Written Report – 6.419x Module 3

Name: ekanshupadhyay

Problem 1
Suggesting Similar Papers

Part (c)

How does the time complexity of your solution involving matrix multiplication in part (a) compare to your friend's algorithm?

First look at each solution

```
def friends_algorithm(A):
"""
Friends method of creating co-citation
matrix from adjacency matrix.

:param A: adjacency matrix
:type A: 2d numpy array
:return : co-citation matrix
:rtype : 2d numpy array
"""
n = A.shape[0]
C = np.zeros(shape=A.shape, dtype=int)

# Go through the rows of A one by one
for i in range(n):
row = A[i, :]

# Find pairs in each row
for j in range(n):
if row[j] > 0:
for k in range(j + 1, n):
if row[k] > 0:
C[j, k] += 1
C[k, j] += 1

return C
```

```
def my_solution(A):
"""
Creating co-citation matrix from
adjacency matrix using matrix
multiplication. Unlike the lecture
exercise, we zero out the diagonal so
the two methods have the same output.

:param A: adjacency matrix
:type A: 2d numpy array
:return : co-citation matrix
:rtype : 2d numpy array
"""
n = A.shape[0]
C = A.T @ A

# Zero out diagonal.
for i in range(n):
C[i, i] = 0

return C
```

In the worst situation, there would be $(n\ 2) = (n)(n-1)/2$ possible pairs if a paper cited every other paper. This equals $O(n^2)$ for an arbitrarily big n. The friend's algorithm is $O(n^3)$ overall since it depends on the order of n2 co-citation pairings for each of the n citing publications.
When the first matrix is transposed, matrix multiplication determines $\vec{aj} \cdot \vec{aj} = (E)^{n}_{i=1}\, a_{ij}^2$ for each cell of the final matrix. There are n^2 cells and this operation has a size of $O(n)$. Overall, it is $O(n^3)$.
Matrix multiplication, however, could be quicker because of vector optimizations in the hardware instruction set and the interpreter's potential ability to look ahead and anticipate repetitive instructions to distribute across numerous threads.

Part (d)

Bibliographic coupling and co-citation can both be taken as an indicator that papers deal with related material. However, they can in practice give noticeably different results. Why? Which measure is more appropriate as an indicator for similarity between papers?

The complement of authorities and hubs can be compared to bibliographic coupling and co-citations, respectively. One paper may be regarded as an authority if other papers cite it. A document may be deemed a hub if it cites several other

papers. Nevertheless, the publications that were paired and co-cited would not be the authorities or hubs, but rather their byproducts.
This comparison shows that they are, in some ways, in opposition to one another. Several sources may be cited in papers to support a single argument, as in a discussion. However it's possible that many more citations support various arguments, therefore they aren't all necessary. Hence, while the general content of co-cited publications may be somewhat similar, they may not all support the same thesis.
In contrast, a paper by itself often makes claims regarding a single subject. So, if this publication is cited in many places, at least some of those sources are discussing the same subject. Hence, a paper with a bibliography should include at least one specific part with arguments that are comparable. Other parts may be similar as they produce additional connections.
The use of bibliographic coupling as a measure of similarity should be preferred.

Problem 2
Investigating a time-varying criminal network

Part (c)
Observe the plot you made in Part (a) Question 1. The number of nodes increases sharply over the first few phases then levels out. Comment on what you think may be causing this effect. Based on your answer, should you adjust your conclusions in Part (b) Question 5?

Wiretaps on well-known network leaders were the first step in the inquiry. It is logical to suppose that the investigation grew as fresh evidence came in, drastically extending with further wiretaps to the periphery actors until each of the core players had been covered. If the assumption in Part (b) Question 5 is true, many of the actors may have consistent involvement but weren't known at the start of the inquiry; in that case, it wouldn't make sense to take the mean throughout every phase. Part (b) Question 5 finds the mean with reference to temporal consistency. But, because the network isn't as stable, we couldn't reliably infer a player's significance from one phase to another.

Part (d)
In the context of criminal networks, what would each of these metrics teach you about the importance of an actor's role in the traffic? In your own words, could you explain the limitations of degree centrality? In your opinion, which one would be most relevant to identify who is running the illegal activities of the group? Please justify.

Degree centrality is straightforward to display, while betweenness and eigenvector centralities need a little more effort. Using my instincts, I created a graph to emphasize these distinctions.

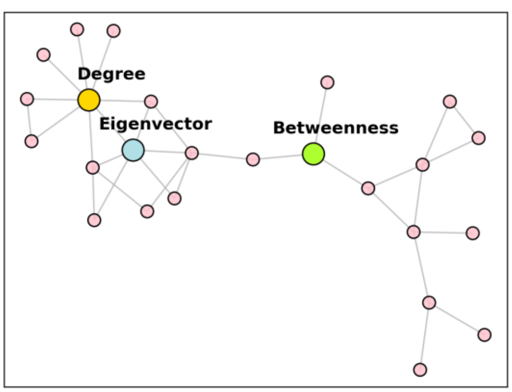


Figure 2.1: Difference between the centralities.

One should be aware that a node may have a high degree yet not appear to be the network's hub or to be closely coupled, for example if all of its nodes are leaves. The manager of a small-scale drug enterprise may theoretically interact with more individuals in a criminal network than a kingpin who uses middlemen.
Betweenness centrality, which counts the number of shortest routes that connect two points in a network, is undoubtedly more influenced by clusters than by the other two.
As there must be an equal number of shortest routes on each side, the node with the highest betweenness centrality must be located somewhat in the middle of the graph. Assume that from Part (b) Question 5, the second-highest node was a middleman for the importation of cocaine between the Serero group and the Colombians. That is what the betweenness identifies—possibly the individuals who link two closely connected networks together rather not necessarily the top player in a tight network.
By evaluating a node's relative connectivity by propagating the degrees of all of its neighbors, neighbors of those neighbors, and so on, eigenvector centrality extends degree centrality.
The iterations make it difficult to read the math after a node's weight starts to influence itself, but the graph above shows that it tends to prefer nodes that are close to nodes with high degrees. As a result, the eigenvector centrality is most suited

for locating the head of operations—that is, the player who is coordinating with several other groups of people—rather than the kingpin.

Part (e)

In real life, the police need to effectively use all the information they have gathered, to identify who is responsible for running the illegal activities of the group. Armed with a qualitative understanding of the centrality metrics from Part (d) and the quantitative analysis from part Part (b) Question 5, integrate and interpret the information you have to identify which players were most central (or important) to the operation.

Although we couldn't drive without a steering wheel or use the internet without an interface, the engine that serves as the hub of all operations and circulates air throughout the company remains its heart. Maybe networking and sustaining collaboration amongst all the components of a criminal organization with their own self interests is the organization's greatest accomplishment. Eigen- vector centrality is the centrality metric that best expresses this.

Ernesto Morales (n12), a liaison to Colombian cartels, is eliminated by the mean eigenvalue centrality in favor of Wallace Lee, an accountant (n85).

Gabrielle Casale (n76) and Bruno de Quinzio (n8), who are in charge of recovering the marijuana, are equally favored by all three measures. Alain Levy (n83), a financier and money mover, is identified by eigenvector centrality as a key actor, but not by betweenness. While his brother Grerard (n86) is more quiet, it's noteworthy to observe that he speaks with the boss considerably more regularly. None of the three metrics of centrality adequately account for edge weight. West End Gang leader Daniel Serero (n1) and his right-hand man Pierre Perlini are at the top of this list (n3). The VP of a financial firm is among the half of the non-trackers on this list, which is interesting[1]. Obviously, cash is just as essential as medicines.

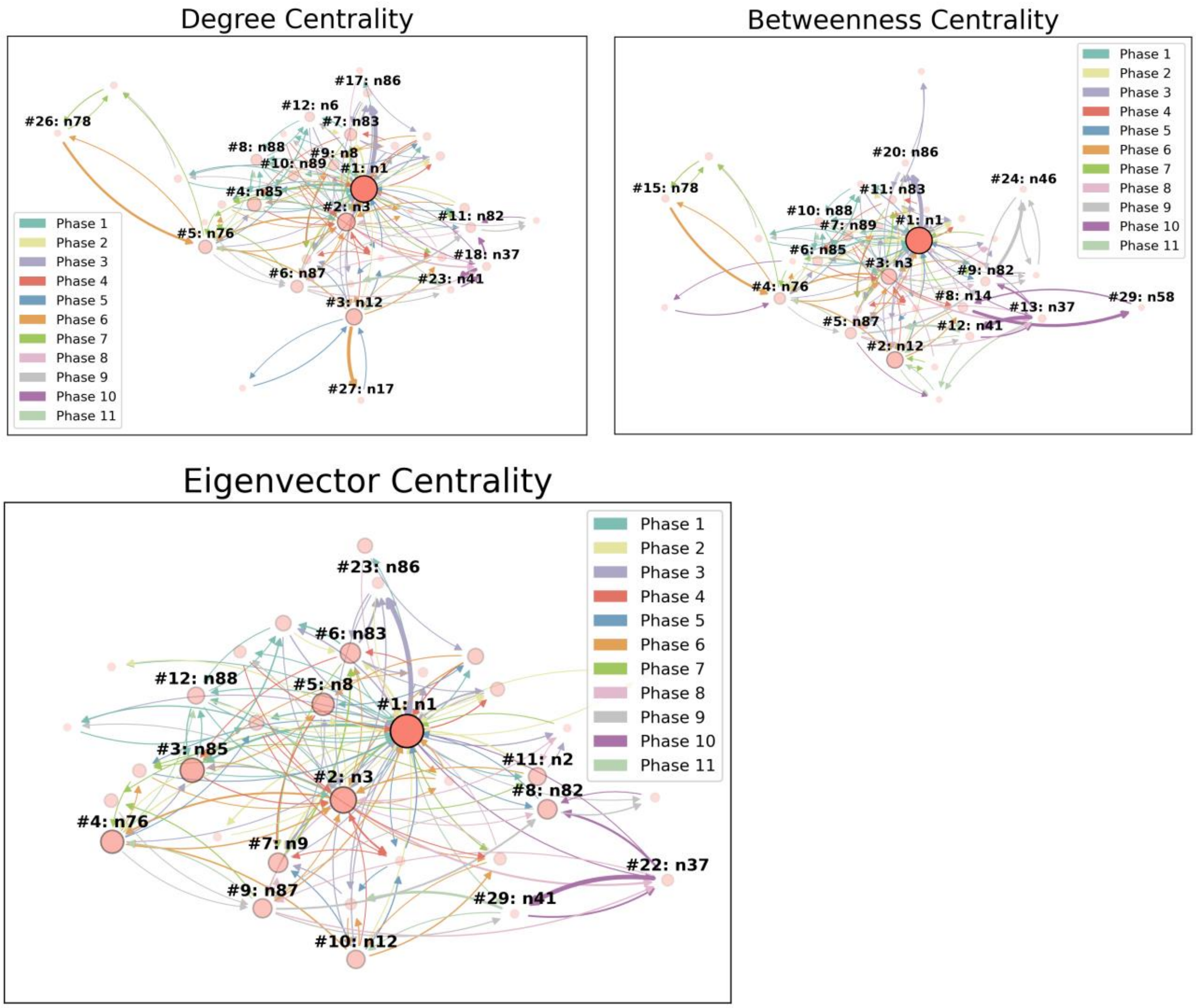


Figure 2.2: Phase at first detected communication and centrality rankings.

Part (f)

The change in the network from Phase 4 to 5 coincides with a major event that took place during the actual investigation. Identify the event and explain how the change in centrality rankings and visual patterns, observed in the network plots above, relates to said event.

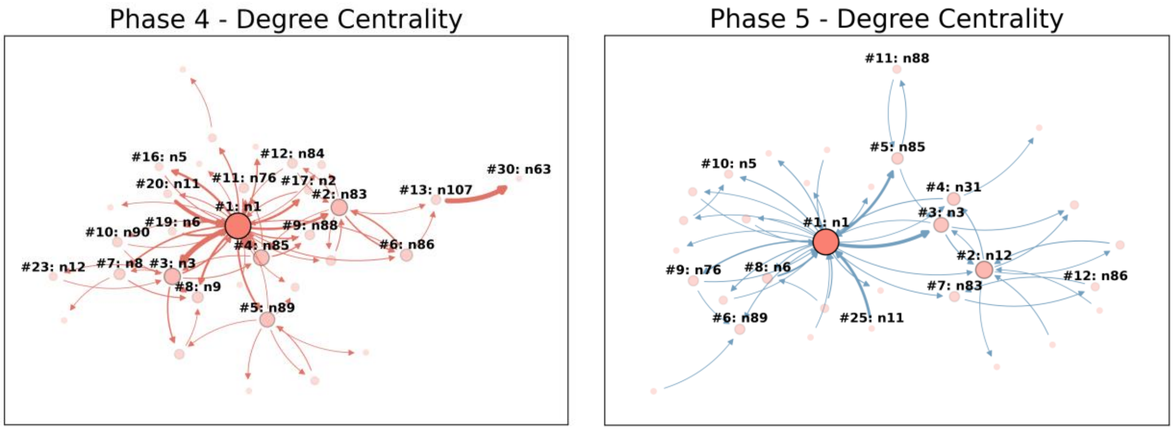


Figure 2.3: Degree centrality before and after phase 5. Top 12 ranks labeled, along with players with high levels of communication.

Although Daniel Serero (n1) was first discovered to have contacted Ernesto Morales (n12) in phase 2, it was Serero who primarily communicated with marijuana brokers Gaspar Lino (n6) and Samir Rabbat (n11) over the course of the investigation, via the accountant Wallace Lee (n85) and investor Alain Levy (n83). After establishing a channel of communication with Morales in phase 4, Pierre Perlini (n3) took over as Serero's primary cocaine contact. Serero handled the marijuana aspect of the business throughout the remaining phases and communicated more often with recovering addicts like Richard Gleeson (n5) and Gabrielle Casale (n76). Phase 4 saw the first 300kg of cannabis being seized, which is likely why the company looked to diversify or realign its operations which increased Perlini's notoriety.
Despite this, comparing centrality rankings within the assignment is inaccurate since the pattern is not as clear-cut (phase 9 is a bit of an exception). Hub and authority ratings indicate that phase 6's tale is more complicated. When considering betweenness, Alain and Gerard Levy (n86), who introduced Lino in the first two stages, became far from one another, and via Morales, both the investigation's wiretapping and the company's network grew. In general, centrality rankings revealed the change at position #2 from Lee as the primary contact for marijuana to Morales as the primary planner of the cocaine import. As a result, Serero handed the communication initiative to Perlini, shifting the burden of accountability.

	n1	n3	n6	n12	n83
1.	0.91	0.00	0.01	N/A	0.04
2.	0.94	0.002	0.00	0.00	0.09
3.	0.83	0.10	0.03	0.00	0.05
4.	0.84	0.09	0.00	0.00	0.08
5.	0.88	0.04	0.00	0.27	0.06
6.	0.54	0.23	0.00	0.38	0.00
7.	0.59	0.07	0.00	0.02	0.00
8.	0.55	0.31	0.00	0.36	0.00
9.	0.25	0.58	0.00	0.36	0.00
10.	0.34	0.00	0.00	0.03	0.03
11.	0.53	0.00	0.00	0.43	0.00

Figure 2.4: Five important nodes at every phase.

Part (g)

While centrality helps explain the evolution of every player's role individually, we need to explore the global trends and incidents in the story in order to understand the behavior of the criminal enterprise. Describe the coarse pattern(s) you observe as the network evolves through the phases. Does the network evolution reflect the background story?

The primary contacts between the European marijuana trade, notably broker Gaspar Lino (n6), and investors and money carriers Alain (n83) and Gerard Levy (n86), who often maintained in touch with VP of prestigious accountant rm Wallace Lee[1] (n85).
They comprised the Serero (n1) organization's main marijuana operation together with the drug traffickers (n5, n8, and n76). The first person to get in touch with Serero was Ernesto Morales (n12) (see phase 2), and after the first seizure at the conclusion of phase 4, Serero most likely delegated the job to his right-hand man Pierre Perlini (n3).
As a result, the organisation had mostly switched its focus to the cocaine trafficking by phase 6. In place of Serero, who was initially making the calls, Perlini took over (see the hub score in phase 6), with Serero managing the leftovers of the marijuana operation. After three seizures in phase 6, the group decided to focus on marijuana instead of Morales, but phase 7 witnessed another seizure, so that too didn't last. By phase 8, a more balanced organisation had reemerged, and it had contacted an unidentified middleman in order to reestablish communication with the cocaine importer Morales.
Phase 9 saw a significant reorganisation: Patrick Lee (n87) engaged in extensive communication with unidentified individuals n41 and n37, who would subsequently act as Morales' go-betweens, taking up Perlini's position in the subsequent stages. Before leaving, Perlini kept in close communication with the investors.
Intriguingly, despite these narcotics seizures, Daniel Serero reportedly only spent three years in prison and amassed $55 million!

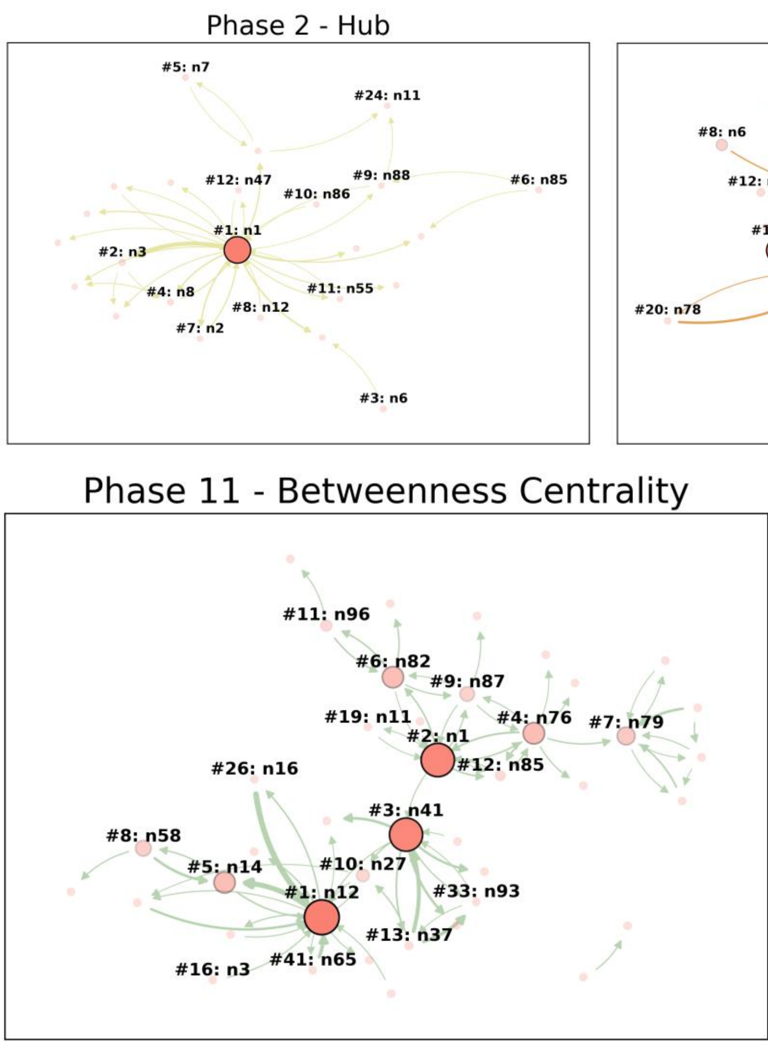


Figure 2.5: Some important phases.

Part (h)

Are there other actors that play an important role but are not on the list of investigation (i.e., actors who are not among the 23 listed above)? List them, and explain why they are important.

Several players who are not among the 23 mentioned players have high centrality rankings or often interacted with significant players. In practically every phase, there were n9 and n19, who often spoke with Serero (n1), Perlini (n3), marijuana recoveries (n5, n8, n76), and cocaine importer Morales (n12). In phase 8, there was n14, who assisted in reestablishing touch with Morales. There was n41, who connected Serero and Morales to the transportation arrangements manager Salvatore Panetta in phase 10, and n37, who did the same in phase 10. (n82).

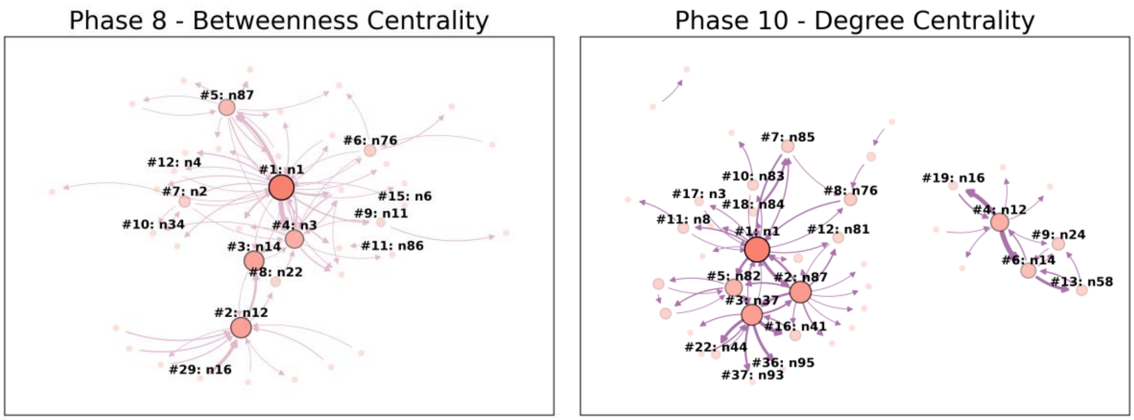


Figure 2.6: Highlighting n14 and n41.

Part (i)

What are the advantages of looking at the directed version vs undirected version of the criminal network?

In a hierarchical social network, like a criminal network, it might be challenging to understand in-degrees and out-degrees in contrast to research citations. Even while the results of that are up to interpretation, at the very least, we could observe who was opening the doors of contact. Giving directions to several unimportant players, for instance, probably doesn't mean much, although making numerous calls to other crucial players might. It would be difficult to equate the latter in terms of relevance to receiving several calls from other significant players.

Also, depending on the relationship's stage—whether it be at the start or finish of a drug deal—the buyer or seller may start the conversation, accordingly.

The left eigenvector centrality would only measure in-degrees for directed graphs, but the right eigenvector centrality would measure out-degrees. Such centrality metrics might be used to assess initiative as a forerunner to network growth or, possibly, power, together with hubs and authorities.

It might be claimed that boosting contacts causes the network's centre to move, and that doing so is more likely to occur than anticipating calls without making an effort. If so, we may use this to examine network dynamics, such as the increasing significance of Ernesto Morales or Pierre Perlini (n3) (n12), or the falling importance of Alain Levy (n83).

Part (j)

Recall the de_nition of hubs and authorities. Compute the hub and authority score of each actor, and for each phase. Using this, what relevant observations can you make on how the relationship between n1 and n3 evolves over the phases. Can you make comparisons to your results in Part (g)?

Pierre Perlini (n3) didn't have a significant impact in the early stages, which may have been because of the early investigation's limitations. On the one side, Daniel Serero (n1) coordinated with a group of investors and money carriers, and on the other, with the marijuana drug runners. Perlini occasionally made touch with them. Phase 2 saw Serero's initial contact with the cocaine importer Ernesto Morales (n12), and phase 3 saw Serero respond. When that happened (see phase 3), Perlini's position grew a little, but Serero continued to make the calls and Perlini remained mostly on the receiving end. Being the main hub, Serero, has an impact on that.

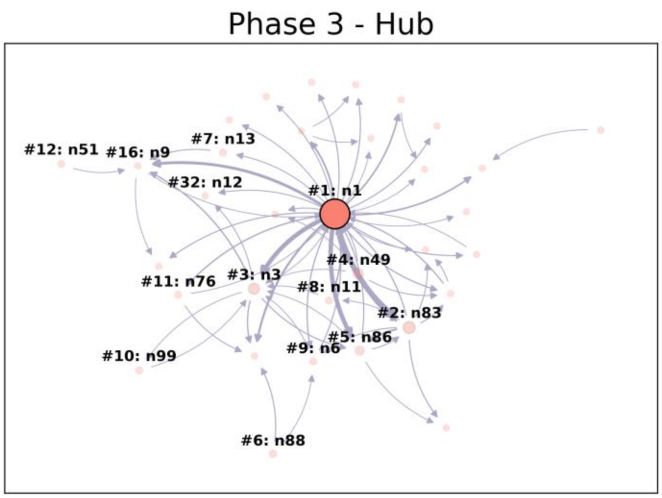
Pierre Perlini (n3) didn't have a significant impact in the early stages, which may have been because of the early investigation's limitations. On the one side, Daniel Serero (n1) coordinated with a group of investors and money carriers, and on the other, with the marijuana drug runners. Perlini occasionally made touch with them. Phase 2 saw Serero's initial

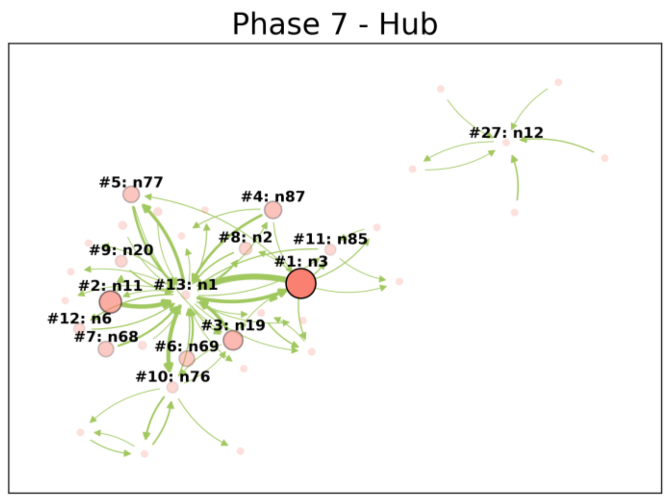
contact with the cocaine importer Ernesto Morales (n12), and phase 3 saw Serero respond. When that happened (see phase 3), Perlini's position grew a little, but Serero continued to make the calls and Perlini remained mostly on the receiving end. Being the main hub, Serero, has an impact on that.

But, it's possible that Perlini's departure had as much to do with the rankings change as his ascent to power did. Serero comfortably cashed on on his connections while he attempted to influence the drug dealers and investors towards the cocaine deal. As a result, Serero got more calls, and by phase 6, he had supplanted Perlini as the top hub. The switch in roles persisted even after phase 6's second wave of seizures; despite briefly losing touch with Morales, Perlini kept in touch with the rest of his network.

Via unidentified middlemen n12 and n14, Ernesto restored contact in phase 8, and investor Patrick Lee (n87) entered the picture. Salvatore Panetta (n82) and Steve Cunhra (n96) were signed by Serero to the roster in order to provide a straight channel into Morales' circle and bypass Perlini. By making these efforts, he regains top hub. Perlini reclaimed supreme control but remaining positioned between Serero and Morales' efforts. Perlini and Lee had extensive conversations, and Lee ultimately invited anonymous players n37 and n41 to take up Perlini's position as primary coordinator between Morales. As a result, Perlini was downgraded to a minor actor by phase 10, with investor Lee and transport Panetta continuing to hold prominent roles until the very end.

It's difficult to determine what exactly occurred to Perlini in the end, whether he voluntarily gave up his post or lost favour. Nonetheless, since I first created the narrative before responding to any of these issues, I'm sure you, dear reader, would agree that this is basically comparable to the analysis in sections (f) and (g).





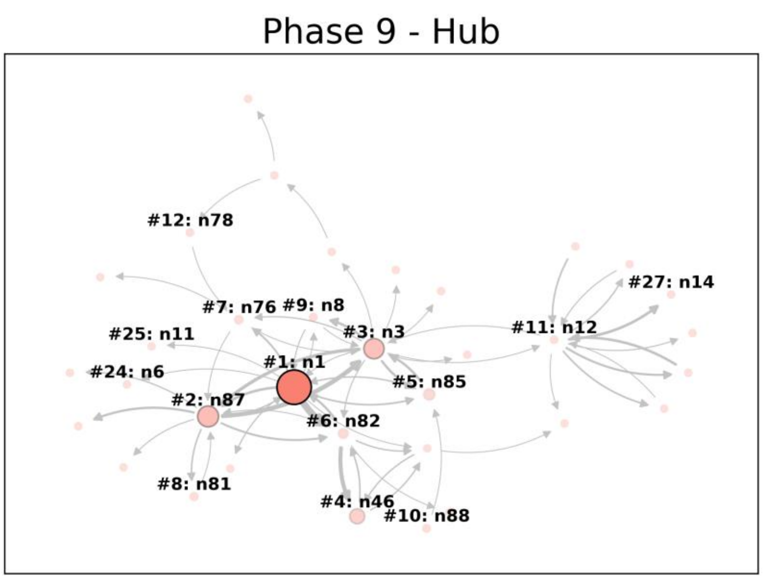


Figure 2.7: Shift of roles between Daniel Serero (n1) and Pierre Perlini (n3).

Phase	Hub: Serero	Hub: Perlini	Authority: Serero	Authority: Perlini
1	**0.7031**	0.0144	0.0118	**0.1357**
2	**0.9730**	0.0076	0.0003	**0.3367**
3	**0.7931**	0.0463	0.0032	**0.1496**
4	**0.8598**	0.0240	0.0022	**0.2755**
5	**0.9065**	0.0105	0.0006	**0.3236**
6	0.0080	**0.1953**	**0.8054**	0.0321
7	0.0068	**0.3433**	**0.7274**	0.0069
8	**0.8259**	0.0174	0.0020	**0.4672**
9	**0.5879**	0.1395	0.0162	**0.0675**
10	0.0000	0.0000	0.0000	0.0000
11	0.0000	**0.0379**	0.0000	0.0000

Figure 2.8: Hub and authority scores between Daniel Serero ($n1$) and Pierre Perlini ($n3$). **Bold** is highest.

Problem 3
Co-offending Network

Part (g)
Plot the degree distribution (or an approximation of it if needed) of G. Comment on the shape of the distribution. Could this graph have come from an Erdos-Renyi model? Why might the degree distribution have this shape?

The graph appears to have a mean degree of 2:945, a range of 2 to 3, and one significant component of 19924, which is consistent with Erdos-behavior Renyi's when np = c < log (n). A deeper look reveals that the tail distribution is too broad to be Poisson, as the Erdos-Renyi approaches for a big n, and that the other components are larger than they should be under the model. This contrast is shown in the graphs together with the log-log power law test, where a straight line indicates that this distribution may be a power law distribution but is definitely not Erdos-Renyi.

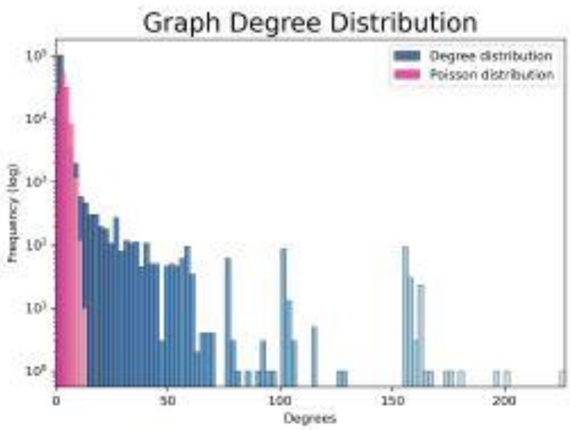


Figure 3.1: Degree distribution has wider tails than Poisson, indicating it's not Erdos-Renyi.

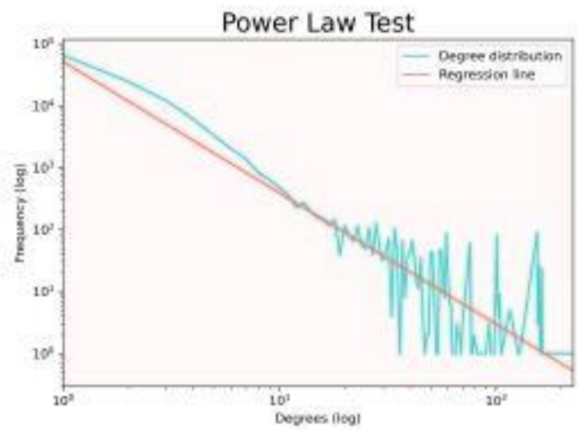
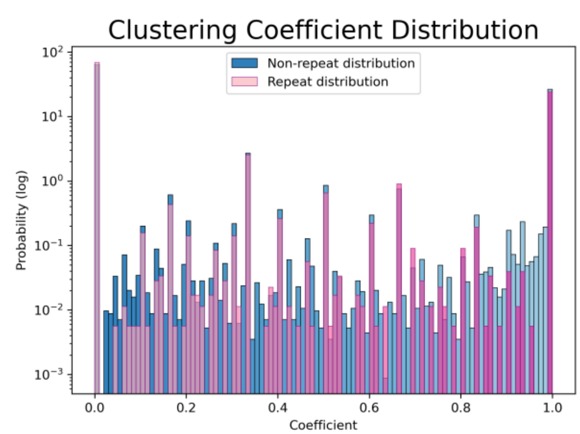


Figure 3.2: Relatively straight log-log plot indicates it might be a power law distribution.

Part (m)
Plot the distribution of clustering coefficients for each node for Gr and Gnr. What shape do the plots make? What does this tell you about the behavior of the actors?

Given that a single criminal event is represented as a fully linked induced subgraph, ends that exclusively associate with individuals from the same group would have a 1:0 clustering coefficient. Otherwise, cliques or interconnected groups of people who have all committed the same crime together would similarly have a coefficient of 1. Yet, 88% of repeat offenders belong to components with a size of 10 or less.



As isolates have been eliminated from these graphs, offenders in a line or star network with a single co-offender or many disconnected co-offenders would have a clustering coefficient of 0:0.
They make up 72% of components with a size of 2.
There may be coefficients between offenders who co-offend with numerous groups, where certain members of each group co-offend with other groups.

35–40% of the outer nodes in the top cluster (the y-axis is log-scale) are stars or line subgraphs with a 0:0 clustering coefficient. Around 25% are in clique subgraphs with a ratio of 1:1. Both non-repeat and repeat co-offenders can agree that this is true. Although the mean degree is greater for non-repeats, repeat offenders have a higher degree distribution (4 vs. 1, respectively). Degrees on average are twice as big for outliers with clustering coefficients between 0.25 and 0.75, at 8.7 degrees compared to 4.3 to 4.4 degrees. additional small.
The outcomes mostly match our predictions. Finding reliable partners should, in my opinion, be among the biggest obstacles to illicit enterprise. Self-protection vigilantes who also have a measure of loyalty and intellect would be necessary. And if one of these groups If it gathers, the FBI could have to disperse it.

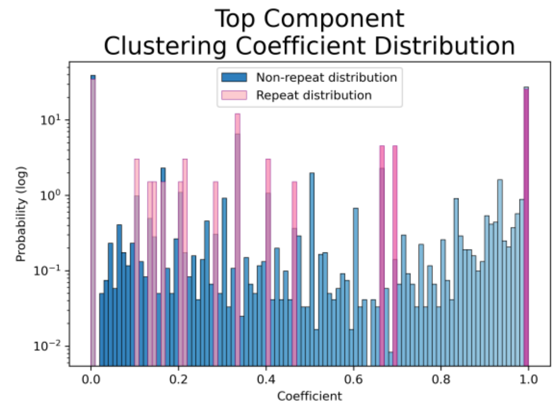


Figure 3.3: Clustering coefficients for repeat and non-repeat co-offenders.

Thus I would assume that most crimes are either performed alone or in small groups, if not both. Because of this, components often have 5 or less members.
I would anticipate that star networks' central nodes are likely gang leadership, whilst cliques are a component of operations, using the CAVIAR network as a point of comparison.

Part (n)

Pick a centrality measure (degree, eigenvector, betweenness, etc) and compute the scores for the top component of Gr and Gnr. Compare the distribution of the centrality across nodes (for example, with summary statistics and/or a histogram). Examine the number of crimes committed by the most central actor in the repeat offender graph, does this support your conclusions?

Because there are so many incredibly tiny values, the distributions are displayed as log-log graphs.
It should come as no surprise that serial offenders may collaborate with others to commit crimes, which is why they score higher in terms of centrality both generally and specifically. The eigenvector centrality (degree is indicated for interest) will be my main focus. Repeat offender 596946 has committed 36 offences and has the greatest eigenvector centrality (0:4056).
I attempted a number of justifications because this is considerably less offences by a single offender. The top 10 offenders have committed between 196-456 offences, all with typical group sizes between 1-2, and their average group size per crime is 1:917.
The offender has 13 neighbors (or degrees), 8 of them are repeat offenders, which is above average but not very high. The average crime rate among the neighbors is 27.69, the average group size is 1.67, and the average degree is 7.69. All of them are adult guys.
All of the neighbors have comparable statistics, as do their neighbors. Moreover, they are all experts in robbery and break-and-enter, with the rare incident of vehicle theft, violence, and drug charges.
Although first I wasn't sure what to anticipate, this makes perfect sense. Long penalties for heinous crimes like murder stop a wide network. Smaller networks are forced to commit crimes with longer punishments, and sex offences are typically personal. With such a vast network of cooperative criminals focusing on just one target, the perpetrator is unmistakably gang-related. It's interesting to note that the robber with the same network is also the actor with the highest degree centrality (15 degrees).

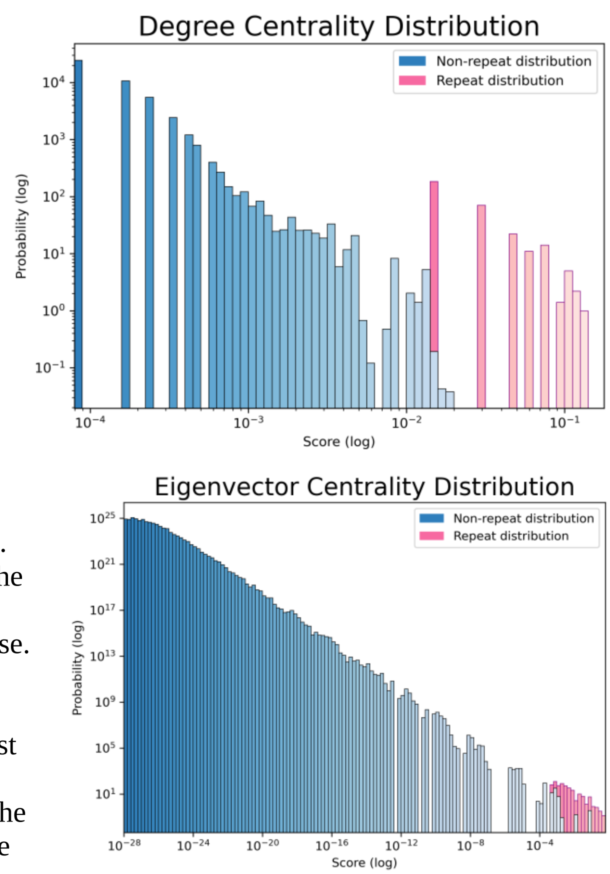


Figure 3.4: Centrality scores for repeat and non-repeat co-offenders.

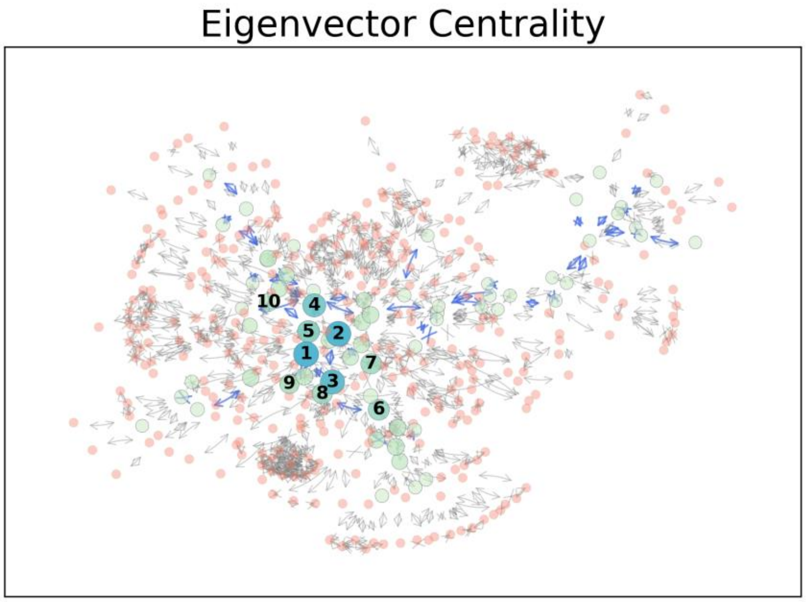


Figure 3.5: Eigenvalue centrality rank. Blue nodes are the repeat co-offender subgraph. Pink nodes are their neighbors, for context. They are mostly car and car property thieves.

Project
Impact of Crime On A Network

I.
Methodology

Different clustering coefficients[m] and centralities[n] have sociological causes, as we have already observed. Also, it's feasible that various criminal activities may be carried out in unique ways. For instance, it has been noted that robbers have a high centrality, which is likely related to the fact that misdemeanors generally carry lesser sentences on their own, that they are harder to prosecute, or that there is a gang aspect that leads many of these criminals to be connected. Personal crimes, such as sexual offences, on the other hand, are typically committed in smaller groups.
I'll be investigating further whether the nature of the crime might affect the co-occurrence network and whether some crimes are more likely to involve local structures like cliques or star graphs. I'll look at a few instances of the tales that local institutions have related to various types of crime.
The only information that is provided must be relative to a node in order for local structures to be predicted. In all honesty, we were unable to collect population statistics for the complete data set. As a result, we ought to employ summary statistics for the node alone, an average of its immediate neighbors, and maybe an average of the connect component it is a part of. Hub and authority ratings are rather unimportant because co-sense is undirected, thus we will focus on clustering coefficient, degree, betweenness, and eigenvector centrality instead. Since they are too prevalent and there would be insufficient structural data to make an observation, small components n 3 are first eliminated. We may as well create MLE predictions of the most typical crime category for nodes in such compositions.
The first three digits of the crime type number, such as df['CrimeType']% 100, are used to group crimes. These offences are divided into similar categories, such as (11) representing various murder kinds, (12) murder conspiracy, (13) sexual offences, (14) assault, (21) theft, (31) prostitution, (41) drug possession, (42) drug tracking, (212) robbery, (213–214) car theft, and so on. We may visually inspect the various components for each crime category to get a sense of whether any local structures exist. The graphs will highlight a specific crime and color-code it on a blue-green scale according to centrality rating, with a pink layer of the nodes that are closest to it for context. The relevant crime's cliques and star graphs will also be coloured. Every node will be identified by the perpetrator's most frequent offence, which is frequently distinct from the sort of crime being investigated. Graph will center on largest betweenness centrality. For bigger components n 20, eigenvector centrality-based graphs will also be produced. The clustering coefficient ought should be roughly visible. Finally, we will look at summary statistics for holistic perspective. They will be grouped by crime type for each of the statistics described above.

II.

Results

There are several instances of local graph topologies that shed light on the current state of nature. Larger components of some crimes have high star graph occurrences, such as breaking transportation rules like the Taxi Transport Act (73007) or the Lotteries or Races Act (73003) or exceeding truck weight restrictions (73005). They are grouped together under the (730) category. Drivers may frequently travel alone or in small groups, and the large distance assures disconnected co-offenders.

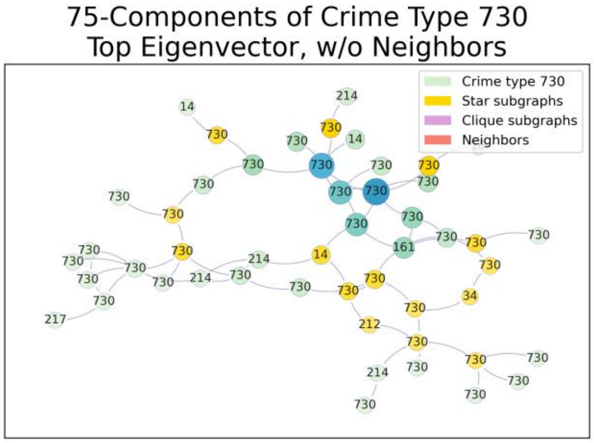


Figure 4.1: Long distance relationships.

The star graphs of prostitutes (31) who are most likely independent and not affiliated with a syndicate typically have tiny component sizes. We might be able to make assumptions about groups even within a particular crime category using these systems.

For instance, prostitutes who are likely connected to criminal organizations sometimes create sizable cliques that are held together by a small number of prominent figures. A huge clique would also include heroin addicts (41), who don't appear to have any trouble ding friends to use with.

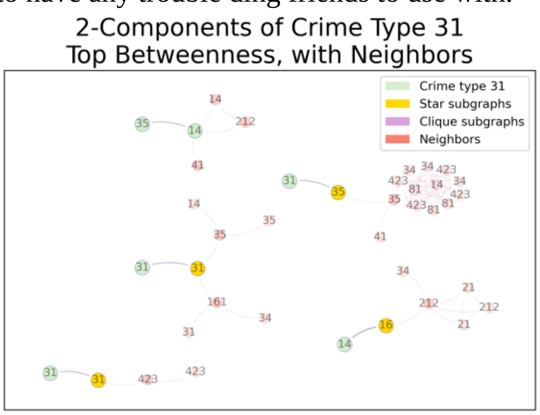


Figure 4.2: Traveling with friends?

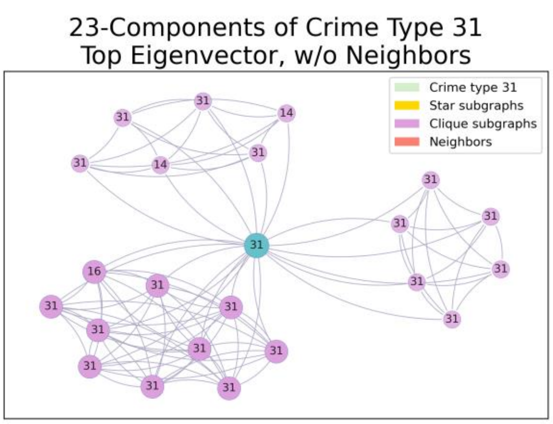


Figure 4.3: Could it be a pimp?

Another example of a large clique would be heroin addicts (41), whom seem to have no trouble ending buddies to use with.

Local structures, however, are more complex than merely whether they resemble star or clique network patterns.

Many networks are close but not precisely the same, so what does it possibly mean that just some people may not know each other? Instead of rigorously grouping coefficients of 0 or 1, I believe we should retain an open mind and search for broader patterns. Large, less related sub-graphs of the same crime might exist. For example, we have this network of robbery (212), and weapons offenses and smuggling (33),respectively.

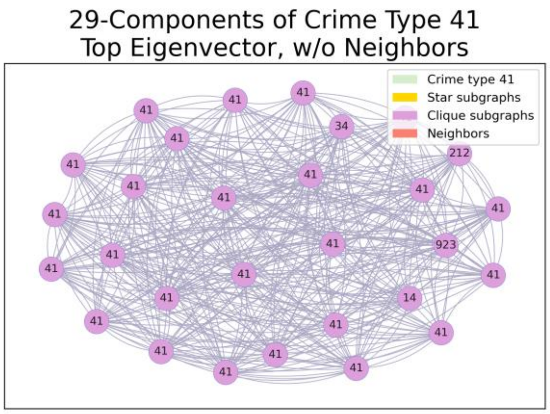


Figure 4.4: I'm sure it was a party.

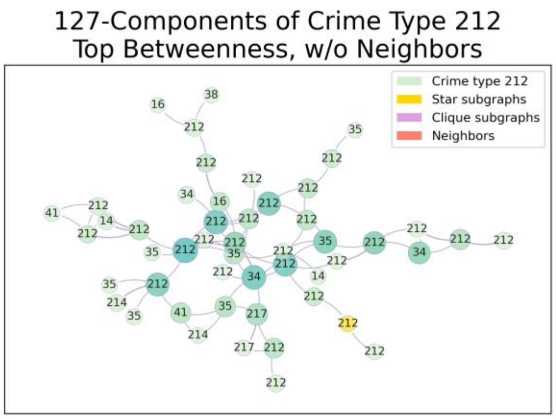


Figure 4.5: It looks street-gang related.

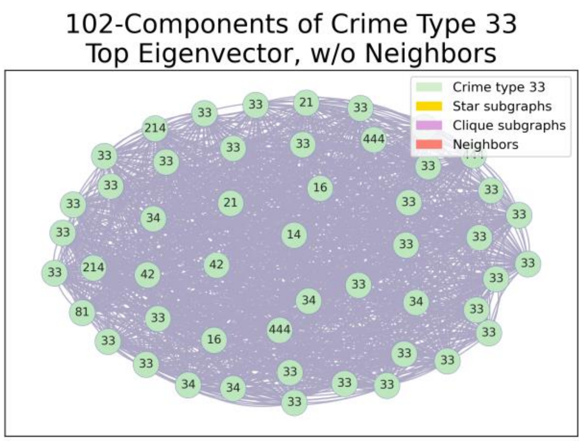


Figure 4.6: More sophisticated gang related.

When we consider their surrounding areas and discover potential connections between different crime categories, it starts to become intriguing. The graphs show that some crimes are frequently performed in combination with other crimes. Two categories of elucidations may be possible here:

1. Organizational patterns that reveal how certain crimes interact with one another.
2. Elements that highlight specific central actors.

For the first, here is a central group of prostitutes (34), with important members related to other prostitution gangs, who also trade in counterfeit currency and legal avoidance. We may assume that the main actors have some connection to management.

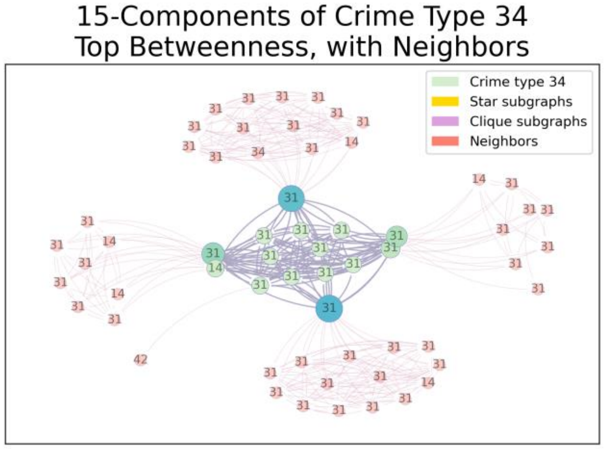


Figure 4.7: Intersection of multiple prostitution groups.

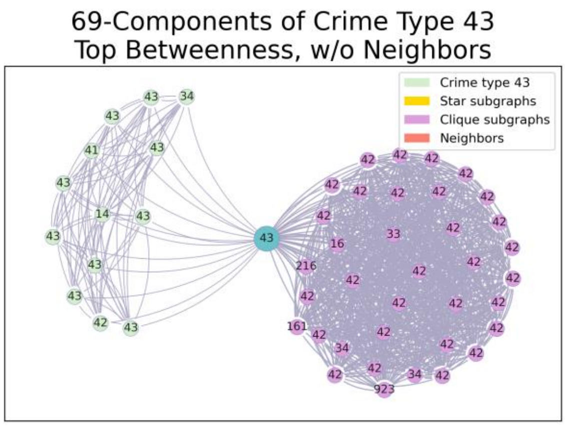


Figure 4.8: The two cultures?

In other developments, the structural connection between the network's trackers (42) and drug importers (43) could be readily seen. The next example displays possibly drug-transporting stolen autos (213) (42). The whole network depicted deals with vehicle theft, but the high eigen-vector centrality subgraph that is emphasized deals with tracking and possession.

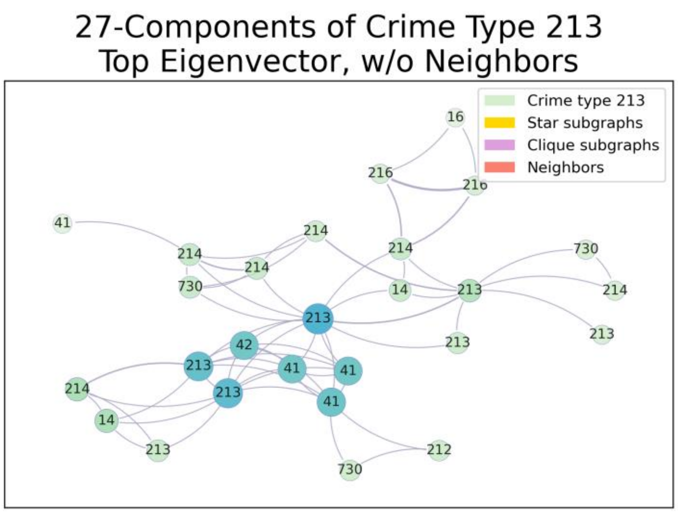


Figure 4.9: The two cultures?

The traffickers and prostitutes establish a clique that is connected by a single central person.

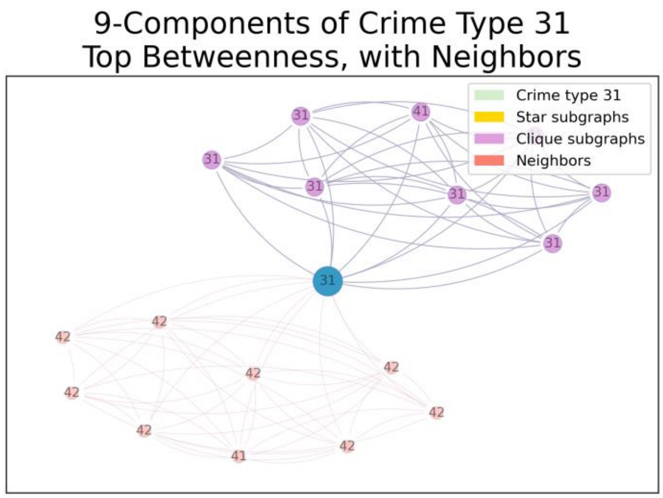


Figure 4.10: Sex, drugs and rock n' roll.

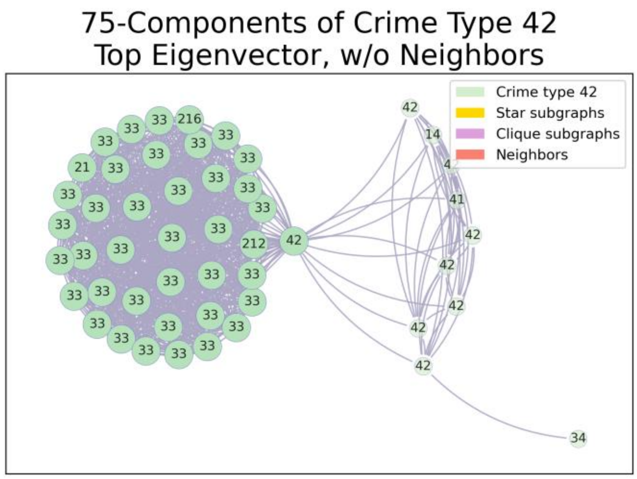


Figure 4.11: Drug kingpin.

Let's look at a few instances of potential criminal business leadership to complete the graphs. The first one (above) features a network for tracking drugs that is supported by an army of armed (33) criminals.

The next individual is the sole link between two intricately linked groups of vehicle thieves (213).

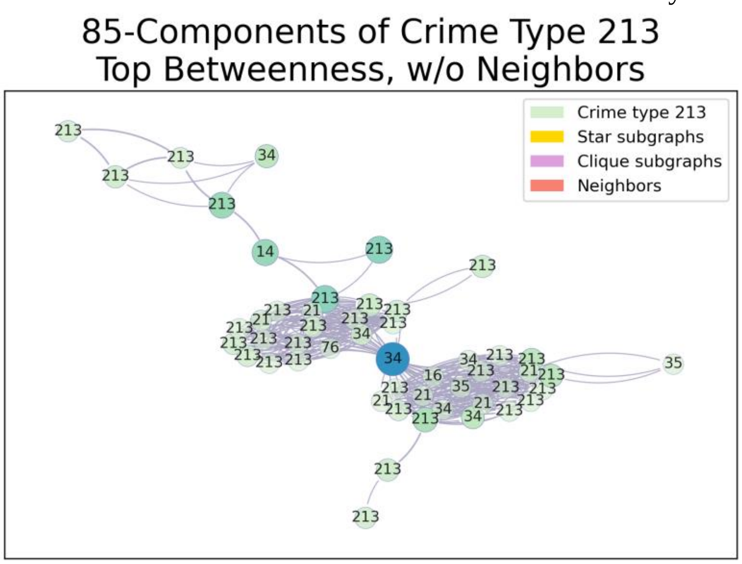


Figure 4.12: He's scoping out your BMW.

Finally, here is a cannabis grower (444) and trafficker (42) that mediates other traffickers, armed criminals (33), people who commit violent assault (14), and members of criminal organizations (399).

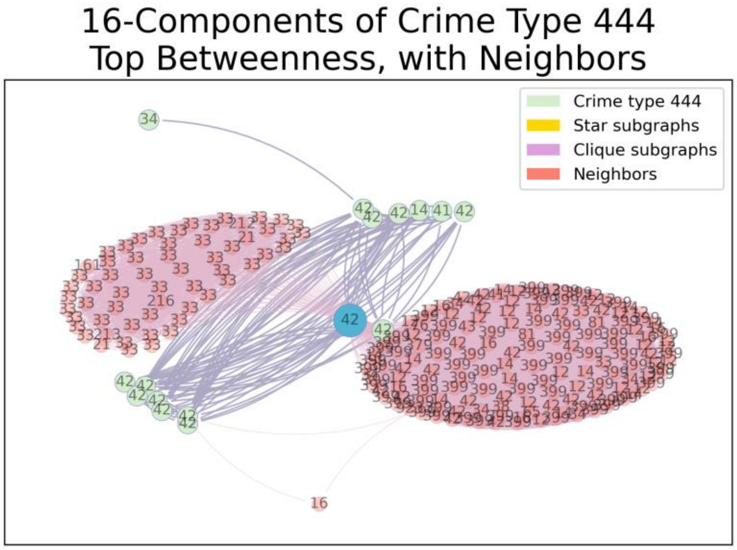


Figure 4.13: He owns all the hydroponic drug houses.

Here are the summary status of crime

	Node				Neighbor				Component			
Crime Type	Clst	Degr	Eigv	Btwn	Clst	Degr	Eigv	Btwn	Clst	Degr	Eigv	Btwn
11	1e-03	7e-05	2e-18	3e-05	1e-03	9e-05	9e-17	6e-05	1e-03	9e-05	2e-04	7e-05
12	2e-03	6e-05	4e-12	5e-05	1e-03	1e-04	3e-11	2e-04	1e-03	1e-04	3e-04	9e-05
13	1e-03	5e-05	1e-11	1e-04	9e-04	9e-05	9e-09	3e-04	1e-03	9e-05	2e-04	7e-05
14	1e-03	6e-05	1e-05	8e-05	1e-03	1e-04	3e-05	2e-04	1e-03	9e-05	2e-04	7e-05
15	1e-03	5e-05	2e-17	9e-06	2e-03	8e-05	2e-17	2e-04	2e-03	1e-04	4e-04	1e-04
16	1e-03	1e-04	2e-06	7e-05	1e-03	1e-04	5e-05	2e-04	1e-03	9e-05	3e-04	8e-05
21	1e-03	1e-04	2e-07	8e-05	1e-03	1e-04	7e-06	2e-04	1e-03	1e-04	3e-04	8e-05
31	2e-03	1e-04	5e-14	4e-05	2e-03	1e-04	1e-13	1e-04	2e-03	1e-04	2e-04	6e-05
32	2e-03	2e-04	6e-17	2e-05	2e-03	2e-04	1e-14	4e-05	2e-03	2e-04	1e-04	3e-05
33	3e-03	1e-04	2e-03	7e-05	3e-03	2e-04	2e-03	2e-04	2e-03	1e-04	3e-04	8e-05
34	1e-03	1e-04	3e-07	8e-05	1e-03	1e-04	4e-05	2e-04	1e-03	1e-04	3e-04	8e-05
35	1e-03	6e-05	2e-06	8e-05	1e-03	1e-04	5e-05	2e-04	1e-03	1e-04	3e-04	8e-05
37	2e-03	5e-05	7e-06	8e-05	2e-03	1e-04	7e-05	2e-04	2e-03	1e-04	3e-04	9e-05
38	3e-03	9e-05	6e-05	6e-05	3e-03	1e-04	9e-05	1e-04	2e-03	1e-04	3e-04	8e-05
41	1e-03	1e-04	3e-07	8e-05	1e-03	1e-04	4e-05	2e-04	1e-03	9e-05	2e-04	7e-05
42	2e-03	2e-04	2e-03	6e-05	2e-03	3e-04	2e-03	2e-04	2e-03	1e-04	3e-04	8e-05
43	3e-03	2e-04	1e-06	2e-05	3e-03	2e-04	2e-06	6e-05	2e-03	1e-04	2e-04	7e-05
44	1e-03	6e-05	2e-18	2e-05	1e-03	9e-05	6e-16	1e-04	1e-03	8e-05	2e-04	5e-05
45	1e-03	4e-05	5e-20	5e-05	8e-04	9e-05	3e-18	1e-04	1e-03	1e-04	3e-04	9e-05
46	2e-03	2e-04	8e-20	5e-05	1e-03	2e-04	2e-18	1e-04	2e-03	1e-04	4e-04	1e-04
49	2e-03	4e-05	1e-22	1e-04	2e-03	6e-05	1e-20	2e-04	3e-03	1e-04	3e-04	8e-05
51	1e-03	9e-05	2e-11	9e-05	1e-03	1e-04	9e-09	3e-04	1e-03	1e-04	2e-04	7e-05
52	2e-03	1e-04	1e-13	9e-05	2e-03	2e-04	2e-12	2e-04	2e-03	1e-04	3e-04	8e-05
53	7e-04	5e-05	3e-16	2e-04	5e-04	8e-05	1e-14	6e-04	1e-03	8e-05	2e-04	6e-05
54	3e-03	7e-05	8e-22	1e-05	3e-03	8e-05	2e-20	9e-06	3e-03	8e-05	1e-04	3e-05
61	3e-03	5e-05	3e-10	1e-05	3e-03	1e-04	4e-09	2e-05	1e-03	9e-05	3e-04	8e-05
62	3e-03	8e-05	2e-27	0e+00	3e-03	8e-05	2e-27	7e-10	2e-03	7e-05	2e-27	5e-10
64	1e-03	8e-05	6e-12	1e-04	1e-03	1e-04	2e-10	3e-04	1e-03	1e-04	3e-04	9e-05
65	3e-03	5e-05	1e-12	3e-05	2e-03	9e-05	6e-11	5e-05	2e-03	8e-05	2e-04	6e-05
76	1e-03	5e-05	7e-07	5e-05	1e-03	1e-04	2e-04	1e-04	1e-03	9e-05	2e-04	7e-05
81	1e-03	8e-05	1e-05	5e-05	1e-03	1e-04	1e-05	1e-04	1e-03	1e-04	2e-04	6e-05
91	4e-04	1e-04	5e-23	5e-04	4e-04	1e-04	2e-21	7e-04	1e-03	1e-04	4e-04	1e-04
92	1e-03	5e-05	7e-14	4e-05	9e-04	9e-05	1e-11	1e-04	1e-03	9e-05	2e-04	7e-05
93	1e-03	6e-05	4e-07	7e-05	1e-03	1e-04	7e-05	2e-04	1e-03	1e-04	2e-04	7e-05
94	4e-04	9e-05	8e-19	1e-04	5e-04	1e-04	1e-16	3e-04	1e-03	1e-04	4e-04	1e-04
145	2e-03	5e-05	3e-11	2e-05	1e-03	9e-05	4e-10	1e-04	1e-03	9e-05	3e-04	8e-05
146	1e-03	6e-05	1e-11	8e-05	1e-03	1e-04	2e-10	3e-04	1e-03	1e-04	3e-04	8e-05
151	2e-03	8e-05	9e-11	1e-04	2e-03	1e-04	4e-08	3e-04	2e-03	1e-04	3e-04	8e-05
161	1e-03	6e-05	2e-09	1e-04	1e-03	1e-04	3e-09	3e-04	1e-03	1e-04	3e-04	8e-05
162	3e-03	1e-04	3e-04	5e-05	3e-03	2e-04	3e-04	3e-04	2e-03	1e-04	6e-04	9e-05
167	1e-03	8e-05	4e-11	2e-04	1e-03	1e-04	2e-10	2e-04	1e-03	1e-04	2e-04	6e-05
211	1e-03	7e-05	7e-09	1e-04	1e-03	1e-04	8e-09	2e-04	1e-03	9e-05	3e-04	8e-05
212	1e-03	6e-05	9e-08	8e-05	1e-03	1e-04	1e-05	2e-04	1e-03	1e-04	2e-04	7e-05

Figure 4.14: Mean clustering coefficient and degree, eigenvector and betweenness centralities per node, neighbors and components. Some of the outliers are highlighted.

Crime Type	Node				Neighbor				Component			
	Clst	Degr	Eigv	Btwn	Clst	Degr	Eigv	Btwn	Clst	Degr	Eigv	Btwn
213	1e-03	7e-05	2e-07	9e-05	1e-03	1e-04	2e-05	2e-04	1e-03	1e-04	3e-04	8e-05
214	1e-03	6e-05	3e-07	8e-05	1e-03	9e-05	2e-05	2e-04	1e-03	9e-05	2e-04	7e-05
216	1e-03	1e-04	1e-05	9e-05	1e-03	1e-04	3e-05	2e-04	1e-03	1e-04	3e-04	8e-05
217	1e-03	7e-05	1e-06	6e-05	1e-03	1e-04	3e-05	2e-04	1e-03	1e-04	2e-04	6e-05
323	3e-02	1e-04	8e-03	7e-06	3e-02	2e-04	9e-03	4e-05	3e-02	2e-04	8e-03	3e-05
345	2e-03	3e-04	2e-15	6e-05	2e-03	3e-04	1e-13	2e-05	2e-03	3e-04	8e-05	2e-05
371	2e-03	1e-04	3e-11	8e-05	2e-03	2e-04	6e-11	2e-04	2e-03	1e-04	2e-04	7e-05
373	2e-03	7e-05	2e-11	7e-05	1e-03	1e-04	1e-09	2e-04	2e-03	9e-05	2e-04	8e-05
379	2e-03	8e-05	5e-14	3e-05	1e-03	1e-04	4e-12	8e-05	1e-03	1e-04	3e-04	8e-05
381	2e-03	5e-05	3e-09	3e-05	2e-03	9e-05	3e-09	2e-04	2e-03	9e-05	2e-04	7e-05
384	7e-03	5e-04	2e-06	5e-05	7e-03	7e-04	2e-06	8e-05	2e-03	2e-04	4e-04	1e-04
399	2e-03	4e-04	2e-04	1e-04	2e-03	4e-04	2e-04	3e-04	2e-03	1e-04	5e-04	1e-04
413	2e-03	8e-05	6e-10	8e-05	2e-03	1e-04	8e-09	3e-04	1e-03	1e-04	3e-04	1e-04
421	3e-03	5e-05	8e-13	3e-05	3e-03	9e-05	2e-11	1e-04	2e-03	9e-05	2e-04	8e-05
422	3e-03	2e-04	4e-03	5e-05	3e-03	3e-04	5e-03	2e-04	2e-03	1e-04	3e-04	8e-05
423	2e-03	2e-04	6e-06	8e-05	2e-03	2e-04	4e-05	2e-04	2e-03	1e-04	3e-04	9e-05
424	2e-03	2e-04	5e-06	5e-05	2e-03	2e-04	7e-05	1e-04	2e-03	1e-04	2e-04	7e-05
425	2e-03	5e-05	2e-04	3e-05	3e-03	9e-05	4e-04	7e-05	3e-03	9e-05	6e-04	7e-05
426	2e-03	2e-03	9e-05	1e-04	2e-03	2e-03	3e-04	2e-04	2e-03	1e-04	6e-04	1e-04
433	4e-04	4e-05	5e-22	1e-05	3e-04	1e-04	3e-20	2e-04	1e-03	1e-04	5e-04	2e-04
444	2e-03	1e-04	3e-03	3e-05	2e-03	2e-04	3e-03	8e-05	2e-03	9e-05	2e-04	5e-05
521	2e-03	5e-05	1e-12	2e-05	2e-03	1e-04	3e-10	2e-04	2e-03	1e-04	2e-04	7e-05
690	1e-03	6e-05	2e-15	2e-05	1e-03	1e-04	6e-14	4e-05	2e-03	1e-04	3e-04	8e-05
710	8e-04	2e-04	2e-14	2e-05	8e-04	3e-04	6e-13	6e-05	1e-03	2e-04	9e-05	3e-05
730	7e-04	6e-05	3e-13	2e-04	7e-04	1e-04	9e-12	4e-04	1e-03	1e-04	3e-04	1e-04
750	1e-03	5e-05	3e-10	4e-05	1e-03	1e-04	6e-08	1e-04	1e-03	9e-05	2e-04	7e-05
911	1e-03	5e-05	7e-21	5e-05	1e-03	9e-05	1e-18	6e-05	1e-03	9e-05	3e-04	8e-05
912	8e-04	9e-05	5e-12	1e-04	9e-04	1e-04	2e-11	2e-04	1e-03	9e-05	3e-04	8e-05
913	1e-03	8e-05	2e-07	9e-05	1e-03	1e-04	6e-05	2e-04	1e-03	1e-04	3e-04	9e-05
923	1e-03	6e-05	3e-07	7e-05	1e-03	9e-05	5e-05	2e-04	1e-03	9e-05	2e-04	6e-05
924	1e-03	6e-05	2e-12	1e-04	1e-03	1e-04	3e-12	2e-04	1e-03	9e-05	2e-04	7e-05
931	9e-04	4e-05	1e-11	3e-05	9e-04	8e-05	8e-10	1e-04	1e-03	8e-05	2e-04	6e-05

Figure 4.15: Results (continued).

III.
Discussion

Based on network structure, results were discovered that only flimsily suggest that crimes are predictable.
Even though there are many isolated instances that point to patterns, it was challenging to detect them from the summary statistics. There hasn't been a definite trend, despite the fact that clique and star-induced subgraphs for various crime categories could be observed across the network. As many criminals commit many offences, various paragraphs overlap. It is probably oversimplified and neglects many of these dynamics to capture only the node, neighbor, and component coefficients and centralities. In a similar vein, it is more comparable to degree centrality than an Eigen-vector. One thing I saw is that I was only able to imagine how valuable the graph may be after thoroughly investigating it. For instance, the police may utilize these networks to examine the methods by which crimes are committed, discover the relationships between offenders, and even foretell the locations and types of upcoming crimes that may occur by evaluating how similar a person or community is to these practices. In hindsight, the findings aren't really shocking; I could foresee the connections between these crimes. But, they are instructive nonetheless, perhaps because they confirm my suspicions. The following stages would be to break down these statistics by node and use the crime categories as labels to train a logistic regression or neural network to see if we could more accurately predict some of the crime kinds. Therefore, using this model as a jumping off point for an inquiry, one may determine whether a criminal on the loose would have additional potential links.

References

[1] Martin Stone. "Canada arrests 30 for drug trafficking," United Press International, April 18, 1996. [Online]. Available: https://www.upi.com/Archives/1996/04/18/Canada-arrests-30-for-drug-trafficking/4693829800000/. [Accessed April 10, 2021].
[2] "Daniel 'The Arab' Serero," Oocities, October, 2009. [Online]. Available: https://www.oocities.org/wiseguywally/DanielSerero.html. [Accessed April 11, 2021].24