance. However, this is less clear for the mean and log-ratio of eigenvector centrality. In general, this reflects that it is harder to capture important criminals quickly when networks are more dense, more clustered, and/or have a higher mean degree. Intuitively, the algorithm gets bogged down in a densely connected part of the network and is unable to proceed to investigating the most important criminals. This appears to be the case even when the strategy gives some weight to maximising the diameter of the graph (and so potentially trying to move out of a dense patch of the network).

Figure 2: Network properties that factor into the success of the balanced diameter strategy with $\alpha = 0.3$ in terms of caught eigenvector centrality.

Figure 3 shows the Al Qaeda network graph and a trace plot with the cumulative eigenvector centrality captured in the first 100 investigations. Clearly, in a majority of runs the algorithm gains no traction. Only 5 of the investigations out of 50 are able to capture any significant number of important criminals here. This shows exactly how successful a cell structure is for a criminal network and why it is so infamous. The network graph visually shows us how Al Qaeda has created a non-small worlds graph that has a reasonable level of clustering, yet is able to maintain a low density and a high diameter. Furthermore, it uses unimportant criminals to create long links between important criminals (the nodes in the centre of the clusters with high eigenvector centrality), demonstrating how weak ties between cells can be leveraged to slow down the algorithm. This characteristic is also reflected by the large ratio between maximum and minimum eigenvector centrality of the nodes.



(a) Greedy random

(b) Al Qaeda Network Graph

Figure 3: Trace plot of 50 simulations of Al Qaeda. The x-axis are the number of investigations and the y-axis is a measure of eigenvector centrality caught.

Returning to figure 2, we also observe that, in general, the performance of the algorithm is roughly consistent across the three types of criminal networks (drugs, gangs, terrorists). The terrorist networks are the most difficult to capture (appear more on the right of the scatter plots), the drug networks are in the middle and the gang networks are the easiest (left). This can be somewhat intuitive as a lot of gang networks are often fairly interconnected with many relationships. On the other hand, drug networks often rely on long chains of communications and are more concerned with being covert. Terrorist networks try to form cells and spend significant effort trying to hide their operations.

The Ndrangheta Mafia and the London gang are outliers that are relatively hard to capture compared to other gang networks (Montagna, Italian, Togo). Notice that Ndranghetha and London are more chaotic and disassortative compared to the other gang networks (Figure 4, Figure 5). The other gang networks exhibit a tree-like structure that branches off of a few key central nodes. So, we find that these more organised networks appear easier to capture perhaps because the strategy can more predictably target the most important nodes in the network.

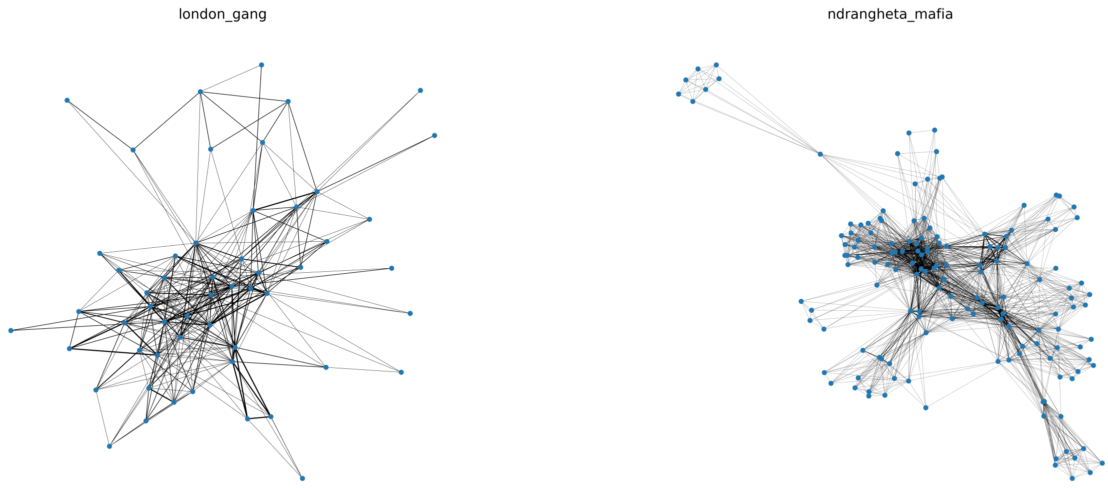


Figure 4

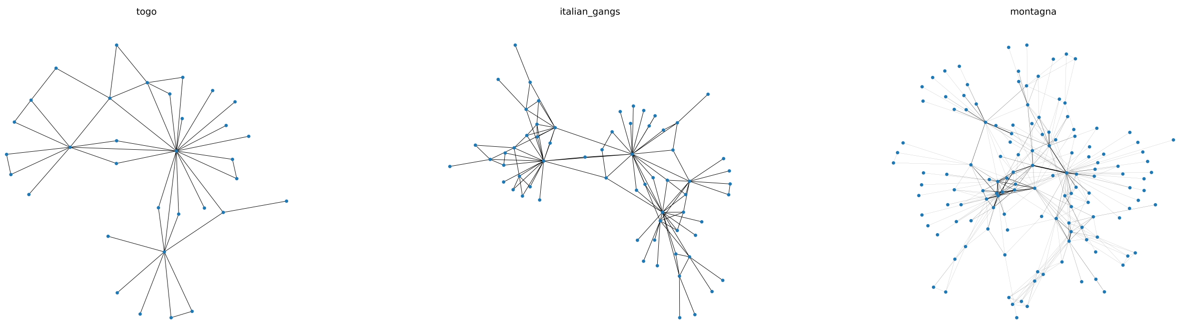


Figure 5

Finally, in Figure 6, we compare the efficacy of different strategies across each of the networks. The naive random strategy serves as a baseline. First, observe the success that a balanced strategy has on the Al Qaeda network. By balancing between depth and greediness we are able to significantly improve the chances of jumping between cells and capturing the important criminals in the network. This strategy is not universally the best (though it is consistently good) and the only other network that it outperforms the other strategies by a large margin is the Ndrangheta Mafia. The balanced diameter strategy here is tuned to capturing this special case of an Al Qaeda network with specific network properties in a non-intuitive way. It is not immediately intuitive why you would forego capturing a criminal with a high likelihood of revealing a number of suspects, as opposed to a suspect who maximises the diameter of the network, but may reveal few suspects. However in the case of Al Qaeda, this is a successful strategy, because it allows the algorithm to build a bridge into another cell. This highlights the importance of law enforcement targeting their a disruption strategy to the expected characteristics of the network (of course, our algorithm does not know these ex ante).

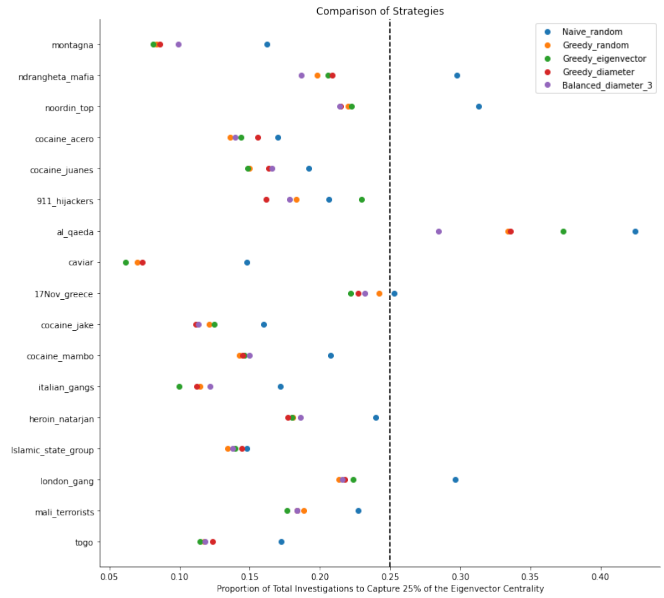


Figure 6: Comparison of strategy performance based on the proportion of investigations needed to catch the first quartile of eigenvector centrality.

V Conclusion and extensions

In this study, we built an algorithm to investigate and capture criminals and applied it to a set of real world criminal networks. Our algorithm was novel because it only received information from the network by interacting with it and, ex ante, did not know its structure. In general, the algorithm successfully captured a large proportion of the criminals, but struggled to 'mop up' the fringe nodes at the edges of the network. We showed that strategies to investigate and capture criminals based on the properties of the network universally outperformed a random baseline.

We were able to pinpoint the structural properties of a network that can help it to avoid capture and remain covert. A particular case was the Al Qaeda network, whose cell structure was very effective at avoiding capture. However, we were also able to show how a particular strategy could overcome much of the difficulty presented by the structure of the Al Qaeda network. This result highlights the importance of designing an investigation strategy that takes into account the expected structure of the network.

The flexibility of our algorithm design means that this work has significant room for extensions. For example, we only use a constant probability of capture model as baseline. However, more complex probability models that interact with the network and which are more reflective of how law enforcement and criminals interact could easily be specified. Furthermore, we made a number of simplifying assumptions on our data for computational tractability. Finding a clever way to incorporate these extra data would be helpful. Finally, there are many more networks that we could explore with our algorithm.

VI Appendix

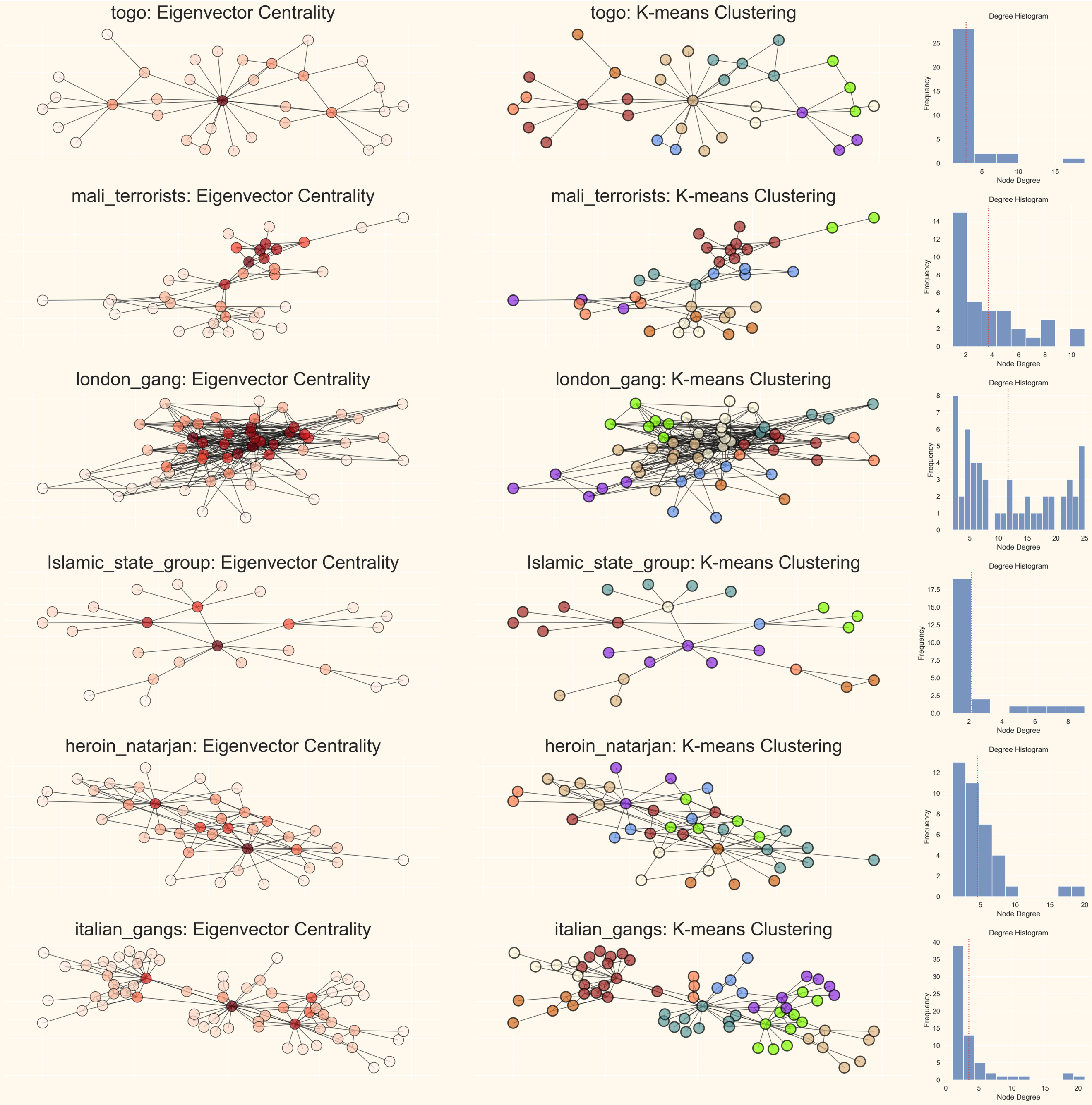


Figure A1: Criminal Networks

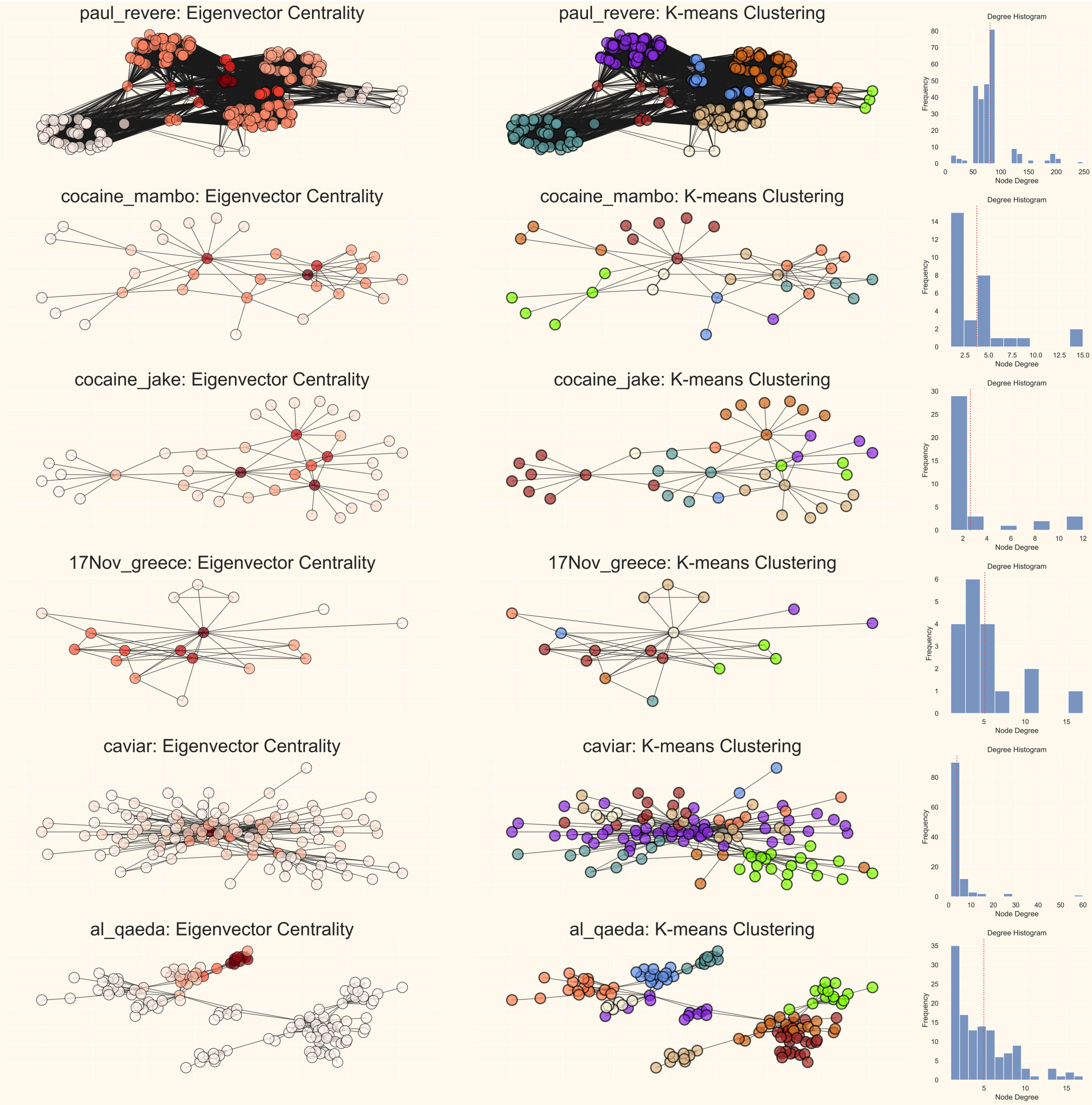


Figure A2: Criminal Networks

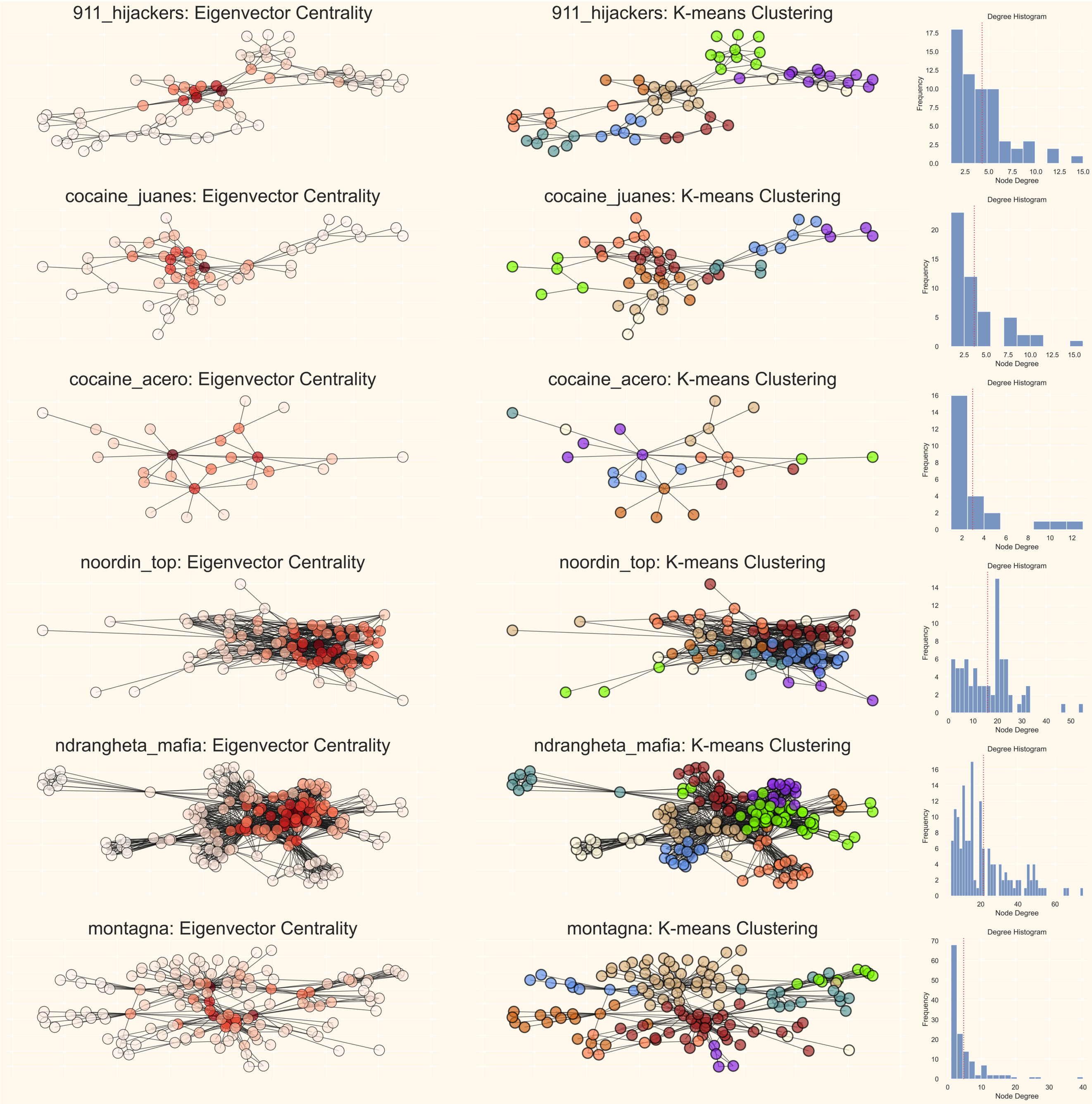


Figure A3: Criminal Networks

Table A1: Naive Random Strategy

Metrics	Proportion of Caught Criminals								Proportion of Caught Centrality								*% Incomplete*
Quantile	0.25		0.50		0.75		1.00		0.25		0.50		0.75		1.00		
Network	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	
17N	36	*17.4*	61	*22.4*	90	*26.6*	116	*29.6*	29	*17.3*	55	*23.7*	78	*26.1*	116	*29.6*	0.0
9/11 Hijackers	119	*29.9*	226	*40.9*	333	*49.7*	439	*56.1*	90	*49.4*	144	*57.5*	218	*59.8*	439	*56.1*	0.0
ISIL	59	*21.8*	113	*29.2*	170	*38.8*	229	*44.4*	33	*18.8*	80	*28.1*	141	*33.8*	229	*44.4*	0.0
Al Qaeda	224	*43.6*	441	*57.2*	647	*68.8*	859	*67.6*	364	*229.8*	452	*248.1*	511	*248.5*	859	*67.6*	5.4
Caviar	223	*42.5*	429	*54.5*	651	*66.9*	875	*68.2*	129	*42.7*	287	*54.7*	507	*65.9*	875	*68.2*	9.4
Operation Acero	55	*21.5*	104	*27.4*	153	*32.6*	211	*39.9*	36	*19.2*	77	*26.7*	124	*30.7*	211	*39.9*	0.0
Operation Jake	84	*27.8*	163	*36.0*	254	*44.0*	341	*52.8*	54	*30.4*	112	*37.4*	194	*41.7*	341	*52.8*	0.0
Operation Juanes	102	*28.2*	199	*36.6*	300	*43.7*	408	*54.9*	78	*44.3*	133	*52.9*	203	*55.5*	408	*54.9*	0.0
Operation Mambo	64	*21.6*	124	*31.0*	184	*34.6*	242	*43.0*	50	*23.9*	98	*34.5*	150	*38.2*	242	*43.0*	0.0
Heroin Natarjan	81	*26.2*	143	*33.3*	206	*37.6*	272	*44.4*	65	*25.3*	123	*33.1*	183	*36.9*	272	*44.4*	0.0
Italian Gangs	139	*34.2*	270	*47.9*	396	*57.5*	532	*67.9*	91	*34.8*	198	*46.5*	328	*55.0*	532	*67.9*	0.0
London Gangs	81	*21.9*	127	*23.9*	173	*26.5*	228	*30.5*	67	*24.2*	107	*25.3*	144	*26.6*	228	*30.5*	0.0
Mali Terrorists	67	*23.2*	136	*32.5*	202	*37.9*	279	*47.8*	63	*35.8*	110	*48.7*	161	*47.3*	279	*47.8*	0.0
Montagna Operation	252	*44.5*	485	*58.1*	723	*69.4*	929	*51.3*	150	*48.7*	310	*62.1*	527	*68.6*	929	*51.3*	45.2
Ndrangheta	161	*27.7*	249	*30.6*	333	*32.5*	429	*37.1*	127	*34.1*	195	*36.7*	267	*36.0*	429	*37.1*	0.0
Malasyian Extremists	104	*23.7*	155	*25.6*	202	*26.6*	280	*35.6*	87	*25.2*	135	*27.6*	176	*27.4*	280	*35.6*	0.0
Project Togo	74	*24.1*	143	*34.0*	209	*41.2*	277	*48.8*	47	*23.3*	108	*30.5*	182	*38.7*	277	*48.8*	0.0

Table A2: Greedy Random Strategy

Metrics	Proportion of Caught Criminals								Proportion of Caught Centrality								*% Incomplete*
Quantile	0.25		0.50		0.75		1.00		0.25		0.50		0.75		1.00		
Network	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	
17N	30	*14.1*	47	*16.3*	75	*20.4*	106	*25.5*	25	*15.6*	39	*17.0*	52	*17.9*	106	*25.5*	0.0
9/11 Hijackers	98	*25.5*	181	*32.9*	290	*42.7*	412	*52.4*	75	*44.7*	96	*46.7*	134	*45.2*	412	*52.4*	0.0
ISIL	54	*20.5*	111	*29.3*	171	*37.5*	231	*45.0*	31	*18.1*	57	*23.1*	139	*34.3*	231	*45.0*	0.0
Al Qaeda	176	*31.9*	381	*51.6*	569	*59.6*	833	*69.7*	278	*186.5*	444	*217.5*	458	*214.2*	833	*69.7*	1.8
Caviar	142	*27.8*	311	*39.1*	577	*63.7*	838	*74.5*	58	*22.9*	125	*27.2*	287	*39.7*	838	*74.5*	3.6
Operation Acero	45	*17.9*	83	*23.5*	141	*33.0*	200	*41.0*	27	*17.7*	50	*20.7*	80	*24.3*	200	*41.0*	0.0
Operation Jake	67	*21.7*	144	*32.6*	242	*42.3*	334	*53.7*	40	*26.0*	58	*27.4*	155	*35.3*	334	*53.7*	0.0
Operation Juanes	80	*25.3*	157	*30.6*	264	*39.9*	385	*52.4*	57	*37.8*	82	*39.0*	116	*39.3*	385	*52.4*	0.0
Operation Mambo	47	*17.1*	90	*22.2*	153	*29.5*	225	*38.7*	32	*18.2*	57	*20.7*	88	*23.3*	225	*38.7*	0.0
Heroin Natarjan	58	*19.2*	100	*24.1*	158	*28.9*	239	*38.4*	43	*19.9*	73	*22.2*	113	*25.6*	239	*38.4*	0.0
Italian Gangs	105	*24.9*	206	*33.5*	341	*43.6*	494	*58.0*	56	*26.8*	124	*33.1*	206	*34.3*	494	*58.0*	0.0
London Gangs	52	*15.3*	77	*15.8*	119	*18.8*	194	*26.9*	41	*15.6*	56	*16.5*	72	*16.3*	194	*26.9*	0.0
Mali Terrorists	55	*19.7*	118	*29.0*	177	*33.8*	265	*42.8*	50	*27.4*	89	*47.7*	120	*39.5*	265	*42.8*	0.0
Montagna Operation	154	*26.3*	323	*36.6*	603	*56.4*	906	*59.7*	75	*30.2*	127	*27.8*	241	*30.6*	906	*59.7*	18.4
Ndrangheta	82	*17.9*	129	*17.7*	208	*19.1*	328	*27.3*	65	*21.6*	84	*21.6*	109	*21.1*	328	*27.3*	0.0
Malasyian Extremists	59	*14.9*	91	*15.4*	131	*16.1*	226	*27.8*	49	*17.5*	69	*17.5*	91	*16.9*	226	*27.8*	0.0
Project Togo	62	*20.3*	127	*31.4*	197	*38.7*	272	*48.8*	32	*16.1*	84	*26.9*	152	*34.6*	272	*48.8*	0.0

Table A3: Greedy Diameter Strategy

Metrics	Proportion of Caught Criminals								Proportion of Caught Centrality								*% Incomplete*
Quantile	0.25		0.50		0.75		1.00		0.25		0.50		0.75		1.00		
Network	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	
17N	36	*17.4*	61	*22.4*	90	*26.6*	116	*29.6*	29	*17.3*	55	*23.7*	78	*26.1*	116	*29.6*	0.0
9/11 Hijackers	119	*29.9*	226	*40.9*	333	*49.7*	439	*56.1*	90	*49.4*	144	*57.5*	218	*59.8*	439	*56.1*	0.0
ISIL	59	*21.8*	113	*29.2*	170	*38.8*	229	*44.4*	33	*18.8*	80	*28.1*	141	*33.8*	229	*44.4*	0.0
Al Qaeda	224	*43.6*	441	*57.2*	647	*68.8*	859	*67.6*	364	*229.8*	452	*248.1*	511	*248.5*	859	*67.6*	5.4
Caviar	223	*42.5*	429	*54.5*	651	*66.9*	875	*68.2*	129	*42.7*	287	*54.7*	507	*65.9*	875	*68.2*	9.4
Operation Acero	55	*21.5*	104	*27.4*	153	*32.6*	211	*39.9*	36	*19.2*	77	*26.7*	124	*30.7*	211	*39.9*	0.0
Operation Jake	84	*27.8*	163	*36.0*	254	*44.0*	341	*52.8*	54	*30.4*	112	*37.4*	194	*41.7*	341	*52.8*	0.0
Operation Juanes	102	*28.2*	199	*36.6*	300	*43.7*	408	*54.9*	78	*44.3*	133	*52.9*	203	*55.5*	408	*54.9*	0.0
Operation Mambo	64	*21.6*	124	*31.0*	184	*34.6*	242	*43.0*	50	*23.9*	98	*34.5*	150	*38.2*	242	*43.0*	0.0
Heroin Natarjan	81	*26.2*	143	*33.3*	206	*37.6*	272	*44.4*	65	*25.3*	123	*33.1*	183	*36.9*	272	*44.4*	0.0
Italian Gangs	139	*34.2*	270	*47.9*	396	*57.5*	532	*67.9*	91	*34.8*	198	*46.5*	328	*55.0*	532	*67.9*	0.0
London Gangs	81	*21.9*	127	*23.9*	173	*26.5*	228	*30.5*	67	*24.2*	107	*25.3*	144	*26.6*	228	*30.5*	0.0
Mali Terrorists	67	*23.2*	136	*32.5*	202	*37.9*	279	*47.8*	63	*35.8*	110	*48.7*	161	*47.3*	279	*47.8*	0.0
Montagna Operation	252	*44.5*	485	*58.1*	723	*69.4*	929	*51.3*	150	*48.7*	310	*62.1*	527	*68.6*	929	*51.3*	45.2
Ndrangheta	161	*27.7*	249	*30.6*	333	*32.5*	429	*37.1*	127	*34.1*	195	*36.7*	267	*36.0*	429	*37.1*	0.0
Malasyian Extremists	104	*23.7*	155	*25.6*	202	*26.6*	280	*35.6*	87	*25.2*	135	*27.6*	176	*27.4*	280	*35.6*	0.0
Project Togo	74	*24.1*	143	*34.0*	209	*41.2*	277	*48.8*	47	*23.3*	108	*30.5*	182	*38.7*	277	*48.8*	0.0

Table A4: Max Diameter Strategy

Metrics	Proportion of Caught Criminals								Proportion of Caught Centrality								*% Incomplete*
Quantile	0.25		0.50		0.75		1.00		0.25		0.50		0.75		1.00		
Network	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	
17N	32	*15.5*	50	*17.9*	79	*23.2*	113	*28.4*	30	*15.8*	42	*17.7*	60	*21.5*	113	*28.4*	0.0
9/11 Hijackers	117	*30.7*	214	*40.7*	322	*52.8*	441	*58.4*	93	*44.0*	131	*52.6*	208	*73.1*	441	*58.4*	0.0
ISIL	57	*22.5*	115	*32.6*	174	*40.1*	232	*45.6*	33	*19.5*	63	*25.6*	141	*36.6*	232	*45.6*	0.0
Al Qaeda	199	*37.4*	403	*54.7*	589	*66.7*	848	*73.3*	307	*202.7*	385	*215.9*	406	*210.9*	848	*73.3*	1.4
Operation Acero	53	*22.0*	95	*29.0*	150	*36.5*	207	*43.1*	34	*18.6*	68	*26.7*	113	*40.0*	207	*43.1*	0.0
Operation Jake	76	*28.1*	152	*36.6*	250	*46.3*	338	*55.1*	54	*32.7*	82	*42.5*	190	*63.8*	338	*55.1*	0.0
Operation Juanes	98	*29.6*	183	*42.0*	291	*50.7*	399	*57.6*	71	*35.8*	115	*48.2*	182	*83.2*	399	*57.6*	0.0
Operation Mambo	58	*21.6*	111	*30.1*	176	*41.0*	242	*47.1*	46	*22.7*	87	*34.5*	145	*61.0*	242	*47.1*	0.0
Heroin Natarjan	76	*25.8*	135	*36.2*	210	*49.6*	271	*50.3*	62	*25.5*	123	*43.9*	199	*62.0*	271	*50.3*	0.0
Italian Gangs	116	*31.3*	223	*43.1*	348	*55.2*	503	*66.7*	81	*42.1*	158	*49.7*	250	*61.8*	503	*66.7*	0.0
London Gangs	84	*28.5*	133	*43.8*	194	*56.3*	241	*49.1*	78	*31.7*	130	*68.6*	187	*75.0*	241	*49.1*	0.0
Mali Terrorists	66	*23.5*	129	*33.3*	189	*37.6*	277	*46.6*	62	*33.7*	103	*45.7*	144	*44.7*	277	*46.6*	0.0
Montagna Operation	191	*35.6*	382	*54.4*	664	*69.2*	919	*52.4*	108	*36.5*	191	*52.8*	398	*157.9*	919	*52.4*	1.6
Ndrangheta	120	*23.5*	189	*31.9*	292	*52.8*	412	*52.6*	98	*22.3*	146	*35.3*	244	*102.8*	412	*52.6*	0.0
Malasyian Extremists	107	*32.7*	163	*51.0*	254	*74.1*	320	*63.1*	102	*36.3*	159	*73.5*	247	*97.6*	320	*63.1*	0.0
Project Togo	68	*23.8*	130	*33.0*	200	*40.1*	276	*47.1*	41	*21.4*	94	*30.9*	166	*40.7*	276	*47.1*	0.0

Table A5: Greedy Eigenvector Strategy

Metrics	Proportion of Caught Criminals								Proportion of Caught Centrality								*% Incomplete*
Quantile	0.25		0.50		0.75		1.00		0.25		0.50		0.75		1.00		
Network	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	
17N	28	*13.3*	45	*15.4*	75	*20.7*	106	*26.0*	23	*14.5*	36	*16.0*	51	*16.9*	106	*26.0*	0.0
9/11 Hijackers	104	*29.1*	189	*36.5*	296	*43.5*	413	*55.0*	94	*56.6*	111	*55.8*	149	*55.4*	413	*55.0*	0.0
ISIL	53	*21.0*	108	*29.8*	166	*37.0*	223	*44.0*	31	*18.6*	53	*21.9*	114	*31.5*	223	*44.0*	0.0
Al Qaeda	177	*30.4*	397	*50.4*	575	*63.3*	816	*71.4*	304	*215.5*	431	*241.0*	444	*237.7*	816	*71.4*	0.6
Caviar	137	*24.9*	301	*38.6*	577	*61.3*	848	*71.5*	52	*18.6*	119	*23.8*	258	*35.6*	848	*71.5*	2.2
Operation Acero	47	*18.2*	83	*23.2*	139	*30.5*	203	*39.8*	29	*17.3*	52	*20.4*	79	*23.7*	203	*39.8*	0.0
Operation Jake	68	*23.5*	143	*33.8*	246	*47.1*	336	*55.8*	41	*26.8*	59	*27.7*	141	*35.5*	336	*55.8*	0.0
Operation Juanes	82	*25.5*	161	*31.3*	268	*42.4*	389	*51.4*	57	*34.7*	82	*36.8*	117	*36.8*	389	*51.4*	0.0
Operation Mambo	48	*18.0*	92	*22.5*	157	*30.6*	229	*41.5*	33	*18.7*	58	*21.6*	88	*23.2*	229	*41.5*	0.0
Heroin Natarjan	58	*18.4*	102	*22.7*	157	*27.9*	235	*37.3*	42	*18.9*	75	*22.2*	112	*23.5*	235	*37.3*	0.0
Italian Gangs	103	*25.6*	201	*33.6*	343	*50.0*	501	*62.1*	49	*21.3*	126	*31.8*	197	*33.9*	501	*62.1*	0.0
London Gangs	53	*16.0*	78	*16.6*	119	*17.8*	190	*24.5*	42	*16.3*	57	*17.0*	74	*16.9*	190	*24.5*	0.0
Mali Terrorists	52	*18.4*	123	*30.6*	177	*34.4*	267	*45.6*	47	*24.1*	102	*56.5*	139	*44.0*	267	*45.6*	0.0
Montagna Operation	155	*25.7*	331	*37.0*	607	*57.2*	906	*57.4*	73	*31.3*	126	*28.7*	237	*31.8*	906	*57.4*	20.0
Ndrangheta	84	*17.9*	131	*17.3*	208	*18.5*	331	*25.5*	68	*23.0*	88	*22.8*	112	*22.1*	331	*25.5*	0.0
Malasyian Extremists	61	*15.7*	93	*16.4*	133	*17.0*	233	*31.3*	52	*17.8*	71	*17.6*	93	*17.7*	233	*31.3*	0.0
Project Togo	61	*20.5*	124	*28.6*	201	*39.4*	274	*46.4*	31	*17.3*	77	*23.8*	136	*29.8*	274	*46.4*	0.0

Table A6: Least Central Strategy

Metrics	Proportion of Caught Criminals								Proportion of Caught Centrality								*% Incomplete*
Quantile	0.25		0.50		0.75		1.00		0.25		0.50		0.75		1.00		
Network	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	
17N	30	*14.5*	48	*18.0*	77	*21.0*	108	*24.9*	24	*15.8*	39	*18.1*	59	*20.2*	108	*24.9*	0.0
9/11 Hijackers	112	*28.1*	209	*38.0*	311	*48.4*	418	*55.5*	80	*37.8*	170	*57.1*	251	*59.7*	418	*55.5*	0.0
ISIL	60	*24.6*	119	*33.8*	179	*41.5*	240	*46.9*	35	*20.4*	89	*31.2*	157	*39.2*	240	*46.9*	0.0
Al Qaeda	204	*40.3*	429	*61.1*	643	*74.3*	841	*68.4*	426	*264.4*	469	*272.4*	486	*270.4*	841	*68.4*	3.0
Caviar	180	*35.7*	374	*50.2*	607	*67.9*	864	*68.6*	96	*38.2*	216	*43.6*	471	*57.3*	864	*68.6*	9.0
Operation Acero	52	*20.2*	98	*27.7*	146	*33.9*	203	*40.7*	29	*18.8*	71	*24.6*	129	*33.3*	203	*40.7*	0.0
Operation Jake	84	*28.9*	163	*38.9*	253	*46.8*	344	*52.7*	54	*31.2*	101	*41.4*	215	*45.2*	344	*52.7*	0.0
Operation Juanes	101	*30.3*	206	*42.3*	298	*48.5*	398	*55.4*	107	*68.6*	169	*83.1*	250	*59.8*	398	*55.4*	0.0
Operation Mambo	52	*20.4*	105	*28.6*	159	*34.4*	223	*42.7*	41	*24.9*	80	*32.7*	119	*32.7*	223	*42.7*	0.0
Heroin Natarjan	70	*23.5*	120	*29.3*	177	*33.7*	252	*43.1*	50	*22.2*	95	*27.5*	149	*32.0*	252	*43.1*	0.0
Italian Gangs	124	*30.6*	253	*46.2*	370	*54.2*	505	*62.6*	93	*34.5*	224	*45.8*	347	*54.8*	505	*62.6*	0.0
London Gangs	79	*26.0*	124	*32.8*	176	*33.6*	224	*33.5*	74	*28.8*	109	*38.4*	153	*45.0*	224	*33.5*	0.0
Mali Terrorists	66	*23.9*	132	*31.7*	204	*40.2*	275	*45.8*	70	*40.8*	116	*55.0*	162	*53.3*	275	*45.8*	0.0
Montagna Operation	216	*41.6*	435	*57.4*	661	*68.8*	925	*52.5*	130	*48.1*	275	*62.8*	503	*66.9*	925	*52.5*	36.8
Ndrangheta	113	*25.5*	202	*29.6*	283	*32.4*	377	*35.9*	108	*48.1*	177	*52.3*	266	*37.2*	377	*35.9*	0.0
Malasyian Extremists	93	*26.0*	154	*29.8*	221	*33.3*	276	*37.3*	85	*26.8*	135	*30.0*	218	*34.6*	276	*37.3*	0.0
Project Togo	69	*24.7*	135	*32.9*	200	*38.8*	280	*46.0*	46	*23.8*	114	*32.2*	189	*37.9*	280	*46.0*	0.0

Table A7: Balanced Diameter (0.1) Strategy

Metrics	Proportion of Caught Criminals								Proportion of Caught Centrality								*% Incomplete*
Quantile	0.25		0.50		0.75		1.00		0.25		0.50		0.75		1.00		
Network	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	
17N	28	*13.6*	44	*15.9*	72	*18.9*	103	*24.0*	23	*14.6*	36	*16.2*	49	*16.8*	103	*24.0*	0.0
9/11 Hijackers	96	*25.6*	179	*31.4*	284	*40.0*	414	*52.0*	69	*38.5*	93	*44.2*	130	*40.5*	414	*52.0*	0.0
ISIL	56	*21.8*	113	*30.4*	173	*37.1*	233	*43.7*	33	*20.0*	63	*25.3*	141	*33.4*	233	*43.7*	0.0
Al Qaeda	173	*31.9*	374	*59.3*	559	*69.6*	827	*82.9*	257	*185.3*	419	*217.6*	433	*213.4*	827	*82.9*	0.6
Operation Acero	44	*17.0*	82	*22.0*	142	*31.9*	204	*39.5*	26	*16.3*	48	*19.0*	80	*22.8*	204	*39.5*	0.0
Operation Jake	68	*23.2*	144	*31.4*	238	*43.4*	325	*49.4*	41	*26.3*	60	*27.0*	156	*34.1*	325	*49.4*	0.0
Operation Juanes	80	*24.3*	156	*29.8*	259	*39.1*	376	*50.0*	58	*36.4*	83	*37.7*	118	*38.2*	376	*50.0*	0.0
Operation Mambo	46	*16.6*	91	*22.0*	150	*28.6*	214	*36.1*	33	*17.1*	57	*19.7*	89	*22.8*	214	*36.1*	0.0
Heroin Natarjan	58	*18.7*	102	*22.9*	159	*28.2*	239	*36.5*	44	*19.3*	74	*21.2*	115	*24.4*	239	*36.5*	0.0
Italian Gangs	105	*25.0*	204	*34.6*	341	*47.0*	497	*60.3*	58	*27.0*	123	*32.0*	205	*35.5*	497	*60.3*	0.0
London Gangs	53	*16.9*	78	*17.5*	120	*19.7*	191	*26.8*	42	*17.3*	57	*18.4*	74	*18.2*	191	*26.8*	0.0
Mali Terrorists	54	*19.1*	122	*28.5*	180	*34.4*	269	*45.0*	49	*26.3*	91	*50.6*	128	*38.6*	269	*45.0*	0.0
Ndrangheta	86	*17.7*	133	*16.7*	209	*16.4*	330	*26.7*	69	*23.0*	89	*22.8*	114	*22.3*	330	*26.7*	0.0
Malasyian Extremists	56	*14.1*	86	*15.0*	127	*15.9*	219	*28.0*	45	*15.1*	65	*15.8*	86	*16.1*	219	*28.0*	0.0
Project Togo	63	*21.4*	126	*28.0*	197	*36.8*	272	*44.8*	33	*18.3*	84	*26.7*	152	*32.1*	272	*44.8*	0.0

Table A8: Balanced Diameter (0.3) Strategy

Metrics	Proportion of Caught Criminals								Proportion of Caught Centrality								*% Incomplete*
Quantile	0.25		0.50		0.75		1.00		0.25		0.50		0.75		1.00		
Network	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	
17N	29	*14.1*	47	*16.8*	76	*19.4*	109	*24.8*	25	*16.0*	38	*17.9*	52	*18.5*	109	*24.8*	0.0
9/11 Hijackers	99	*28.4*	187	*34.4*	289	*41.4*	422	*56.1*	75	*43.4*	96	*46.7*	137	*46.9*	422	*56.1*	0.0
ISIL	53	*20.0*	111	*30.2*	172	*38.7*	232	*44.8*	32	*17.7*	58	*23.1*	139	*34.0*	232	*44.8*	0.0
Al Qaeda	169	*33.0*	373	*54.7*	552	*67.4*	821	*74.2*	233	*185.3*	393	*225.6*	407	*221.3*	821	*74.2*	0.0
Operation Acero	47	*16.8*	85	*23.6*	140	*31.0*	203	*40.6*	28	*17.5*	53	*19.1*	84	*25.2*	203	*40.6*	0.0
Operation Jake	66	*22.2*	146	*32.8*	251	*43.4*	347	*52.9*	39	*23.9*	59	*25.4*	158	*35.1*	347	*52.9*	0.0
Operation Juanes	84	*25.9*	161	*27.2*	264	*39.4*	383	*47.3*	63	*36.4*	88	*37.9*	122	*37.9*	383	*47.3*	0.0
Operation Mambo	48	*17.1*	94	*22.1*	154	*29.9*	225	*40.5*	33	*18.6*	59	*20.7*	91	*23.1*	225	*40.5*	0.0
Heroin Natarjan	62	*20.9*	108	*25.3*	162	*30.2*	244	*43.8*	45	*20.8*	78	*23.6*	120	*26.4*	244	*43.8*	0.0
Italian Gangs	106	*24.8*	205	*34.1*	344	*49.9*	502	*65.4*	61	*27.7*	124	*32.8*	206	*35.3*	502	*65.4*	0.0
London Gangs	51	*15.8*	76	*16.5*	117	*19.2*	188	*26.5*	40	*16.4*	55	*17.1*	72	*17.1*	188	*26.5*	0.0
Mali Terrorists	52	*18.5*	116	*28.0*	171	*30.7*	262	*41.1*	48	*26.2*	88	*48.3*	122	*36.9*	262	*41.1*	0.0
Montagna Operation	172	*20.1*	357	*40.1*	656	*62.0*	954	*42.3*	94	*39.9*	146	*23.3*	274	*42.9*	954	*42.3*	0.4
Ndrangheta	82	*5.5*	129	*4.6*	207	*10.7*	336	*22.8*	62	*11.1*	81	*10.4*	106	*8.2*	336	*22.8*	0.0
Malasyian Extremists	59	*14.5*	92	*14.8*	133	*15.7*	236	*29.1*	50	*17.3*	70	*17.3*	93	*16.5*	236	*29.1*	0.0
Project Togo	63	*21.0*	129	*30.2*	201	*36.6*	278	*43.7*	32	*16.8*	86	*26.5*	157	*35.0*	278	*43.7*	0.0

Table A9: Balanced Diameter (0.5) Strategy

Metrics	Proportion of Caught Criminals								Proportion of Caught Centrality								*% Incomplete*
Quantile	0.25		0.50		0.75		1.00		0.25		0.50		0.75		1.00		
Network	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	
17N	30	*13.9*	47	*16.4*	77	*20.7*	109	*25.9*	26	*15.2*	39	*17.4*	54	*18.2*	109	*25.9*	0.0
9/11 Hijackers	102	*26.1*	190	*30.2*	291	*36.5*	417	*47.9*	77	*41.8*	95	*45.5*	146	*36.0*	417	*47.9*	0.0
ISIL	57	*21.2*	115	*31.5*	173	*37.2*	231	*45.2*	34	*18.5*	63	*24.4*	143	*35.2*	231	*45.2*	0.0
Al Qaeda	175	*30.8*	375	*45.2*	549	*52.7*	819	*72.1*	277	*179.1*	386	*186.6*	399	*185.4*	819	*72.1*	0.2
Operation Acero	52	*15.7*	89	*20.4*	142	*28.2*	207	*38.8*	34	*15.1*	59	*18.4*	92	*20.2*	207	*38.8*	0.0
Operation Jake	69	*23.0*	140	*31.0*	240	*43.3*	331	*53.7*	45	*27.2*	64	*27.7*	153	*34.8*	331	*53.7*	0.0
Operation Juanes	82	*25.9*	160	*31.2*	264	*39.3*	382	*53.8*	62	*39.9*	86	*42.3*	120	*40.3*	382	*53.8*	0.0
Operation Mambo	51	*18.9*	95	*23.6*	155	*30.4*	223	*41.1*	37	*21.0*	63	*22.5*	95	*25.3*	223	*41.1*	0.0
Heroin Natarjan	63	*19.9*	108	*25.2*	163	*30.9*	237	*39.7*	47	*20.2*	80	*23.2*	121	*27.0*	237	*39.7*	0.0
Italian Gangs	108	*26.7*	209	*34.9*	342	*46.9*	504	*59.4*	62	*29.2*	134	*35.6*	221	*37.6*	504	*59.4*	0.0
London Gangs	53	*15.7*	77	*16.2*	119	*18.6*	189	*25.8*	42	*16.3*	57	*17.2*	73	*16.9*	189	*25.8*	0.0
Mali Terrorists	57	*19.1*	124	*27.4*	183	*32.9*	280	*46.2*	52	*26.8*	92	*48.9*	130	*36.8*	280	*46.2*	0.0
Montagna Operation	148	*21.7*	315	*36.5*	607	*81.5*	912	*49.5*	83	*34.3*	125	*22.3*	243	*34.2*	912	*49.5*	0.4
Ndrangheta	81	*17.0*	130	*15.6*	209	*17.6*	329	*26.4*	61	*19.4*	82	*19.2*	107	*18.2*	329	*26.4*	0.0
Malasyian Extremists	60	*13.7*	92	*13.2*	132	*14.3*	227	*25.7*	50	*16.1*	70	*15.8*	92	*14.7*	227	*25.7*	0.0
Project Togo	65	*22.0*	129	*30.9*	201	*40.0*	280	*48.1*	36	*19.5*	89	*28.7*	160	*35.6*	280	*48.1*	0.0

Table A10: Balanced Diameter (0.7) Strategy

Metrics	Proportion of Caught Criminals								Proportion of Caught Centrality								*% Incomplete*
Quantile	0.25		0.50		0.75		1.00		0.25		0.50		0.75		1.00		
Network	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	
17N	33	*16.0*	49	*18.3*	77	*21.9*	110	*27.9*	31	*16.6*	42	*18.7*	55	*19.5*	110	*27.9*	0.0
9/11 Hijackers	113	*31.9*	199	*35.3*	297	*42.6*	426	*54.2*	92	*41.0*	117	*47.0*	167	*46.3*	426	*54.2*	0.0
ISIL	58	*23.8*	114	*33.7*	175	*39.3*	235	*45.8*	34	*20.0*	65	*28.1*	142	*37.6*	235	*45.8*	0.0
Al Qaeda	191	*35.0*	398	*54.7*	565	*59.7*	826	*73.4*	316	*183.6*	405	*188.5*	417	*187.7*	826	*73.4*	0.0
Operation Acero	53	*21.7*	93	*26.3*	145	*31.7*	207	*42.5*	33	*17.1*	65	*24.2*	100	*26.8*	207	*42.5*	0.0
Operation Jake	71	*22.8*	144	*32.5*	241	*43.7*	333	*52.5*	49	*25.8*	72	*31.7*	156	*35.2*	333	*52.5*	0.0
Operation Juanes	100	*27.9*	177	*32.4*	280	*41.8*	390	*48.2*	75	*36.6*	111	*37.5*	147	*35.6*	390	*48.2*	0.0
Operation Mambo	61	*23.1*	109	*29.6*	166	*33.8*	237	*40.5*	49	*23.1*	81	*30.9*	117	*34.7*	237	*40.5*	0.0
Heroin Natarjan	75	*24.5*	123	*31.7*	177	*35.5*	249	*41.3*	61	*24.1*	107	*37.0*	145	*41.9*	249	*41.3*	0.0
Italian Gangs	117	*28.1*	221	*38.0*	355	*49.2*	514	*64.0*	76	*33.8*	148	*37.0*	235	*39.9*	514	*64.0*	0.0
London Gangs	82	*30.1*	106	*31.9*	146	*31.3*	213	*34.7*	77	*32.6*	92	*34.9*	106	*33.7*	213	*34.7*	0.0
Mali Terrorists	64	*22.9*	127	*30.1*	184	*34.4*	270	*44.8*	63	*30.3*	105	*47.3*	141	*42.9*	270	*44.8*	0.0
Montagna Operation	156	*31.9*	334	*39.7*	602	*32.6*	906	*35.6*	97	*31.5*	142	*35.2*	260	*45.6*	906	*35.6*	0.2
Ndrangheta	104	*13.2*	149	*14.5*	222	*20.0*	349	*40.7*	88	*17.3*	107	*17.6*	130	*17.2*	349	*40.7*	0.0
Malasyian Extremists	101	*31.8*	131	*31.7*	167	*30.8*	261	*35.5*	96	*34.0*	113	*33.6*	135	*32.6*	261	*35.5*	0.0
Project Togo	68	*23.4*	131	*31.9*	201	*40.5*	277	*47.3*	42	*22.4*	94	*30.1*	165	*41.2*	277	*47.3*	0.0

Table A11: Balanced Diameter (0.9) Strategy

Metrics	Proportion of Caught Criminals								Proportion of Caught Centrality								*% Incomplete*
Quantile	0.25		0.50		0.75		1.00		0.25		0.50		0.75		1.00		
Network	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	
17N	32	*16.0*	49	*19.2*	75	*21.9*	107	*28.2*	29	*16.3*	41	*18.9*	58	*21.4*	107	*28.2*	0.0
9/11 Hijackers	115	*30.2*	213	*38.2*	313	*47.1*	440	*55.6*	92	*44.2*	129	*57.9*	199	*71.3*	440	*55.6*	0.0
ISIL	58	*21.0*	117	*32.7*	178	*42.0*	237	*48.5*	34	*18.2*	64	*24.3*	146	*38.2*	237	*48.5*	0.0
Al Qaeda	196	*38.3*	402	*56.9*	579	*66.3*	823	*74.1*	316	*193.8*	421	*204.7*	447	*198.5*	823	*74.1*	0.2
Caviar	158	*48.3*	339	*93.0*	602	*94.6*	865	*68.5*	88	*56.5*	201	*159.6*	363	*166.2*	865	*68.5*	0.4
Operation Acero	52	*19.8*	93	*25.2*	149	*35.8*	211	*39.5*	32	*16.8*	67	*24.1*	111	*37.3*	211	*39.5*	0.0
Operation Jake	75	*26.4*	158	*39.9*	258	*52.4*	348	*61.3*	52	*30.9*	85	*48.7*	196	*69.1*	348	*61.3*	0.0
Operation Juanes	96	*30.3*	182	*41.0*	288	*53.9*	397	*65.2*	73	*38.6*	113	*42.9*	179	*80.9*	397	*65.2*	0.0
Operation Mambo	61	*23.3*	114	*33.2*	180	*45.3*	242	*49.8*	51	*24.6*	93	*39.7*	156	*73.1*	242	*49.8*	0.0
Heroin Natarjan	82	*28.0*	140	*36.7*	210	*46.1*	272	*47.7*	68	*27.5*	128	*42.1*	198	*56.0*	272	*47.7*	0.0
Italian Gangs	118	*32.1*	223	*42.2*	351	*50.6*	508	*63.0*	80	*41.8*	155	*48.5*	245	*55.4*	508	*63.0*	0.0
London Gangs	85	*29.3*	128	*42.0*	166	*45.8*	225	*41.9*	79	*31.8*	120	*58.4*	141	*57.6*	225	*41.9*	0.0
Mali Terrorists	65	*23.7*	127	*31.0*	188	*36.7*	276	*46.7*	63	*33.2*	102	*45.6*	144	*42.0*	276	*46.7*	0.0
Malasyian Extremists	103	*25.8*	168	*57.9*	209	*57.9*	291	*46.4*	95	*24.6*	155	*73.7*	183	*72.2*	291	*46.4*	0.0
Project Togo	68	*24.0*	133	*32.1*	203	*40.4*	275	*46.0*	42	*21.7*	97	*31.4*	168	*40.7*	275	*46.0*	0.0

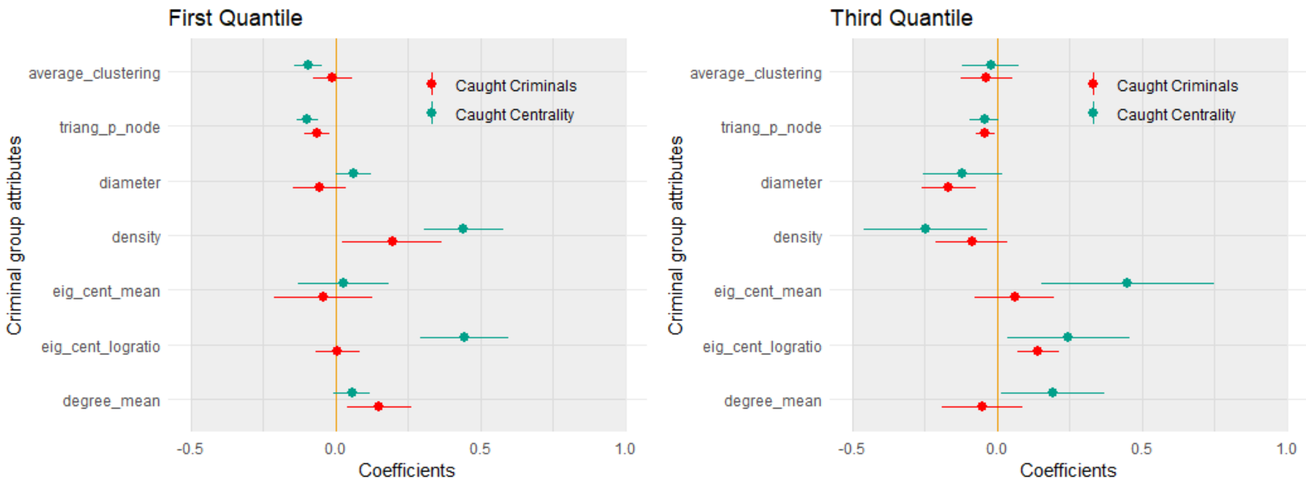


Figure A4: The effect of graph characteristics on investigation effectiveness

The figures report coefficients for the number of scaled simulations it takes to capture a quarter of a criminal network (across the first and third quantiles). The effects are in standard deviation units, and strategies are included as fixed effects. Bars depict the 95 percent confidence intervals calculated from standard errors.

References

Carley, K. M., Lee, J.-S., and Krackhardt, D. (2002). Destabilizing networks. *Connections*, 24(3):79–92.

Crenshaw, M. (2010). Mapping terrorist organizations. *Unpublished working paper.*

Ficara, A., Curreri, F., Cavallaro, L., De Meo, P., Fiumara, G., Bagdasar, O., and Liotta, A. (2021). Social network analysis: the use of graph distances to compare artificial and criminal networks. *Journal of Smart Environments and Green Computing*, 1(3):159–172.

Harris-Hogan, S. (2012). Australian neo-jihadist terrorism: Mapping the network and cell analysis using wiretap evidence. *Studies in Conflict & Terrorism*, 35(4):298–314.

Helfstein, S. and Wright, D. (2011). Covert or convenient? evolution of terror attack networks. *Journal of Conflict Resolution*, 55(5):785–813.

Koschade, S. (2006). A social network analysis of jemaah islamiyah: The applications to counterterrorism and intelligence. *Studies in Conflict & Terrorism*, 29(6):559–575.

Lauchs, M., Keast, R., and Yousefpour, N. (2011). Corrupt police networks: uncovering hidden relationship patterns, functions and roles. *Policing & society*, 21(1):110–127.

Martins, A. F. (2021). Dismantling and modeling cocaine trafficking networks. *Blog post.*

Morselli, C., Giguère, C., and Petit, K. (2007). The efficiency/security trade-off in criminal networks. *Social networks*, 29(1):143–153.

Oliver, K., Crossley, N., Edwards, G., Koskinen, J., Everett, M., and Broccatelli, C. (2014). Covert networks: structures, processes and types. *Unpublished manuscript, University of Manchester, Manchester, UK*, pages 4–13.

Varese, F. (2010). *Organized crime: Critical concepts in criminology.* Routledge.