Written Report – 6.419x Module 3

Name: (Naichun Chen)

§ Problem 1: Suggesting Similar Papers

Part (c):(2 points) Include your answer to this question in your written report. (100 word limit.)

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

As above, for a brief introduction to the big-O notation, refer to the optional problem 1.7 in Module 1.

Solution: Both of the two methods are O(n^3).

Compared to the friend's algorithm, more related libraries are developed with matrix operations, the simplicity and optimization in computing, so we could solve the problem in an efficient way in matrix operations.

2. Part (d):(3 points)Include your answer to this question in your written report. (200 word limit.)

Bibliographic coupling and cocitation 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?

Solution:

Bibliographic coupling is two papers that cite the same set of other papers, in network analysis, the edge weights represent the number of shared citations between the two papers. The similarity can be measured by these shared citations in the papers, which is highly related to the availability of references in the time when the papers were published. Even if the arriving paper is published recently, the Bibliographic coupling is not updated within time. So it's more static and stable than Cociation. It's a good indicator for the start point and initializer in the research field.

Co-citation is two papers that are both cited by the same paper, in network analysis, weights represent the number of times the papers are cited. The similarity can be measured by the frequency of two papers being cited. When a new paper is published, the trend is also changed. So, it's a dynamic measure and a good indicator of trend and influence of the research field.

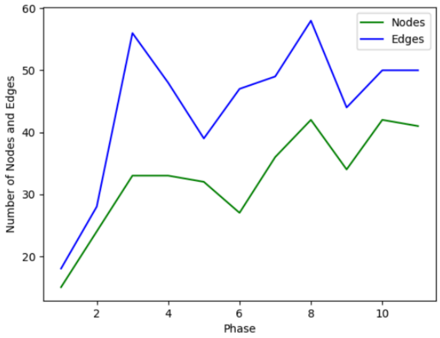
Hence, Bibliographic coupling is a good indicator to find the similarities of the foundation between two papers at a certain period of time. On the other hand, co-citation analysis focus more on dynamically impacts of the research field, so it's a good indicator for the similarities of trend and insights evolving with time between two papers.

§ **Problem 2: Investigating a time-varying criminal network**

Part (c):(2 points)Include your answer to this question in your written report. (100 words, 200 word limit.)

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?

Solution:



The dramatic increase of the nodes in the beginning phases indicate the expansion of the criminal network. The criminal organization hired more members (nodes) and re-organized the network(edges) to adjust the investigation of the police office. The evolving of the networks reflects the adaptation of criminal networks to against police force.

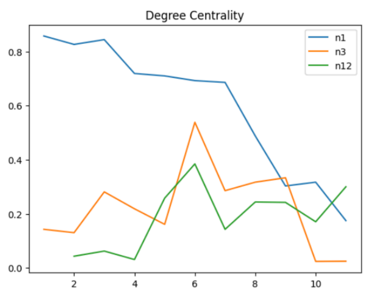
In the later process, there is more stability and no substantial increase, due to the reinforcement of security and the movement of the operation, such as after the first seizure, traffickers reoriented to import cocaine to other places.

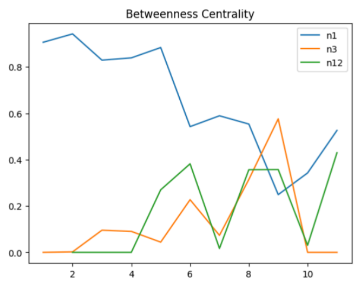
In Part(b) Question5, we are exploring the temporal consistency of a player's centrality, dramatically expanding if the initial phases might affect the measurement of the centrality. But the networks become more and more stable in the end, so it is still effective in the overall process.

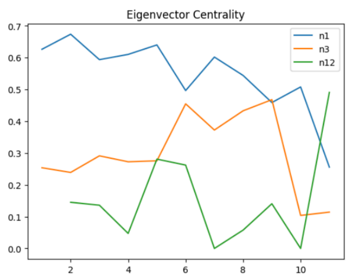
Part (d):(5 points)Include your answer to this question in your written report. (300 words, 400 word limit.)

In the context of criminal networks, what would each of these metrics (including degree, betweenness, and eigenvector centrality) 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.

Solution:







Degree Centrality reflects the influence by connecting one node to the others. If one node has higher value than the other, it implies the node connects to more nodes, but it ignores the whole structure.

Betweenness Centrality reflects the control of the information. It quantitatively identifies the position and leadership. If one crucial node has been deleted , then the operation will be interrupted.

Eigenvector Centrality reflects the centerness of the network. The higher the eigenvector centrality, the higher the power. The most important one highly connects to the rest of influential nodes.

Degree centrality and eigenvector centrality highlight direct and overall influence, these metrics imply the importance of an actor's role in the traffic evolving all the time. BUT **betweenness centrality** will be most relevant to identify who is running the illegal activities of the group, once the crucial node is removed, the whole network will malfunction.

Part (e):(3 points)Include your answer to this question in your written report. (100 words, 200 word limit)

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.

Hint: Note that the definition of a player's "importance" (i.e. how central they are) can vary based on the question you are trying to answer. Begin by defining what makes a player important to the group (in your opinion) ; use your answers from Part (d) to identify which metric(s) are relevant based on your definition and then, use your quantitative analysis to identify the central and peripheral traffickers. You may also perform a different quantitative analysis, if your definition of importance requires it.

Solution:

```
Highest Mean Centrality :
 Degree -
   index      Mean   Std. dev  Variance  Activity
0     n1  0.601485  0.240848  0.058008      11.0
4     n3  0.223505  0.150110  0.022533      11.0
25   n12  0.170893  0.125645  0.015787      10.0
7    n85  0.118010  0.049330  0.002433      11.0
22   n76  0.112235  0.073632  0.005422      10.0
3    n83  0.095836  0.076965  0.005924      11.0
Betweenness -
   index      Mean   Std. dev  Variance  Activity
0     n1  0.655051  0.238535  0.056899      11.0
25   n12  0.167562  0.187327  0.035092      10.0
4     n3  0.129403  0.179745  0.032308      11.0
22   n76  0.083791  0.099159  0.009833      10.0
54   n87  0.061327  0.092958  0.008641       6.0
84   n41  0.050369  0.167055  0.027907       3.0
Eigenvector -
   index      Mean   Std. dev  Variance  Activity
0     n1  0.546391  0.117044  0.013699      11.0
4     n3  0.298095  0.124629  0.015533      11.0
7    n85  0.190612  0.065860  0.004338      11.0
22   n76  0.165877  0.080282  0.006445      10.0
3    n83  0.153522  0.083390  0.006954      11.0
14    n8  0.152394  0.067531  0.004560      10.0
```

Degree Centrality : n1 has the highest value, which represents the highest direct connection.

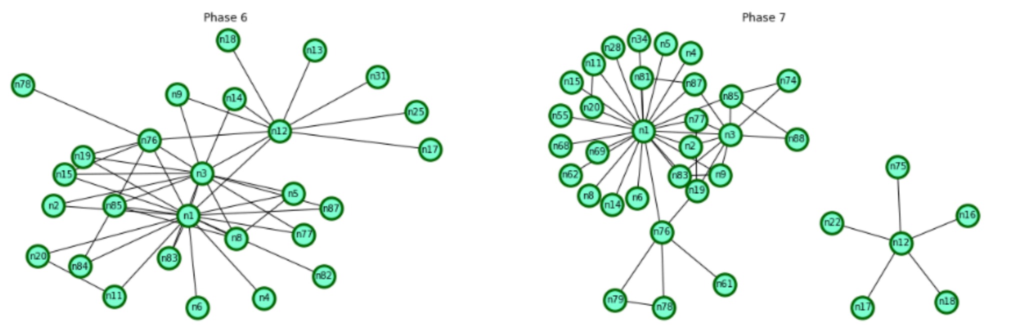
Betweenness Centrality: n1, n3, which show the highest value in three of the measurements, which represent the controlness of the criminal information network.

Eigenvector Centrality: n1 also represents the highest value which reflects the connection with other central people.
Daniel Serero (n1) : Mastermind of the network, played an important role in the beginning of the phases, but Pierre Perlini (n3) : Principal lieutenant of Serero, became more crucial in the end.
In general, n1, and n3 are most relevant to running the illegal activities of the group.
highlights **n1** is the most central and crucial traffickers with the broadest connection and controlness.

Part (f) Question 2:(3 points)Include your answer to this question in your written report. (200 words, 300 word limit.)

The change in the network from Phase X to X+1 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.

Solution:



The major change of network happened in **phase 6 and phase 7**, indicating that major events that took place during this period, matches the timeline of traffickers reoriented to cocaine import from Colombia, transiting through the United States. The whole structure has been reorganized from **centralization to distribution** after interruption and investigation from the police.

Phase 6: n1 and **n3** have the authority as the mastermind of a drug network in downtown Montréal, attempting to import marijuana to Canada from Morocco.

Phase 7: The transition from Marijuana to Cocaine Trafficking. **n12(Ernesto Morales)** as a new principal organizer involved in the cocaine import, intermediary between the Colombians and the Serero organization.

Part (g):(4 points)Include your answer to this question in your written report. (200 words, 300 word limit.)

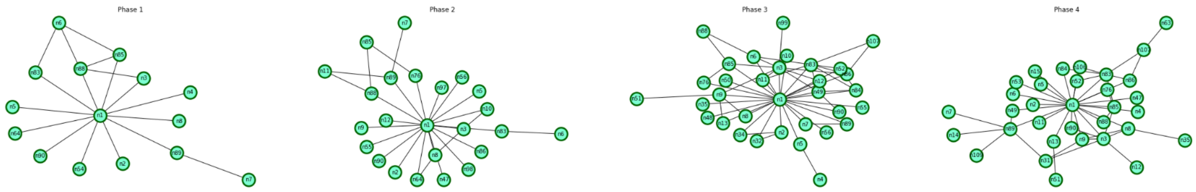
While centrality helps explain the evolution of every player's role individually, we need to explore the global trends and incidents in the story in order to understand the behavior of the criminal enterprise.

Describe the coarse pattern(s) you ***observe as the network evolves through the phases****. Does the network evolution reflect the background story?*

Hint: Look at the set of actors involved at each phase, and describe how the composition of the graph is changing. Investigate when important actors seem to change roles by their movement within the hierarchy. Correlate your observations with the information that the police provided in the setup to this homework problem.

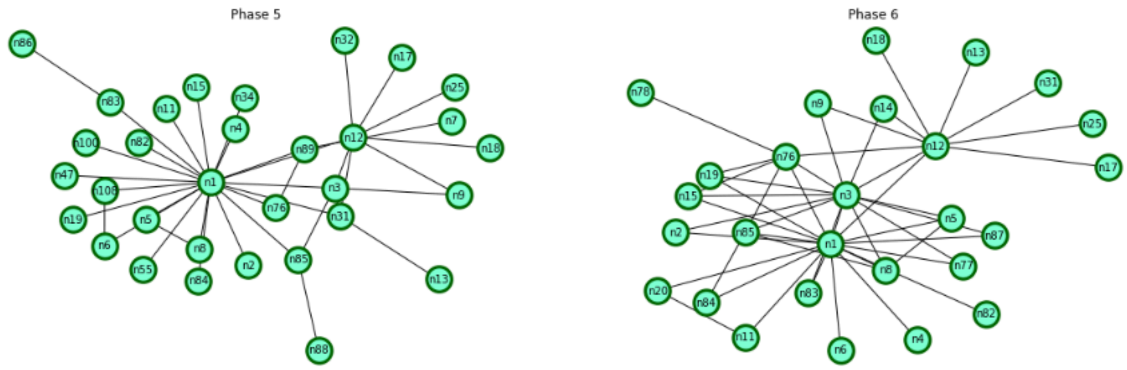
Yes, the network evolves through the phase that reflects the background story.

Phase 1-4:



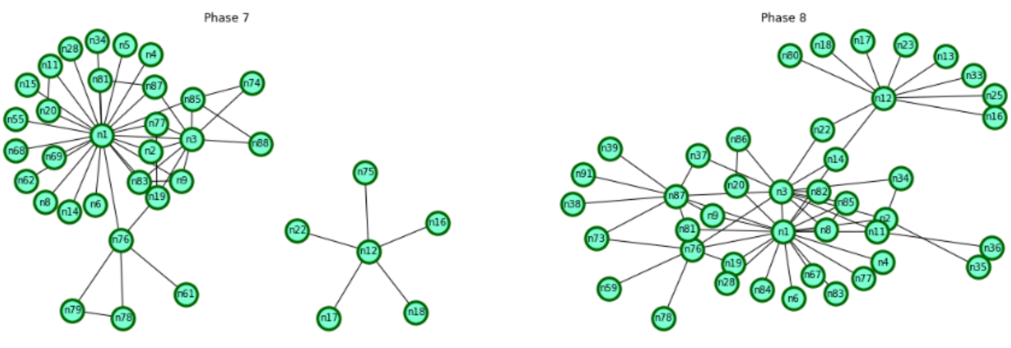
In this period, the expansion and the centralization is obvious. As a mastermind of the network, n1 shows a strong ability in coordination and recruiting. More and more members joined the network.

Phase 5-6:

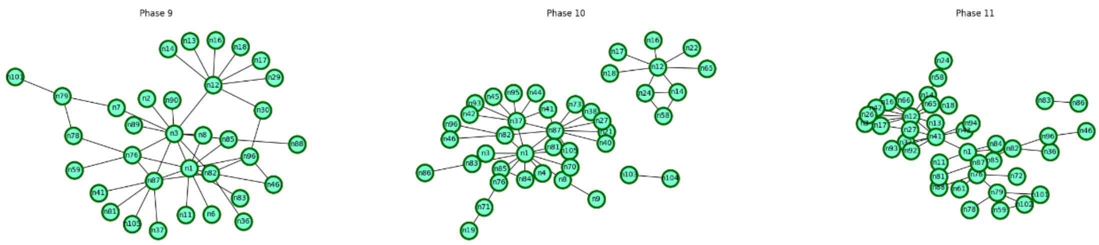


After the first seizure, happening in phase 4, n12 started to play a connecting role as principal organizer of the cocaine import, intermediary between the Colombians and the Serero organization, in order to adaptation to police interventions

Phase 7-8:



In this period, the network shows the shift of the power. n12 as a principal organizer of the cocaine import becomes more prominent in the whole structure. Decentralization to support the new trafficking strategic shifts.



Phase 9-11:

Hierarchical structure has been replaced in this period, which shows adaptive responsiveness. The whole network becomes more resilient and decentralized, in order to maintain the operation without arresting.

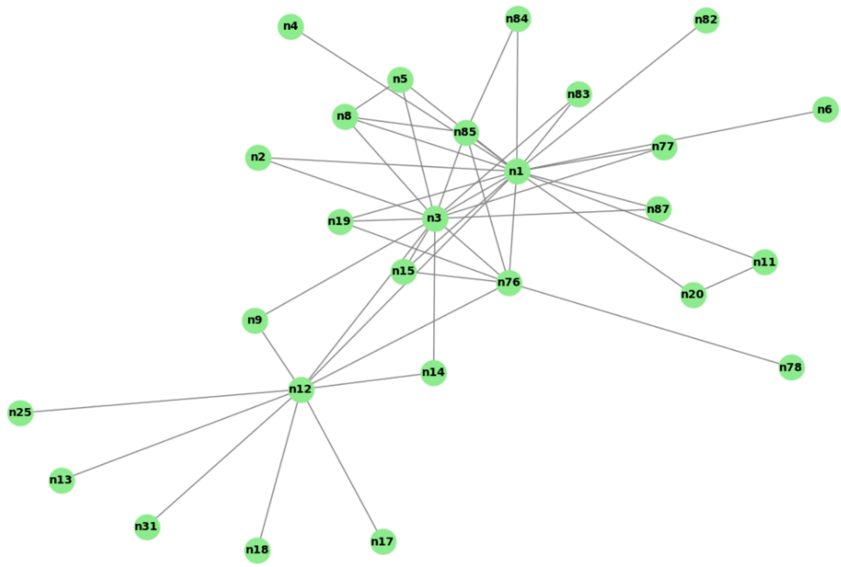
Part (h):(2 points)Include your answer to this question in your written report. (50 words, 100 word limit.)

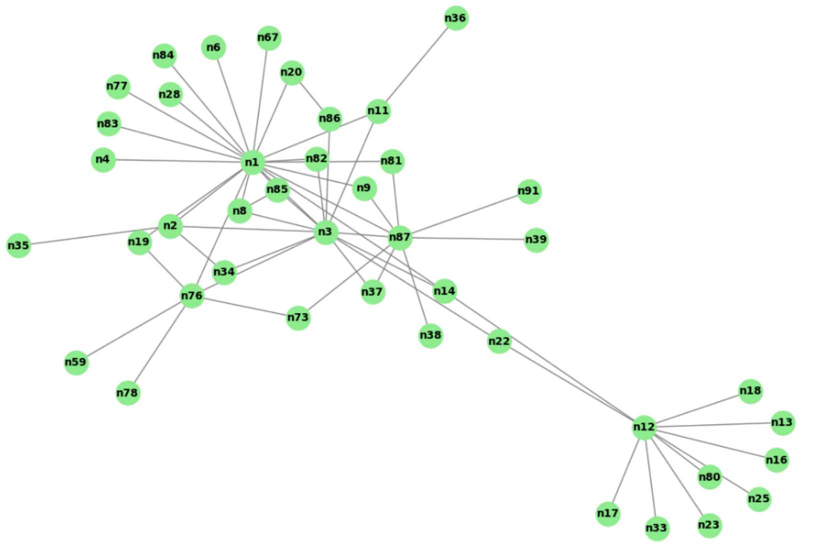
Are there other actors that play an important role but are not on the list of investigation (i.e., actors who are not among the 23 listed above) ? List them, and explain why they are important.

Top Betweenness Centrality: n1, n12, n3, n76, n87, n41, n89, n14, n83, n82

	index	Mean	Std. dev	Variance	Activity
0	n1	0.655051	0.238535	0.056899	11.0
25	n12	0.167562	0.187327	0.035092	10.0
4	n3	0.129403	0.179745	0.032308	11.0
22	n76	0.083791	0.099159	0.009833	10.0
54	n87	0.061327	0.092958	0.008641	6.0
84	n41	0.050369	0.167055	0.027907	3.0
2	n89	0.047948	0.072882	0.005312	6.0
42	n14	0.032671	0.081351	0.006618	7.0
3	n83	0.031785	0.034215	0.001171	11.0
50	n82	0.029196	0.052430	0.002749	6.0

n14, and n41 have high **Betweenness Centrality**, which are strategic connectors within the network, but not in the lists. If we remove the n41 and



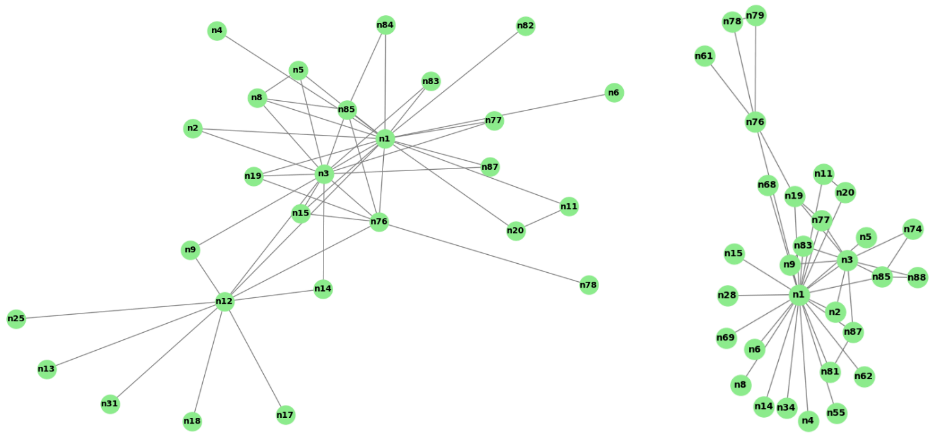


From the graph, we could also see n14 plays a crucial role in certain phases.

Top 10 Eigenvector Centrality: n1, n3, n85, n76, n83, n8, n12, n87, n2, n9

	index	Mean	Std. dev	Variance	Activity
0	n1	0.546391	0.117044	0.013699	11.0
4	n3	0.298095	0.124629	0.015533	11.0
7	n85	0.190612	0.065860	0.004338	11.0
22	n76	0.165877	0.080282	0.006445	10.0
3	n83	0.153522	0.083390	0.006954	11.0
14	n8	0.152394	0.067531	0.004560	10.0
25	n12	0.141893	0.152335	0.023206	10.0
54	n87	0.141080	0.163028	0.026578	6.0
9	**n2**	**0.114302**	**0.062870**	**0.003953**	**9.0**
23	**n9**	**0.100680**	**0.081816**	**0.006694**	**8.0**

n2, n9 also in top 10 values, which indicate influence due to their connections, but not in the lists.



In addition, **n9 and n2** also show a highly significant connection with other central nodes and sub-groups in the graph.

The remaining two questions will concern the directed graphs derived from the CAVIAR data.

Part (i):(2 points)Include your answer to this question in your written report. (150 words, 250 word limit.)

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

Hint: If we were to study the directed version of the graph, instead of the undirected, what would you learn from comparing the in-degree and out-degree centralities of each actor? Similarly, what would you learn from the left- and right-eigenvector centralities, respectively?

Solution:

The directed graphs (*in-degree and out-degree centralities, left- and right-eigenvector centralities)* represent the dynamic relationship between criminals.

The keypoints of information can be carried by directed version of the criminal work as follows:

1. **Flow and Distribution:** The resources and information are passed by out-degree centralities.
2. **Control and Center:** The power of people who had central control is indicated by the out-degree centralities.
3. **Influence:** The people directly affecting more people in the criminal network could be shown in right eigenvector centralities.
4. **Importance:** The people who play a crucial role in the criminal network could be shown in left eigenvector centralities.
5. **Asymmetric relationship:** The hierarchy of structure is shown in the graph. Some people are more centrical and irreplaceable, in contrast, outsiders are more vulnerable in the graph. Marginal people are replaced at different phases.

Part (j):(4 points)Include your answer to this question in your written report. (300 words, 400 word limit)

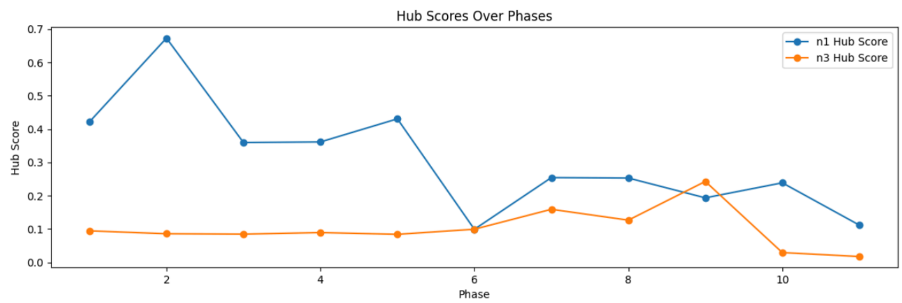
Recall the definition of hubs and authorities. Compute the hub and authority score of each actor, and for each phase. (Remember to load the adjacency data again this time using create_using = nx.DiGraph().)

With networkx you can use the nx.algorithms.link_analysis.hits function, set max_iter=1000000 for best results.

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)?

Solution:

```python
# Compute HITS scores for each phase
hits_scores = {}
for phase_num in range(1, 12):
    hits = nx.hits(G[phase_num], max_iter=1000000)
    hits_scores[phase_num] = hits

# Extract hub and authority scores for n1 and n3
for phase_num in range(1, 12):
    hub_n1 = hits_scores[phase_num][0].get("n1", 0)
    authority_n1 = hits_scores[phase_num][1].get("n1", 0)
    hub_n3 = hits_scores[phase_num][0].get("n3", 0)
    authority_n3 = hits_scores[phase_num][1].get("n3", 0)
    print(f"Phase {phase_num}:")
    print(f"  n1 - Hub: {hub_n1:.4f}, Authority: {authority_n1:.4f}")
    print(f"  n3 - Hub: {hub_n3:.4f}, Authority: {authority_n3:.4f}")

# Analysis of the relationship evolution between n1 and n3
```



Hub Scores: n1 (Daniel Serero) has the highest score in most of the phases, indicating he is the core of the criminal network, but his role as hub is decreasing with time. **n3 (Pierre Perlini)** plays a role as a crucial connector to others, even exceeding n1 in phase9.

Authority Scores: The authority score of **n1 (Daniel Serero)** is lower than **n3 (Pierre Perlini)** in the beginning phases, but **n1** surpasses **n3** in phase6. The evolving of the authority indicates n3 has most power in the start, and **n1** overtake the power from **n3.**

Comparisons to the results in Part (g), From Hub Scores and Authority Scores, indicates the leadership and the irreplaceable connection of n1.

§ Problem 4

The last part of this assignment is an open-ended project. Choose a sociologically interesting question about either the CAVIAR network or the facebook or twitter network from the recitation notebook section (for the social network data, go to https://snap.stanford.edu/data/ego-Facebook.html and https://snap.stanford.edu/data/ego-Twitter.html, or any other publicly available network data set, e.g. those at https://snap.stanford.edu/data/index.html.

Try to answer your own question using the data. You can subset the data in whichever way you desire as long as it is (sociologically) meaningful.

Think of how you may want to ***subset the data*** *in the context of the CAVIAR or the social networks, or the publicly available network data set you have chosen.*

Sociologically Interesting Question:

How has the immediate network of central criminal figures in the CAVIAR network evolved in response to police operations?

(2 points) Describes ***methodology for network analysis.***

(2 points) Grader is convinced that the methodology makes sense for the question to be answered.

(2 points) Presents results, ***including figures and/or statistics****, which address the question of interest.*

(2 points) The ***described methodology has been applied*** *in complete and the results shown (that is, the author did not forget to include anything they discussed in the methodology.)*

Adequately discusses the results obtained.

(2 points) Question does not need to be successfully answered, but the grader should be convinced that the author has answered the question to the ***best ability of the methodology presented.***

(1 point) Provides commentary on what was discovered, what were ***the limitations of the methods,*** *what may have been surprising to discover, etc.*

(1 point) Award this point if the question was successfully answered to the ***grader's satisfaction.***

Solution:

Methodology

1. **Data Loading and Preprocessing**

   The information is loaded with csv, and preprocess with the index.

2. **Network Visualization**

   For each phase, a network is constructed using NetworkX. Nodes represent players, and edges represent the relationships.

3. **Degree Centrality Analysis and Community Detection**

   Values of the Degree Centrality and Community Detection could indicate the importance of the figures.

4. **Key figures Centrality Visualization**

   Get the most central figures and compute them in Visualization.

5. **Immediate Network (Ego Network) Analysis**

The key players(n1, n3, n12) got from step4, the immediate network includes the node itself, all directly connected nodes, and the edges connecting these neighbors to the ego node.

Step 1: Data Loading and Preprocessing

```
import pandas as pd
import networkx as nx
import matplotlib.pyplot as plt
from networkx.algorithms import community
from collections import Counter

# Load data and create graphs
phases = {}
G = {}
for i in range(1, 12):
    var_name = "phase" + str(i)
    file_name = "https://raw.githubusercontent.com/ragini30/Networks-Homework/main/" + var_name + ".csv"
    phases[i] = pd.read_csv(file_name, index_col=["players"])
    phases[i].columns = "n" + phases[i].columns
    phases[i].index = phases[i].columns
    phases[i][phases[i] > 0] = 1
    G[i] = nx.from_pandas_adjacency(phases[i])
    G[i].name = var_name
```

Step 2: Network Visualization

Analyze Changes in Network Metrics

```
# Function to plot network
def plot_network(G, phase_num):
    pos = nx.spring_layout(G)  # positions for all nodes
    plt.figure(figsize=(12, 8))
    nx.draw_networkx_nodes(G, pos, node_size=500)
    nx.draw_networkx_edges(G, pos, width=1.0, alpha=0.5)
    nx.draw_networkx_labels(G, pos, font_size=12)
    plt.title(f"Network Visualization for Phase {phase_num}")
    plt.show()
```

Step 3: Degree Centrality Analysis and Community Detection

Evaluate changes in metrics such as degree centrality, such as Betweenness Centrality Analysis, Community Detection, and the role of these figures.

```python
# Function to analyze network
def analyze_network(G, phase_num):
    print(f"--- Phase {phase_num} ---")
    print(f"Number of nodes: {G.number_of_nodes()}")
    print(f"Number of edges: {G.number_of_edges()}")

    degrees = dict(G.degree())
    betweenness = nx.betweenness_centrality(G)
    top_degrees = sorted(degrees.items(), key=lambda x: x[1], reverse=True)[:5]
    top_betweenness = sorted(betweenness.items(), key=lambda x: x[1], reverse=True)[:5]

    print("Top 5 nodes by degree:")
    for node, degree in top_degrees:
        print(f"Node: {node}, Degree: {degree}")

    print("\nTop 5 nodes by betweenness centrality:")
    for node, bc in top_betweenness:
        print(f"Node: {node}, Betweenness Centrality: {bc}")

    # Community detection
    communities = community.greedy_modularity_communities(G)
    print("\nCommunities detected:")
    for i, comm in enumerate(communities):
        print(f"Community {i+1}: {comm}")

    # Plot the network
    plot_network(G, phase_num)

# Analyze and plot each phase
for phase_num in range(1, 12):
    analyze_network(G[phase_num], phase_num)
```

And the results for each phases as follows:

```
--- Phase 1 ---
Number of nodes: 15
Number of edges: 18
Top 5 nodes by degree:
Node: n1, Degree: 12
Node: n88, Degree: 4
Node: n85, Degree: 3
Node: n6, Degree: 3
Node: n89, Degree: 2

Top 5 nodes by betweenness centrality:
Node: n1, Betweenness Centrality: 0.9065934065934067
Node: n89, Betweenness Centrality: 0.14285714285714288
Node: n88, Betweenness Centrality: 0.053113553113553126
Node: n83, Betweenness Centrality: 0.03663003663003663
Node: n85, Betweenness Centrality: 0.03663003663003663

Communities detected:
Community 1: frozenset({'n90', 'n2', 'n8', 'n4', 'n5', 'n54', 'n1', 'n64'})
Community 2: frozenset({'n88', 'n83', 'n3', 'n6', 'n85'})
Community 3: frozenset({'n7', 'n89'})
```

```
--- Phase 2 ---
Number of nodes: 24
Number of edges: 28
Top 5 nodes by degree:
Node: n1, Degree: 19
Node: n89, Degree: 3
Node: n3, Degree: 3
Node: n88, Degree: 3
Node: n8, Degree: 3

Top 5 nodes by betweenness centrality:
Node: n1, Betweenness Centrality: 0.942687747035573
Node: n89, Betweenness Centrality: 0.12384716732542819
Node: n83, Betweenness Centrality: 0.08695652173913043
Node: n88, Betweenness Centrality: 0.08300395256916995
Node: n76, Betweenness Centrality: 0.038208168642951255

Communities detected:
Community 1: frozenset({'n90', 'n98', 'n2', 'n97', 'n1', 'n47', 'n86', 'n9', 'n55', 'n12', 'n5', 'n56'})
Community 2: frozenset({'n7', 'n88', 'n89', 'n11', 'n85', 'n76'})
Community 3: frozenset({'n10', 'n64', 'n8', 'n3'})
Community 4: frozenset({'n6', 'n83'})

--- Phase 3 ---
Number of nodes: 33
Number of edges: 56
Top 5 nodes by degree:
Node: n1, Degree: 27
Node: n3, Degree: 9
Node: n83, Degree: 8
Node: n9, Degree: 5
Node: n49, Degree: 5

Top 5 nodes by betweenness centrality:
Node: n1, Betweenness Centrality: 0.829502688172043
Node: n3, Betweenness Centrality: 0.09549731182795697
Node: n9, Betweenness Centrality: 0.06754032258064516
Node: n5, Betweenness Centrality: 0.0625
Node: n83, Betweenness Centrality: 0.046572580645161295

Communities detected:
Community 1: frozenset({'n10', 'n89', 'n50', 'n32', 'n48', 'n2', 'n34', 'n1', 'n35', 'n7', 'n55', 'n56'})
Community 2: frozenset({'n88', 'n11', 'n3', 'n12', 'n6', 'n99', 'n85', 'n76'})
Community 3: frozenset({'n49', 'n90', 'n83', 'n52', 'n107', 'n84', 'n86'})
Community 4: frozenset({'n13', 'n8', 'n9', 'n51'})
Community 5: frozenset({'n4', 'n5'})


--- Phase 4 ---
Number of nodes: 33
Number of edges: 48
Top 5 nodes by degree:
Node: n1, Degree: 23
Node: n83, Degree: 7
Node: n3, Degree: 7
Node: n89, Degree: 6
Node: n85, Degree: 5

Top 5 nodes by betweenness centrality:
Node: n1, Betweenness Centrality: 0.8393097158218126
Node: n89, Betweenness Centrality: 0.19621255760368667
Node: n3, Betweenness Centrality: 0.09043778801843318
Node: n83, Betweenness Centrality: 0.0795890937019969
Node: n8, Betweenness Centrality: 0.0625

Communities detected:
Community 1: frozenset({'n88', 'n11', 'n2', 'n47', 'n15', 'n1', 'n4', 'n53', 'n6', 'n5', 'n85', 'n76'})
Community 2: frozenset({'n9', 'n90', 'n8', 'n3', 'n12', 'n35', 'n31'})
Community 3: frozenset({'n63', 'n83', 'n52', 'n107', 'n84', 'n106', 'n86'})
Community 4: frozenset({'n49', 'n7', 'n89', 'n109', 'n14'})
Community 5: frozenset({'n13', 'n51'})


--- Phase 5 ---
Number of nodes: 32
Number of edges: 39
Top 5 nodes by degree:
Node: n1, Degree: 22
Node: n12, Degree: 8
Node: n3, Degree: 5
Node: n31, Degree: 4
Node: n89, Degree: 3

Top 5 nodes by betweenness centrality:
Node: n1, Betweenness Centrality: 0.8838709677419355
Node: n12, Betweenness Centrality: 0.2698924731182796
Node: n89, Betweenness Centrality: 0.06451612903225806
Node: n83, Betweenness Centrality: 0.06451612903225806
Node: n85, Betweenness Centrality: 0.06451612903225806

Communities detected:
Community 1: frozenset({'n11', 'n108', 'n34', 'n82', 'n8', 'n4', 'n5', 'n100', 'n84', 'n2', 'n47', 'n15', 'n1', 'n55', 'n19', 'n6'})
Community 2: frozenset({'n32', 'n18', 'n17', 'n9', 'n31', 'n13', 'n25', 'n3', 'n12'})
Community 3: frozenset({'n7', 'n89', 'n76'})
Community 4: frozenset({'n88', 'n85'})
Community 5: frozenset({'n86', 'n83'})
```

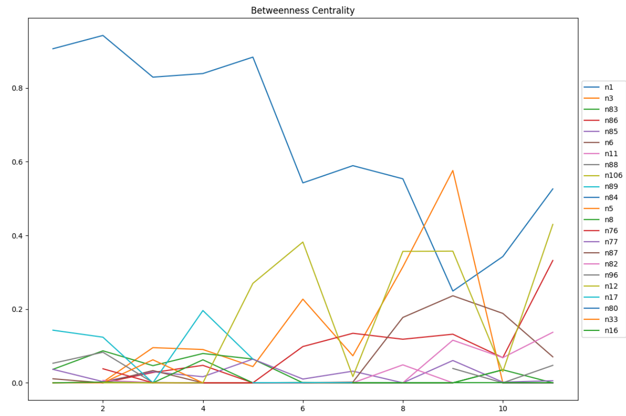
```
--- Phase 6 ---
Number of nodes: 27
Number of edges: 47
Top 5 nodes by degree:
Node: n1, Degree: 18
Node: n3, Degree: 14
Node: n12, Degree: 10
Node: n76, Degree: 7
Node: n85, Degree: 5

Top 5 nodes by betweenness centrality:
Node: n1, Betweenness Centrality: 0.5425641025641026
Node: n12, Betweenness Centrality: 0.3820512820512821
Node: n3, Betweenness Centrality: 0.22717948717948716
Node: n76, Betweenness Centrality: 0.09846153846153845
Node: n85, Betweenness Centrality: 0.010256410256410255

Communities detected:
Community 1: frozenset({'n13', 'n25', 'n18', 'n12', 'n14', 'n17', 'n9', 'n31'})
Community 2: frozenset({'n87', 'n11', 'n6', 'n4', 'n82', 'n20', 'n1'})
Community 3: frozenset({'n8', 'n2', 'n3', 'n83', 'n5', 'n77'})
Community 4: frozenset({'n78', 'n19', 'n15', 'n76'})
Community 5: frozenset({'n84', 'n85'})

--- Phase 7 ---
Number of nodes: 36
Number of edges: 49
Top 5 nodes by degree:
Node: n1, Degree: 24
Node: n3, Degree: 10
Node: n76, Degree: 5
Node: n12, Degree: 5
Node: n85, Degree: 4

Top 5 nodes by betweenness centrality:
Node: n1, Betweenness Centrality: 0.5893557422969188
Node: n76, Betweenness Centrality: 0.13445378151260506
Node: n3, Betweenness Centrality: 0.07338935574229691
Node: n85, Betweenness Centrality: 0.03165266106442576
Node: n12, Betweenness Centrality: 0.016806722689075633

Communities detected:
Community 1: frozenset({'n11', 'n69', 'n34', 'n1', 'n15', 'n6', 'n55', 'n62', 'n8', 'n20', 'n28', 'n68', 'n4', 'n14', 'n5'})
Community 2: frozenset({'n88', 'n2', 'n83', 'n3', 'n85', 'n74', 'n9'})
Community 3: frozenset({'n78', 'n79', 'n19', 'n61', 'n77', 'n76'})
Community 4: frozenset({'n16', 'n18', 'n75', 'n22', 'n12', 'n17'})
Community 5: frozenset({'n87', 'n81'})

--- Phase 8 ---
Number of nodes: 42
Number of edges: 58
Top 5 nodes by degree:
Node: n1, Degree: 20
Node: n3, Degree: 13
Node: n12, Degree: 10
Node: n87, Degree: 9
Node: n76, Degree: 6

Top 5 nodes by betweenness centrality:
Node: n1, Betweenness Centrality: 0.553658536585366
Node: n12, Betweenness Centrality: 0.3567073170731707
Node: n3, Betweenness Centrality: 0.31463414634146336
Node: n14, Betweenness Centrality: 0.2634146341463415
Node: n87, Betweenness Centrality: 0.1774390243902439

Communities detected:
Community 1: frozenset({'n16', 'n23', 'n18', 'n12', 'n17', 'n80', 'n13', 'n25', 'n22', 'n33', 'n14'})
Community 2: frozenset({'n11', 'n83', 'n36', 'n77', 'n1', 'n6', 'n67', 'n28', 'n4', 'n84'})
Community 3: frozenset({'n2', 'n82', 'n34', 'n35', 'n86', 'n8', 'n3', 'n20', 'n85'})
Community 4: frozenset({'n87', 'n38', 'n39', 'n37', 'n91', 'n81', 'n9'})
Community 5: frozenset({'n78', 'n59', 'n19', 'n73', 'n76'})

--- Phase 9 ---
Number of nodes: 34
Number of edges: 44
Top 5 nodes by degree:
Node: n3, Degree: 11
Node: n1, Degree: 10
Node: n12, Degree: 8
Node: n87, Degree: 8
Node: n82, Degree: 6

Top 5 nodes by betweenness centrality:
Node: n3, Betweenness Centrality: 0.5762310606060607
Node: n12, Betweenness Centrality: 0.3573232323232323
Node: n1, Betweenness Centrality: 0.2490530303030303
Node: n87, Betweenness Centrality: 0.2362689393939394
Node: n76, Betweenness Centrality: 0.13194444444444445

Communities detected:
Community 1: frozenset({'n13', 'n16', 'n29', 'n18', 'n12', 'n14', 'n17'})
Community 2: frozenset({'n88', 'n89', 'n90', 'n2', 'n3', 'n85'})
Community 3: frozenset({'n7', 'n78', 'n59', 'n79', 'n101', 'n76'})
Community 4: frozenset({'n11', 'n8', 'n83', 'n6', 'n1'})
Community 5: frozenset({'n87', 'n37', 'n41', 'n81', 'n105'})
Community 6: frozenset({'n96', 'n46', 'n82', 'n36', 'n30'})
```

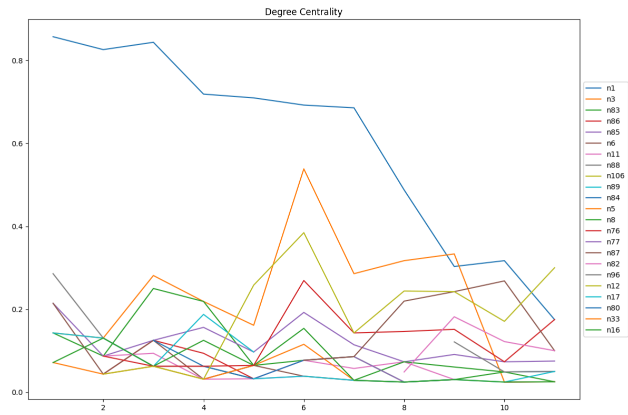
```
--- Phase 10 ---
Number of nodes: 42
Number of edges: 50
Top 5 nodes by degree:
Node: n1, Degree: 13
Node: n87, Degree: 11
Node: n37, Degree: 9
Node: n12, Degree: 7
Node: n82, Degree: 5

Top 5 nodes by betweenness centrality:
Node: n1, Betweenness Centrality: 0.3426829268292683
Node: n87, Betweenness Centrality: 0.18841463414634146
Node: n37, Betweenness Centrality: 0.174390243902439
Node: n76, Betweenness Centrality: 0.06829268292682927
Node: n82, Betweenness Centrality: 0.06829268292682927

Communities detected:
Community 1: frozenset({'n70', 'n83', 'n1', 'n71', 'n86', 'n9', 'n19', 'n8', 'n3', 'n4', 'n84', 'n85', 'n76'})
Community 2: frozenset({'n65', 'n16', 'n18', 'n24', 'n17', 'n58', 'n22', 'n12', 'n14'})
Community 3: frozenset({'n27', 'n81', 'n73', 'n40', 'n21', 'n87', 'n38', 'n105'})
Community 4: frozenset({'n45', 'n42', 'n37', 'n93', 'n95', 'n41', 'n44'})
Community 5: frozenset({'n96', 'n46', 'n82'})
Community 6: frozenset({'n103', 'n104'})

--- Phase 11 ---
Number of nodes: 41
Number of edges: 50
Top 5 nodes by degree:
Node: n12, Degree: 12
Node: n41, Degree: 9
Node: n1, Degree: 7
Node: n76, Degree: 7
Node: n79, Degree: 5

Top 5 nodes by betweenness centrality:
Node: n41, Betweenness Centrality: 0.554059829059829
Node: n1, Betweenness Centrality: 0.5262820512820513
Node: n12, Betweenness Centrality: 0.42991452991452994
Node: n76, Betweenness Centrality: 0.33205128205128204
Node: n79, Betweenness Centrality: 0.17884615384615385

Communities detected:
Community 1: frozenset({'n88', 'n11', 'n81', 'n1', 'n87', 'n61', 'n84', 'n72', 'n85', 'n76'})
Community 2: frozenset({'n65', 'n16', 'n66', 'n18', 'n17', 'n13', 'n42', 'n3', 'n12', 'n26'})
Community 3: frozenset({'n94', 'n43', 'n93', 'n37', 'n27', 'n41', 'n92'})
Community 4: frozenset({'n59', 'n79', 'n102', 'n101', 'n78'})
Community 5: frozenset({'n96', 'n36', 'n46', 'n82'})
Community 6: frozenset({'n14', 'n24', 'n58'})
Community 7: frozenset({'n86', 'n83'})
```

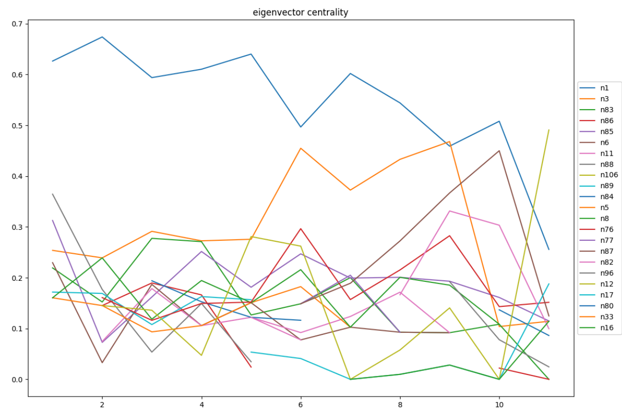
Step 4: 23 Key figures Centrality Visualization
The role of 23 of the players [1, 3, 83, 86,85,6,11,88,106,89,84,5,8,76,77,87,82,96,12,17,80,33,16] in "Serero organization"

Compute degree centrality, betweenness centrality and eigenvector centrality and Visualized for the 23 players across all phases

```
def compute_metric(metric_fn):
    return { i: metric_fn(G[i]) for i in range(1,12) }

deg = compute_metric(nx.degree_centrality)  ## networkx already calculates normali
bet = compute_metric(lambda g: nx.betweenness_centrality(g, normalized = True))
eig = compute_metric(nx.eigenvector_centrality)  ## normalized value returned

deg_df = pd.DataFrame.from_dict(deg, orient='index')
bet_df = pd.DataFrame.from_dict(bet, orient='index')
eig_df = pd.DataFrame.from_dict(eig, orient='index')
```





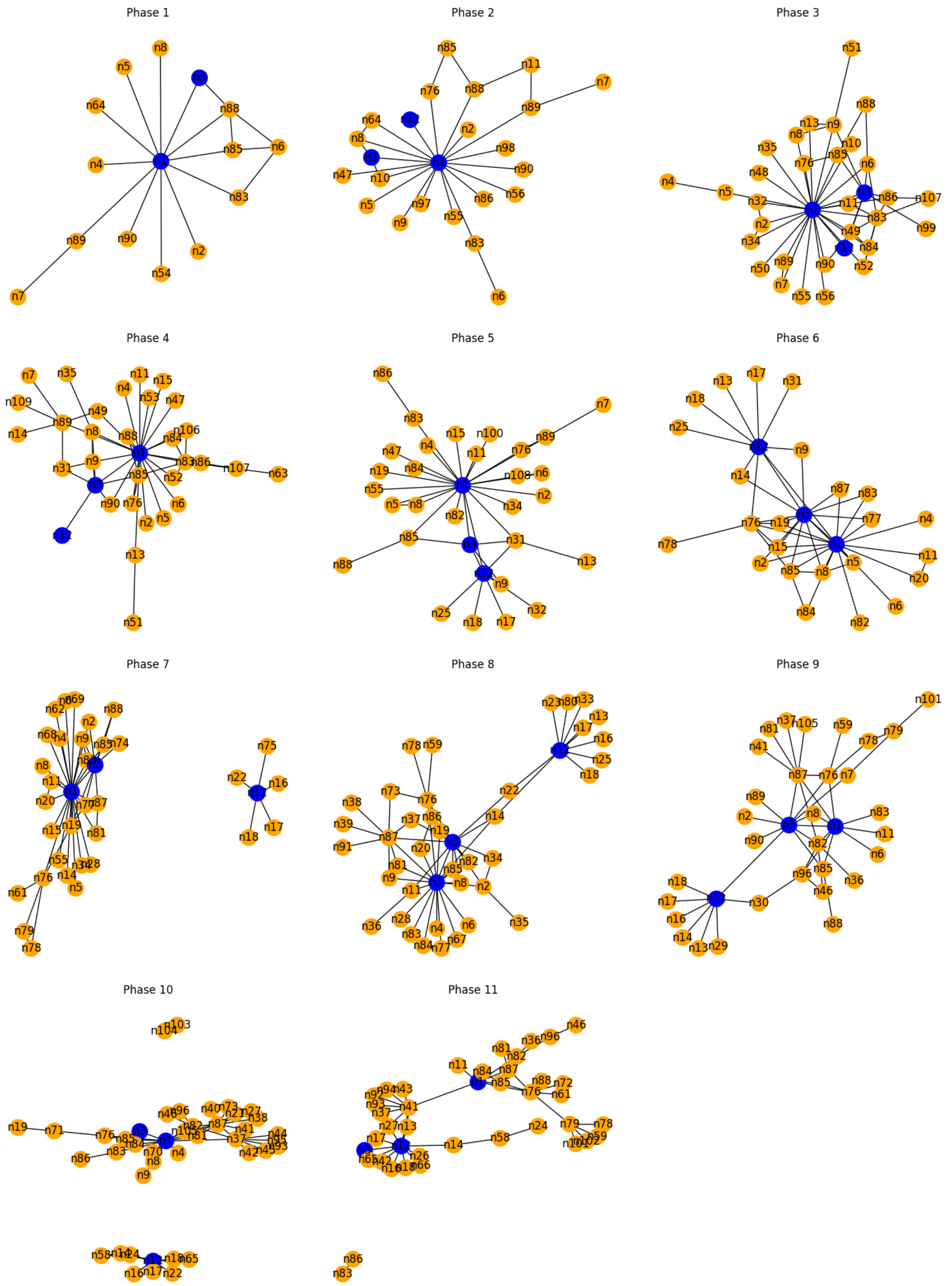


We got the top 3 central figures: [n1, n3, n12] and visualized the relationship with other figures in time.

Step 5: Temporal network of central criminal figures

The central criminal figures are marked as red and visualized with temporal network analysis.

```
# Plot all phases with highlighted nodes
fig, axes = plt.subplots(4, 3, figsize=(15, 20))
axes = axes.flatten()

key_players = {'n1', 'n3', 'n12'}

for i in range(1, 12):
    graph = G[i]
    color = ['red' if node in key_players else 'green' for node in graph]
    nx.draw_spring(graph, node_color=color, with_labels=True, ax=axes[i-1])
    axes[i-1].set_title(f"Phase {i}")

# Remove any empty subplot (in case of 11 phases)
for j in range(i, len(axes)):
    fig.delaxes(axes[j])

plt.tight_layout()
plt.show()
```



Analyzing the Evolving Network:

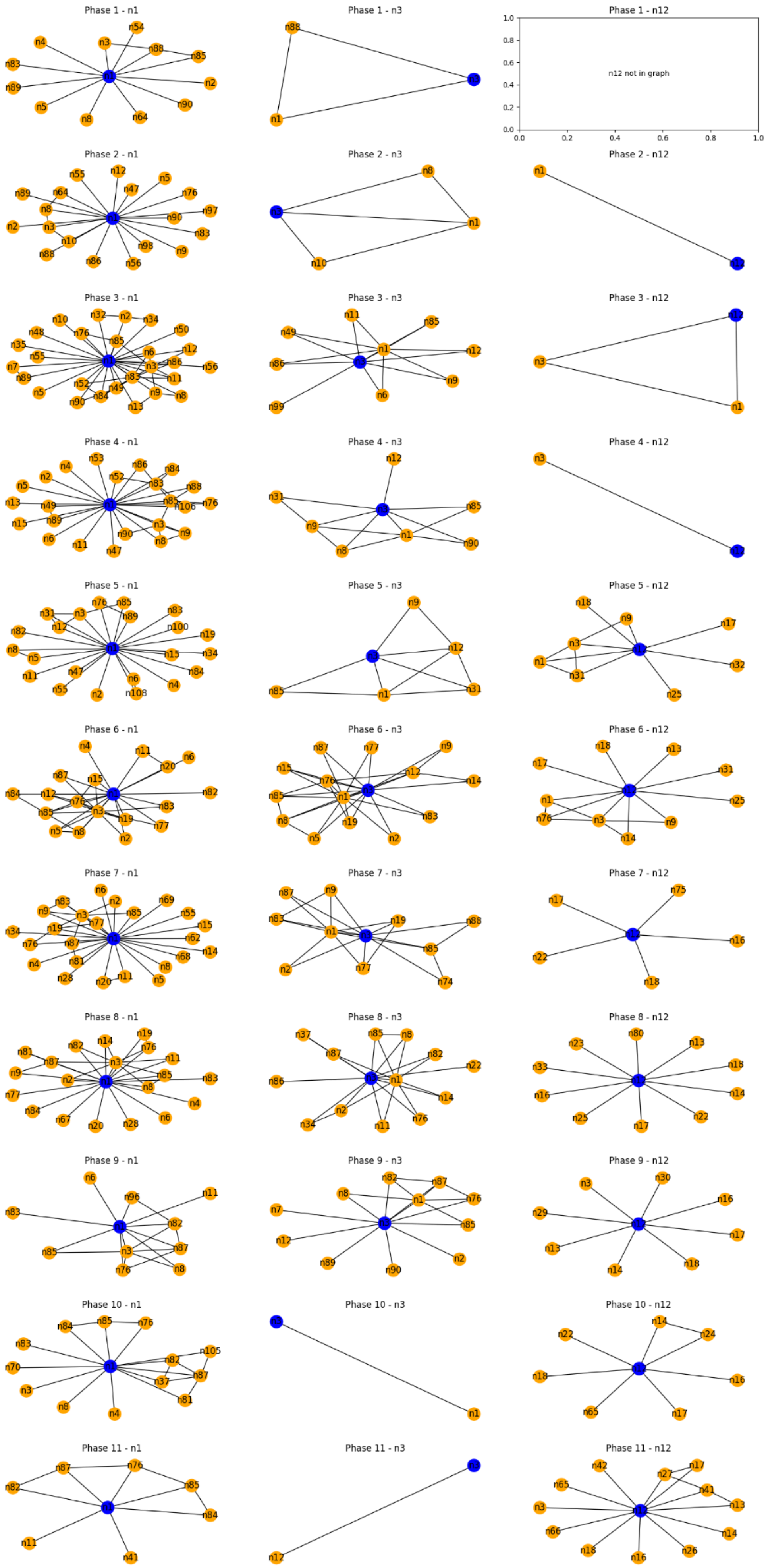
The evolution of the networks were analyzed to understand how the network adapted to police pressure. n1 is highly connected to financial supporters (such as n85, 86, 83) in all phases, even the power and control has already taken over in the late phases.
After the investigation and the change of the shipping location,we can see the pattern shifted from initial centralizations to decentralization and compartmentalization.

Step 6:Immediate Networks for Key Players:

Immediate network, or called ego-network, is a graph representation consisting of the key figure (node) itself and all other relatives(nodes) directly connected.

```
import pandas as pd
import networkx as nx
import matplotlib.pyplot as plt

import pandas as pd
import networkx as nx
import matplotlib.pyplot as plt


# Function to plot immediate network of a node
def plot_immediate_network(graph, node, ax, title):
    if node in graph:
        ego_graph = nx.ego_graph(graph, node)
        color = ['red' if n == node else 'green' for n in ego_graph]
        nx.draw_spring(ego_graph, node_color=color, with_labels=True, ax=ax)
    else:
        ax.text(0.5, 0.5, f"{node} not in graph", horizontalalignment='center', verticalalignment='center')
    ax.set_title(title)

# Nodes for which to compute immediate networks
nodes_of_interest = ['n1', 'n3', 'n12']

# Create subplots
fig, axes = plt.subplots(11, len(nodes_of_interest), figsize=(15, 30))

# Plot immediate networks for each node in each phase
for i in range(1, 12):
    graph = G[i]
    for j, node in enumerate(nodes_of_interest):
        plot_immediate_network(graph, node, axes[i-1][j], f"Phase {i} - {node}")

# Adjust layout
plt.tight_layout()
plt.show()
```

We get subgraphs containing central figures and their immediate neighbors.



Phase 1 - n1
Phase 1 - n3
Phase 1 - n12
n12 not in graph
Phase 2 - n1
Phase 2 - n3
Phase 2 - n12
Phase 3 - n1
Phase 3 - n3
Phase 3 - n12
Phase 4 - n1
Phase 4 - n3
Phase 4 - n12
Phase 5 - n1
Phase 5 - n3
Phase 5 - n12
Phase 6 - n1
Phase 6 - n3
Phase 6 - n12
Phase 7 - n1
Phase 7 - n3
Phase 7 - n12
Phase 8 - n1
Phase 8 - n3
Phase 8 - n12
Phase 9 - n1
Phase 9 - n3
Phase 9 - n12
Phase 10 - n1
Phase 10 - n3
Phase 10 - n12
Phase 11 - n1
Phase 11 - n3
Phase 11 - n12

Discussion of the results:
The shift of power is visualized in the graph, all the figures are highly centralized around n1 in the beginning of the phases, but after the first seizure, the emergence of the n12, start to connect with more figures and form his own subgroups and decentralization towards cocaine trafficking.

The limitations of the methods:

1. **Incompleteness of Data:** Only part of data is shown in the network, lack of the whole
2. **Local Structure:** The method shows the local structure, unable to represent the whole picture of the global perspective.

Surprising Findings:
From each phase, even the police got a couple of seizures and the drugs, but the resilience and the adaption of the whole network evolving rapidly, especially in phase 6 and phase 9, the adaptation of the network shows the extreme change within traffickers after operations.
The distinct phases of decentralization and compartmentalization reflect that the whole operation is designed to fight with police force.

Phase 4	1 seizure	$2,500,000	300 kg of marijuana
Phase 6	3 seizures	$1,300,000	2 x 15 kg of marijuana + 1 x 2 kg of cocaine
Phase 7	1 seizure	$3,500,000	401 kg of marijuana
Phase 8	1 seizure	$360,000	9 kg of cocaine
Phase 9	2 seizures	$4,300,000	2 kg of cocaine + 1 x 500 kg marijuana
Phase 10	1 seizure	$18,700,000	2200 kg of marijuana
Phase 11	2 seizures	$1,300,000	12 kg of cocaine + 11 kg of cocaine

Conclusion:

Each type of the visualization enables us to examine the changes in the local and global structure of the network, observing how participants adapt and resilient when faced with increasing pressures from the operation.

Overall, the perspective of the evolving of the crime has been clearly shown in the methodology applied, which is revealing the dynamic of the operation. Even though there are some limitations, it could still address couple of insights and identify key players from **Immediate Networks.**

Reference

[1] Co-citation, bibliographic coupling and leading authors, institutions and countries in the 50 years of Technological Forecasting and Social Change, Volume 165,2021,120487,ISSN 0040-1625, https://doi.org/10.1016/j.techfore.2020.120487.

[2]Caviar Case Study.[Online].https://www.kaggle.com/code/gargmanas/caviar-case-study

[3] Social Network Analysis with Network.[Online].https://domino.ai/blog/social-network-analysis-with-networkx