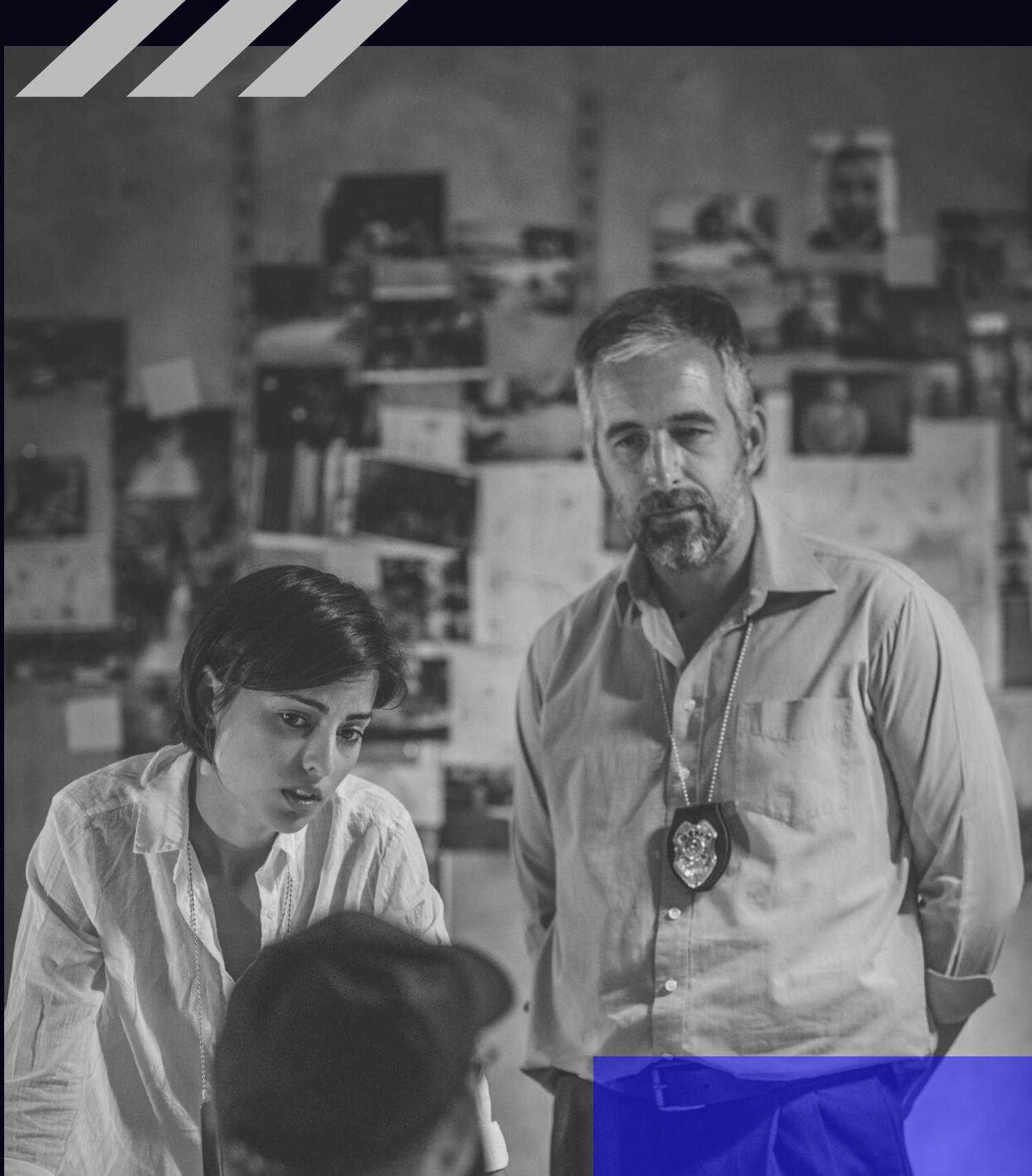


Drug Criminal Network

BADS7201: Social Network Analysis



Objective

"Analyze changes in the network structure
and predict link in the network."



Problem

"Drugs are a danger to society. Therefore, it is necessary to study the drug delivery network. in order to suppress and eliminate."



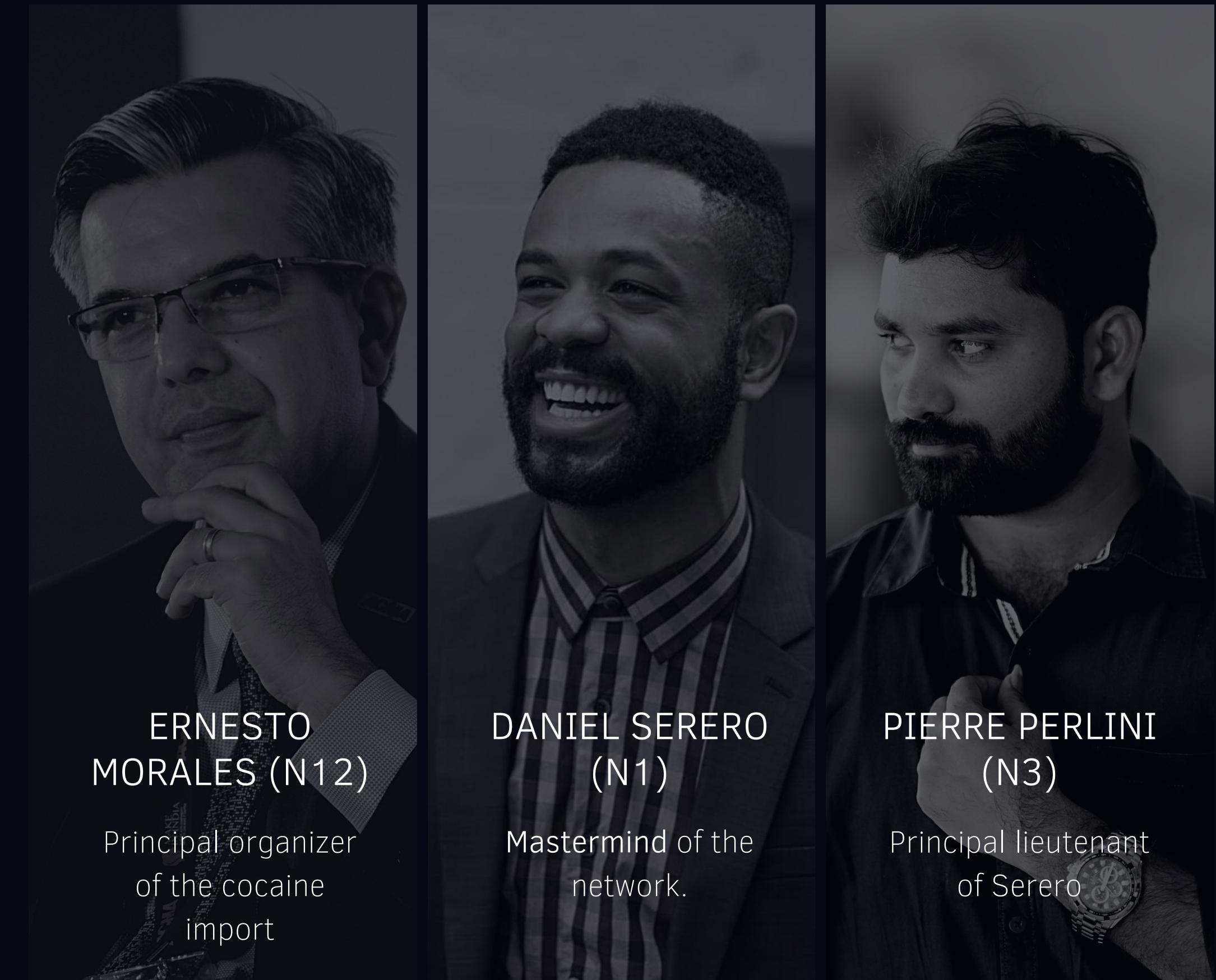
CAVIAR project

The project operated by the Montréal police and the Royal Canadian Mounted Police of Canada. The aim is to seize the drugs without arresting the perpetrators, created the opportunity to study a criminal network.

About Dataset

- THE NETWORK CONSISTS OF 110 (NUMBERED) PLAYERS
- PLAYERS 1-82 ARE THE TRAFFICKERS. PLAYERS 83-110 ARE THE NON-TRAFFICKERS
- 11 WIRETAP WARRANTS, 2 MONTHS PER WIRETAP PHASE
- SOME PHASES INCLUDED NO SEIZURES, AND OTHERS INCLUDED MULTIPLE.

The Criminal



Sample Data

Phase 2

has 24 Nodes and 28 Edges

	n1	n4	n89	n83	n3	n5	n88	n85	n90	n2	n7	n54	n6	n64	n8
n1	0	1	4	0	4	2	2	9	1	2	0	2	0	1	1
n4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n89	1	0	0	0	0	0	0	0	0	0	3	0	0	0	0
n83	1	0	0	0	0	0	0	0	0	0	0	0	5	0	0
n3	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0
n5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n88	1	0	0	0	1	0	0	3	0	0	0	0	1	0	0
n85	1	0	0	0	0	0	2	0	0	0	0	0	5	0	0
n90	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n7	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
n54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n6	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
n64	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

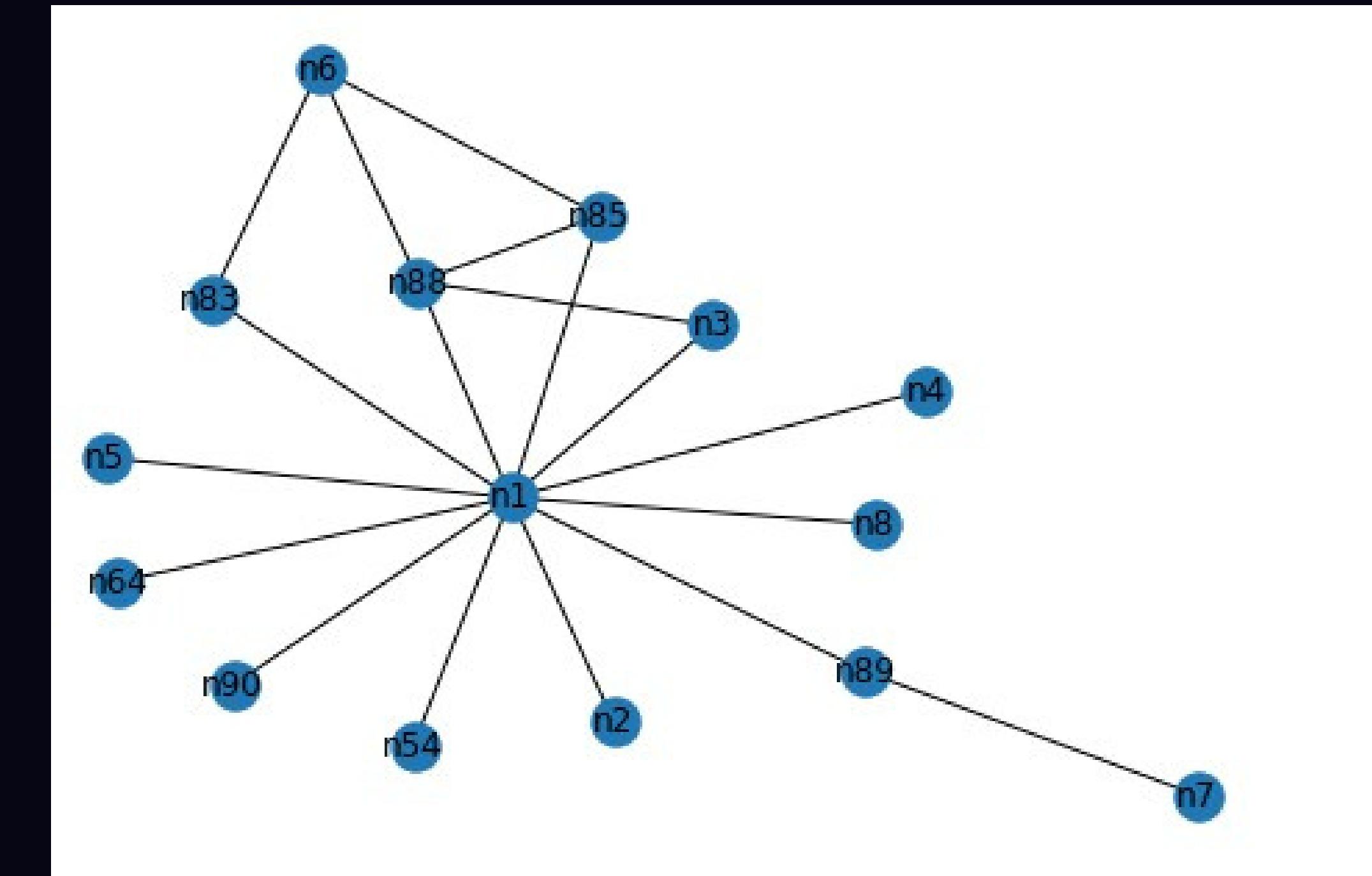
Sample Data

Phase 2 (Plot)

has 24 Nodes and 28 Edges

Node >> Criminal

Edge >> Phone Call



Caviar Network

Seizure no.	No. nodes	No. edges	Avg degree	Calls per node
1	15	26	2.4	3.9
2	24	35	2.3	5.0
3	33	68	3.4	8.5
4	33	65	2.9	13.8
5	32	50	2.4	5.0
6	27	68	3.4	24.7
7	33	60	3.4	9.6
8	42	75	2.7	7.9
9	34	60	2.6	9.3
10	42	68	2.4	11.4
11	41	72	2.4	8.6

CAVIAR project

METHODOLOGY

BADS7201: SOCIAL NETWORK ANALYSIS

METHODOLOGY

Set yourself apart from other game streamers.

1

Analyze changes in the network structure

- Plot Graph
- Calculate Network Centerarity
- Community Detection

2

Link Prediction

- Using the following algorithm
- Jaccard Coefficient
 - Adamic Adar
 - Preferential Attachment

Evaluate result with ROC AUC

CALCULATE NETWORK CENTRALITY

▶ **Betweenness centrality**
ความเป็นจุดศูนย์กลางโดยวัดการเป็นจุดคั่นกลาง หรือ ตำแหน่งที่เป็นสะพานเชื่อมจุดยอดต่างๆ (Bridgers) เข้าหากันโดยพิจารณาจากสัดส่วนของระยะทางที่สั้นที่สุด (Geodesic Path) ในการเชื่อมโยงระหว่างจุด Node แต่ละคู่

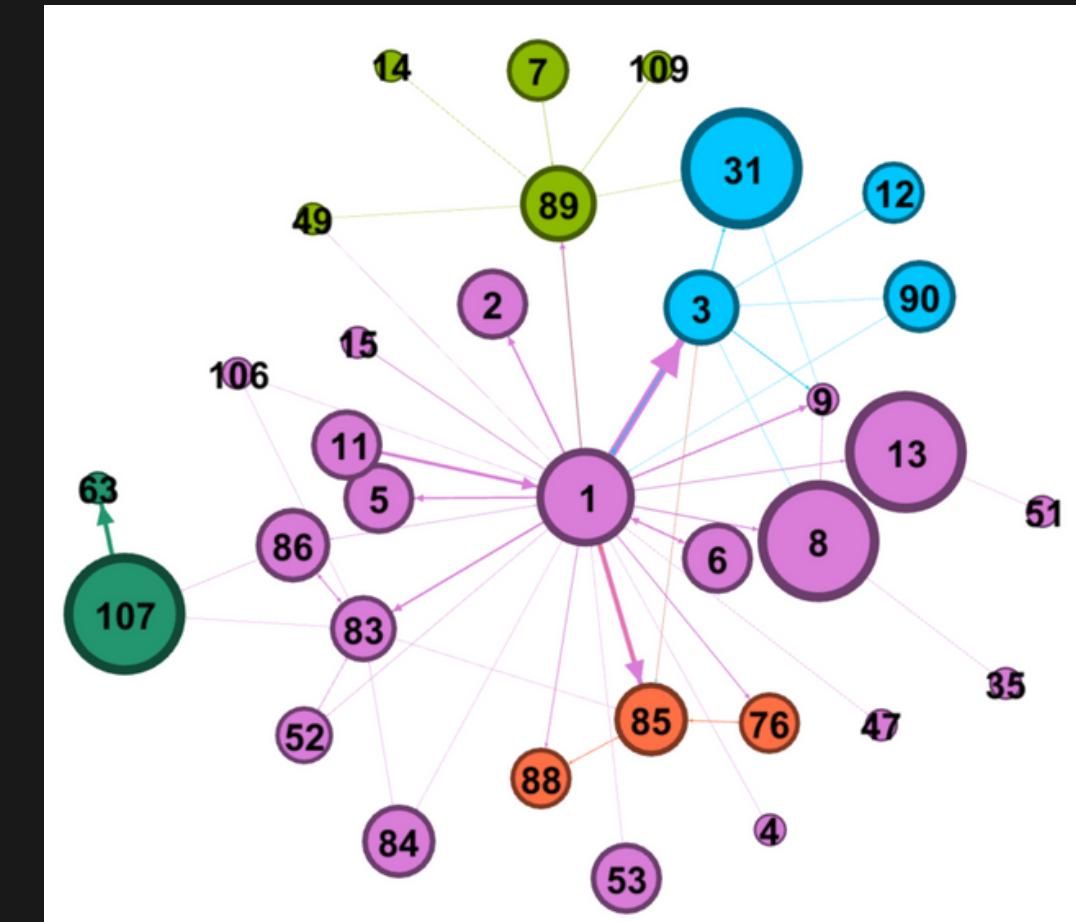
▶ **Closeness centrality**
ความเป็นจุดศูนย์กลางโดยวัดจากความใกล้ชิดกับ Node อื่นๆ และใช้ระยะทางที่สั้นที่สุดในการเข้าถึง (Geodesic Path Distance) โดยวัดได้จากจำนวนเส้นเชื่อมโยงทั้งหมด กี่ครั้งในการเดินทางจากจุดยอดหนึ่งไปยังอีกจุดยอดหนึ่ง โดยการลากผ่านจุดยอดอื่นๆ ด้วยเส้นทางที่สั้นที่สุด

▶ **Degree centrality**
ความเป็นจุดศูนย์กลางโดยวัดจากดีกรี เป็นการคำนวณค่า การเป็นจุดศูนย์กลางของการเชื่อมโยง (Hub) ซึ่งเป็นตำแหน่งที่มีอิทธิพลสูงสุดของการเชื่อมโยง โดยวัดได้จากจำนวนเส้นเชื่อมโยงทั้งหมดที่โยงมาจากจุดยอดอื่นๆ

CALCULATE USING "GEPHI"

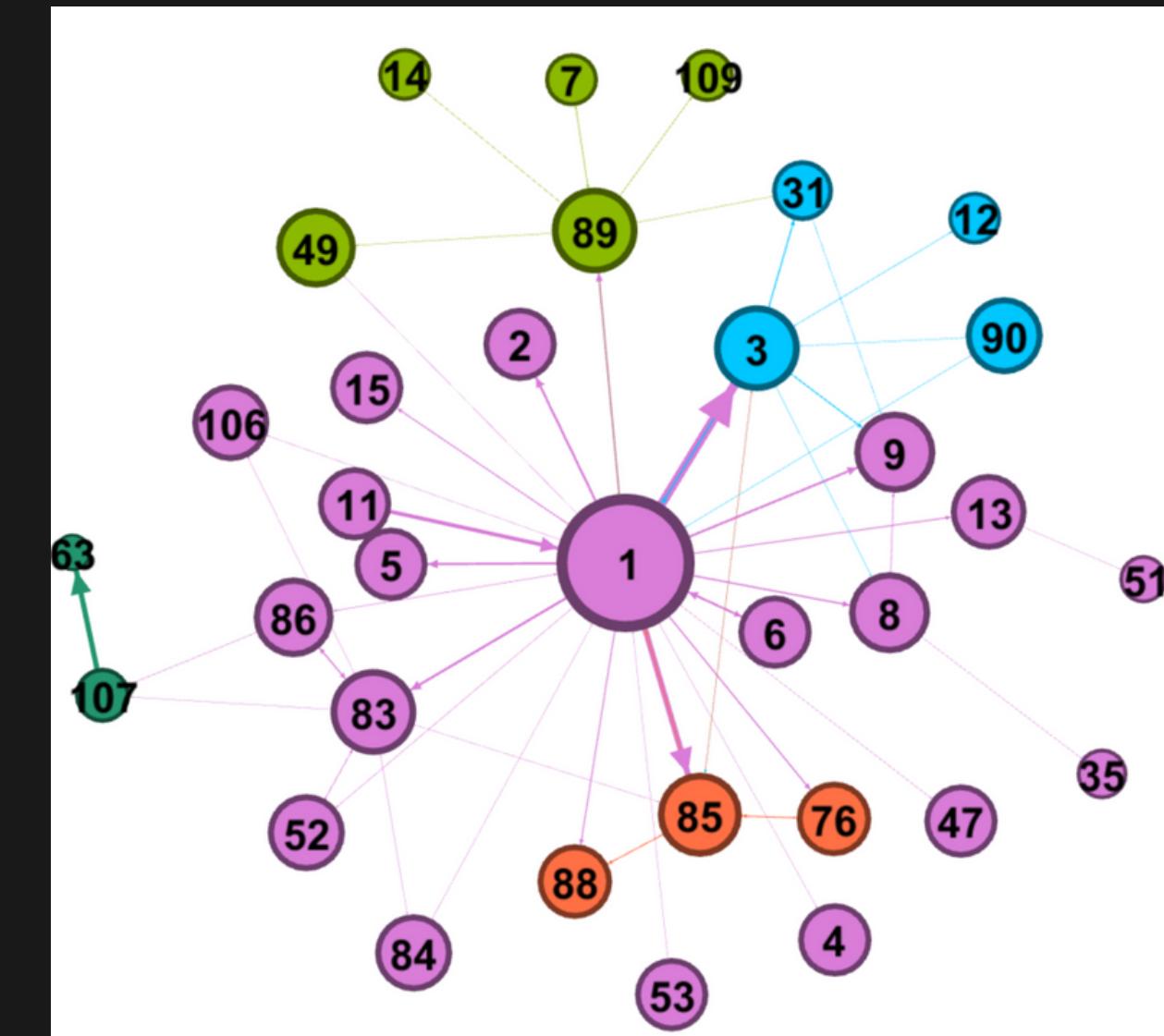
Although, the graph we plot are directed graph but we calculate centrality on undirected graph. Here why.....

Label	betweenesscentrality	closenesscentrality	Degree
8	18	1	4
13	18	1	2
31	1	1	3
107	18	1	3
1	425.5	0.738095	32
89	82	0.492063	8
3	64	0.484375	10
86	26.166667	0.469697	5
85	85.666667	0.462687	9



Calculated on Directed Graph in Gephi

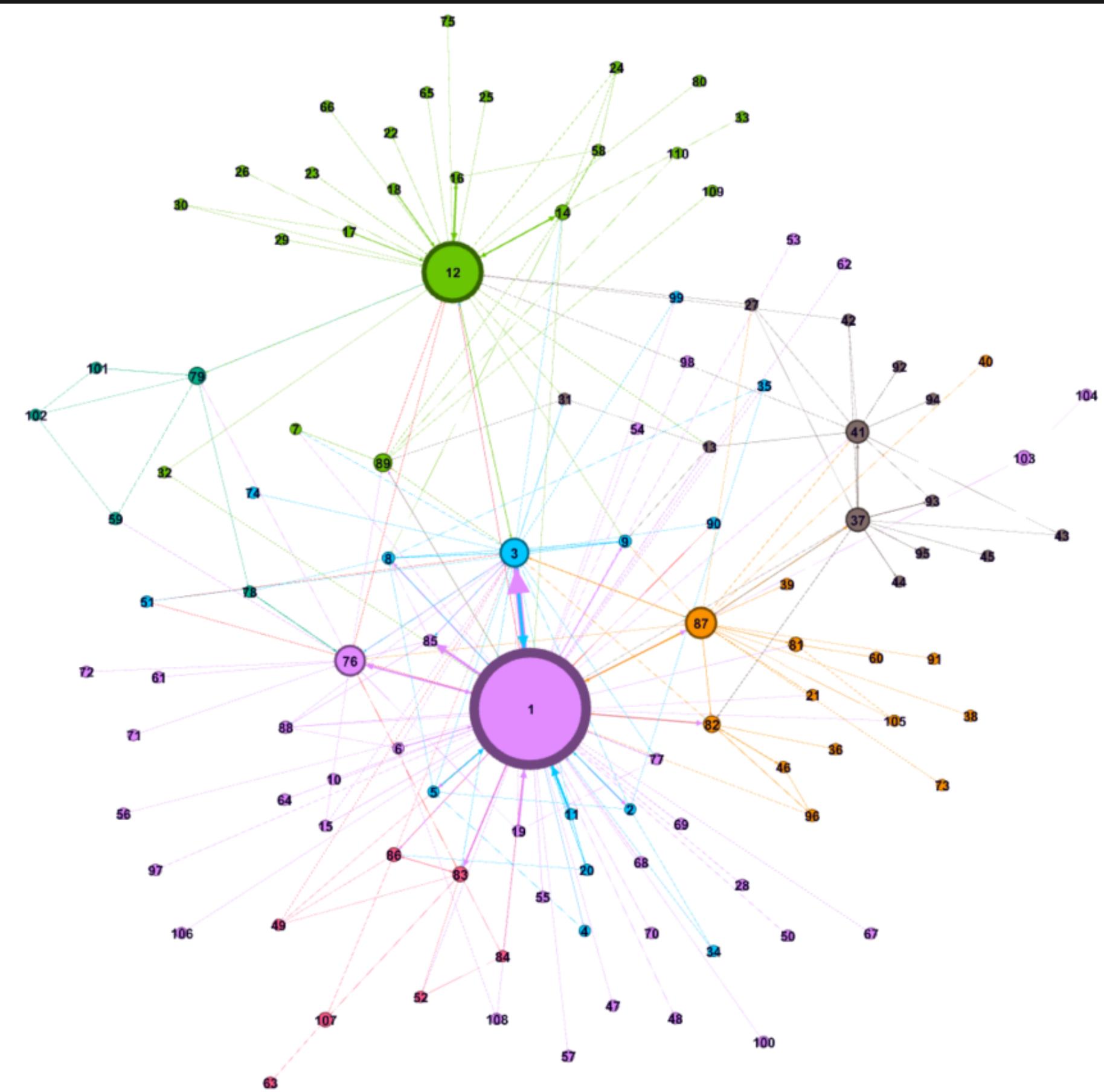
Label	betweenesscentrality	closenesscentrality	Degree
1	0.83931	0.761905	32
89	0.196213	0.5	8
83	0.079589	0.5	10
3	0.090438	0.5	10
85	0.016537	0.492308	9
8	0.0625	0.477612	4
9	0.014761	0.477612	4
86	0.047427	0.470588	5
49	0	0.470588	2



Calculated on Undirected Graph in Gephi

MAKE MORE SENSE !!

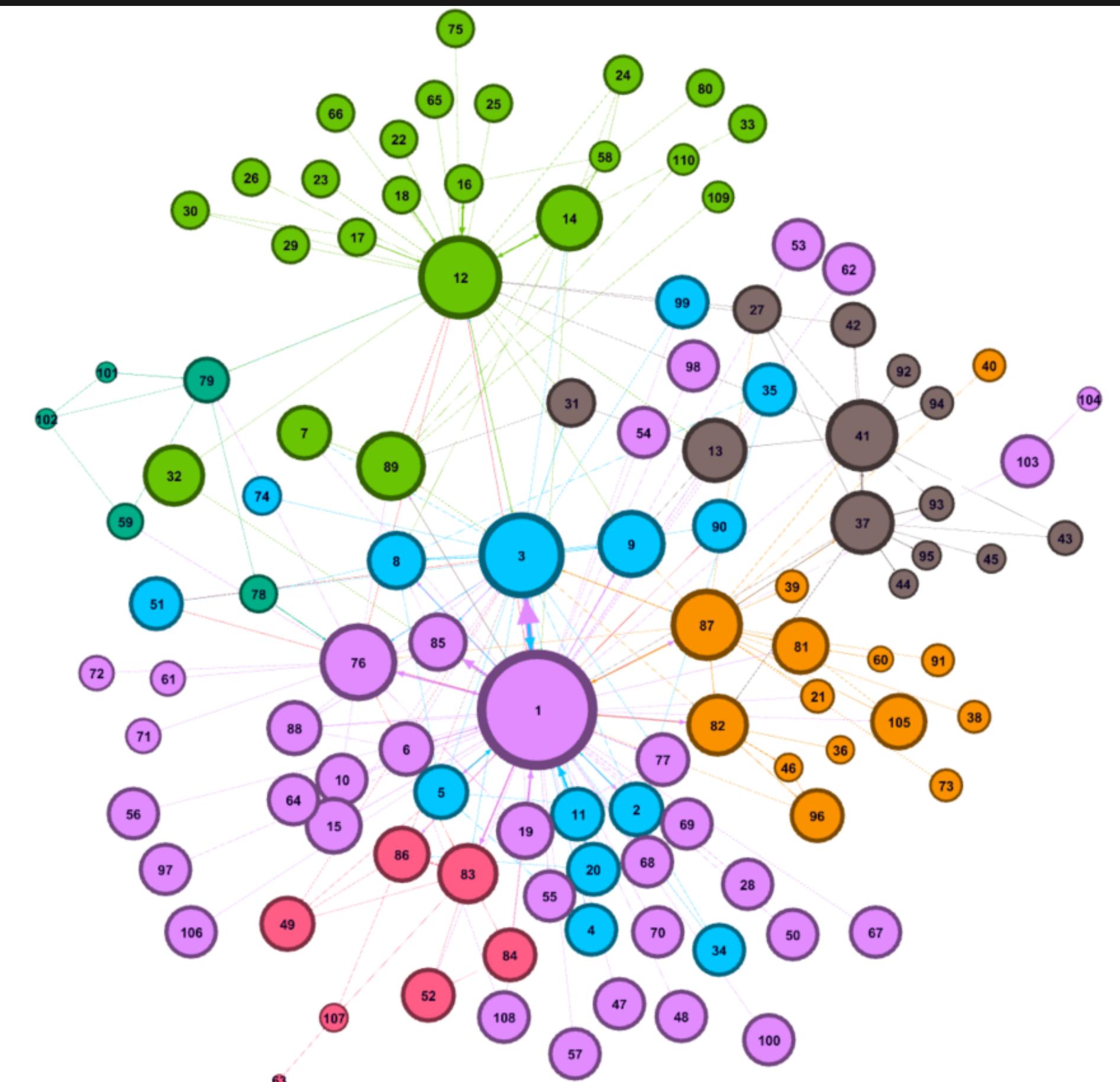
FULL DATASET - BETWEENNESS CENTRALITY



Betweenness centrality

N1 - Daniel Serero, the mastermind	0.637459
N12 - Principal organizer of the cocaine import	0.292142
N87 - Investor	0.119279
N76 - Charged with recuperating the marijuana	0.115199
N3 - Principal lieutenant of Serero	0.105205

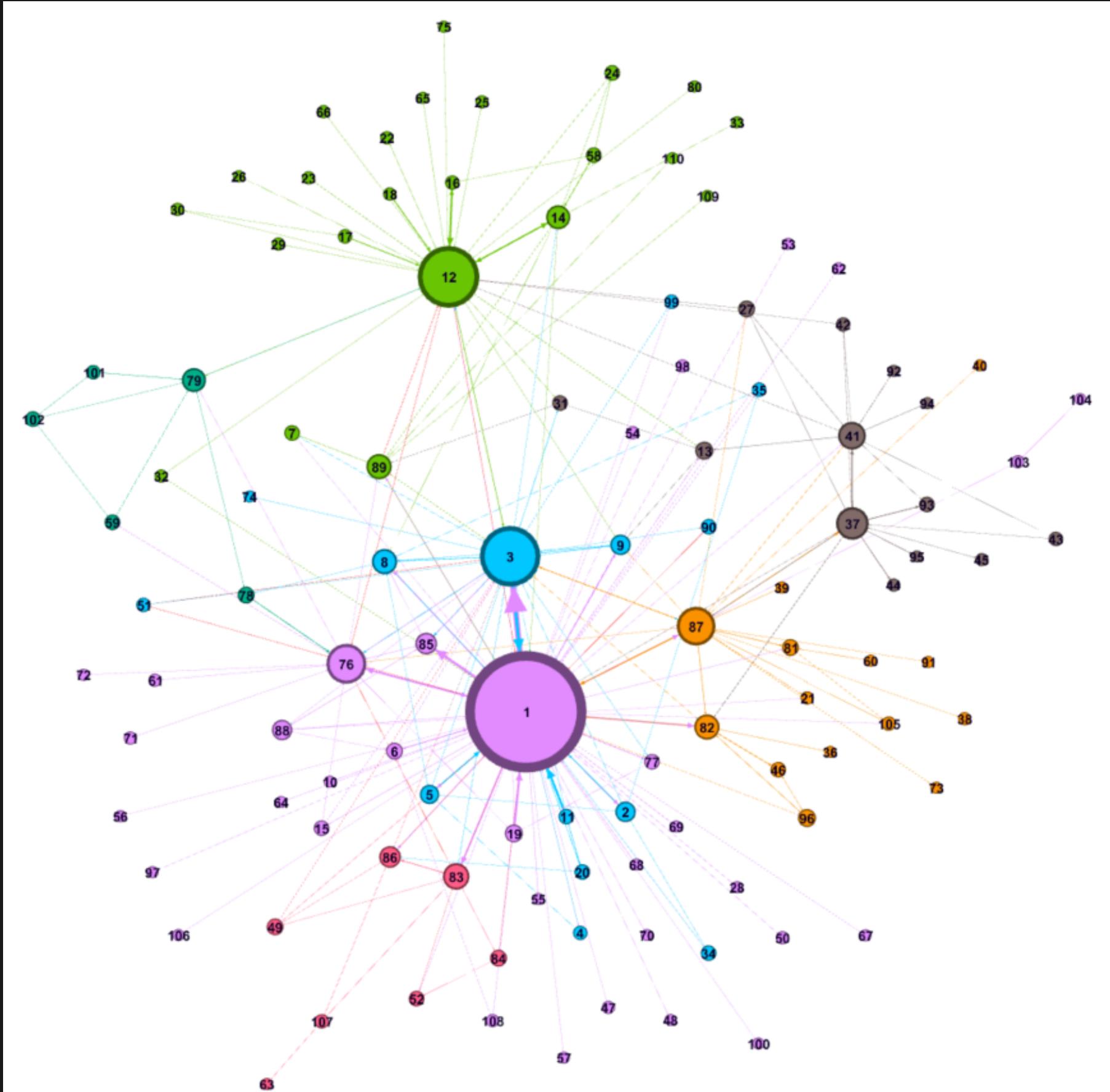
FULL DATASET - CLOSENESS CENTRALITY



Closeness centrality

N1 - Daniel Serero, the mastermind	0.677019
N3 - Principal lieutenant of Serero	0.536946
N12 - Principal organizer of the cocaine import	0.529126
N76 - Charged with recuperating the marijuana	0.50463
N41 - Non trafficcker	0.484444

FULL DATASET - DEGREE CENTERALITY



Degree centrality

N1 - Daniel Serero, the mastermind	91
N12 - Principal organizer of the cocaine import	42
N3 - Principal lieutenant of Serero	41
N76 - Charged with recuperating the marijuana.	24
N87 - Investor.	23



CRIMINAL NETWORK IN EACH PHASE

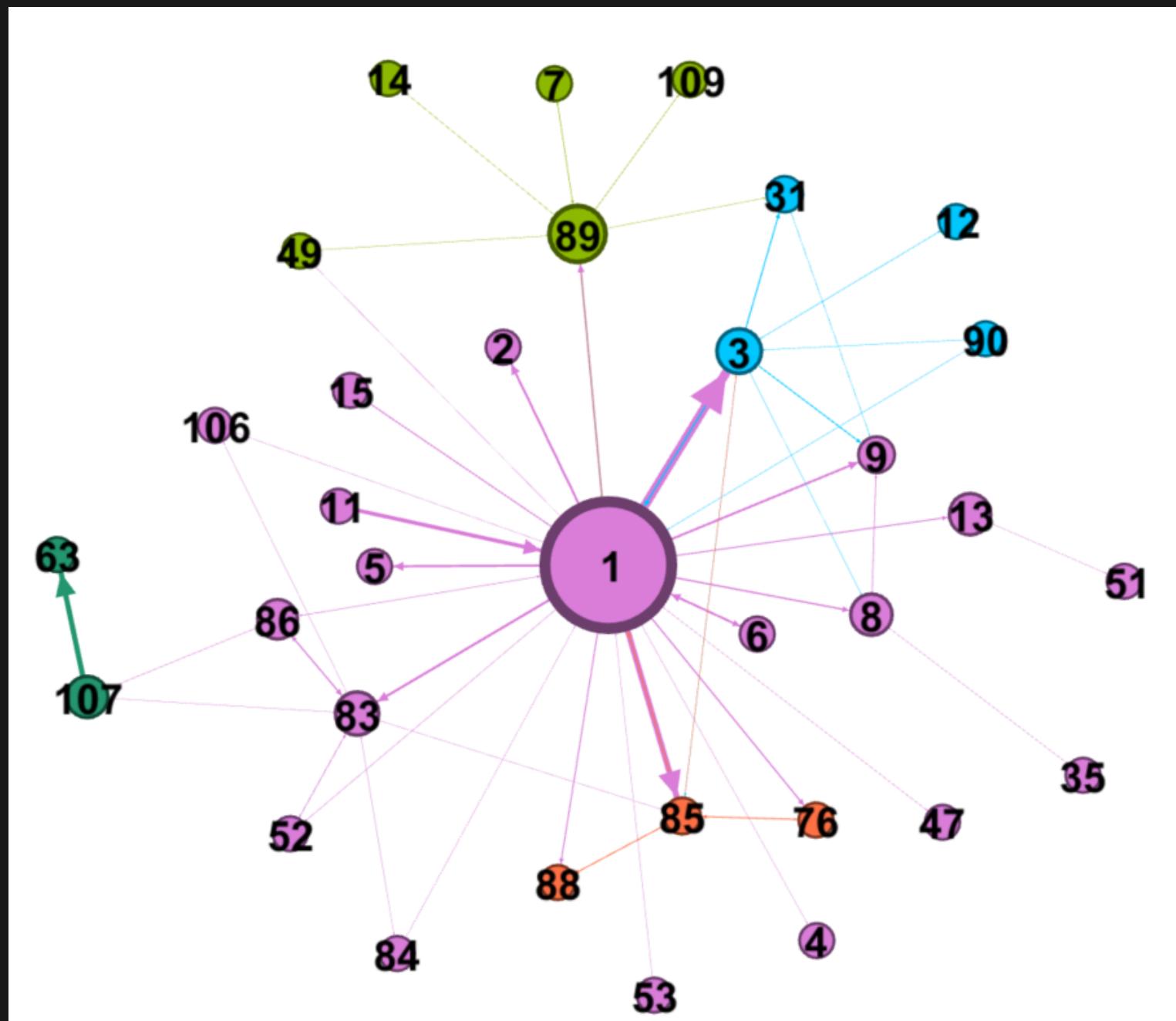
From our study, we found that in phase 4 and 5 the network had changed the most.

By checking the ground truth, it has been found that the reason a major change has occurred in phase 4 is because of the first seizure. Which then resulted in a contact with N12 of Columbia to import cocaine through the United States.

In the next slide, the network in phase 4 and 5 will be explained.



PHASE 4



Betweenness centrality

N1 - Daniel Serero, the mastermind	0.83931
N89 - Investor	0.196213
N3 - Principal lieutenant of Serero	0.090438
N83 - Investors and transporters of money	0.079589
N8 - Charged with recuperating the marijuana.	0.0625

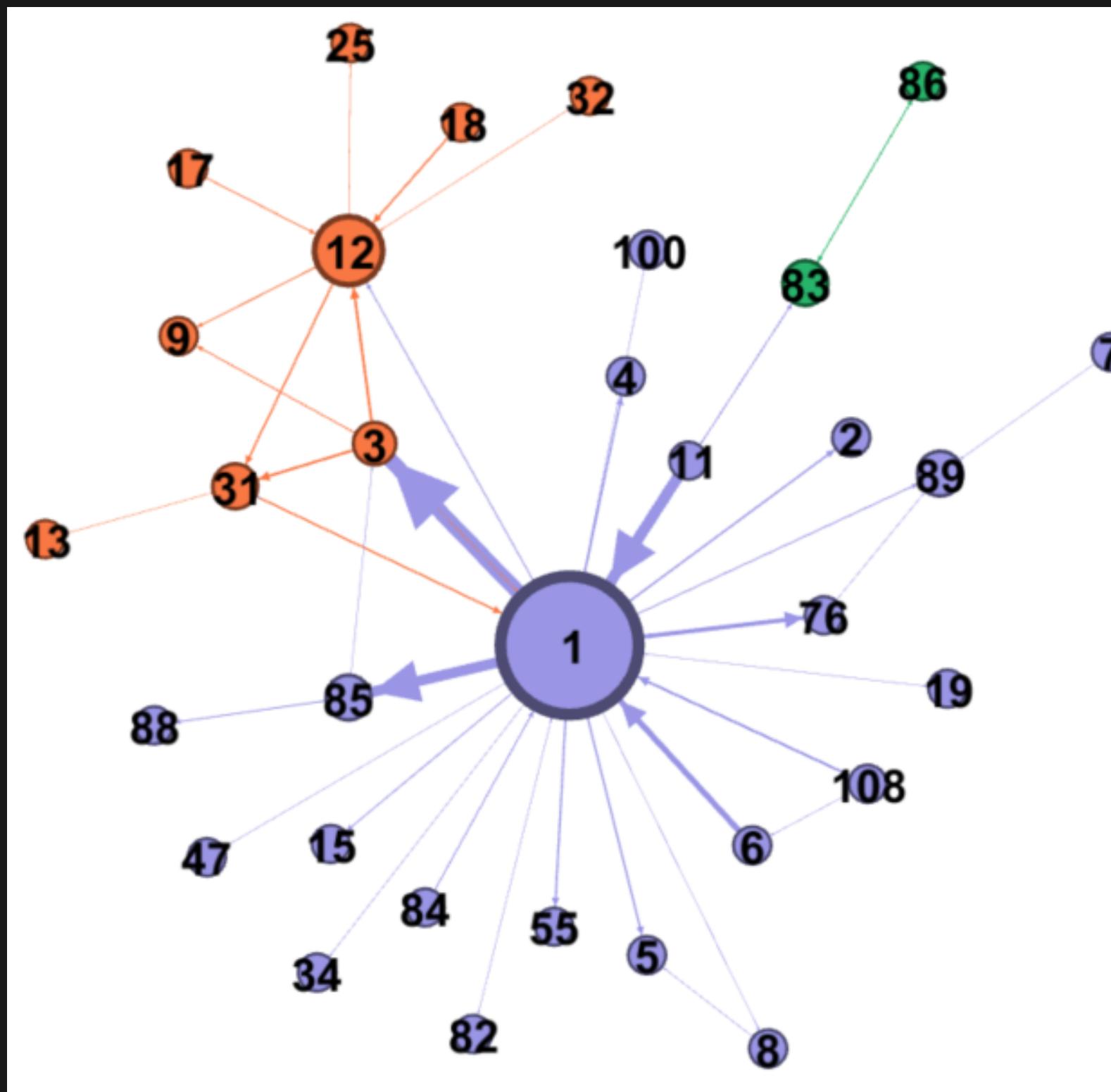
Closeness centrality

N1 - Daniel Serero, the mastermind	0.761905
N89 - Investor	0.5
N3 - Principal lieutenant of Serero	0.5
N83 - Investors and transporters of money	0.5
N85 - Takes care of financial affairs.	0.492308

Degree centrality

N1 - Daniel Serero, the mastermind	32
N3 - Principal lieutenant of Serero	10
N83 - Investors and transporters of money	10
N85 - Takes care of financial affairs.	9
N89 - Investor	8

PHASE 5



Betweenness centrality

N1 - Daniel Serero, the mastermind	0.883871
N12 - Principal organizer of the cocaine import	0.269892
N31 - Non-trafficker	0.064516
N85 - Takes care of financial affairs.	0.064516
N89 - Investor	0.064516

Closeness centrality

N1 - Daniel Serero, the mastermind	0.775
N12 - Principal organizer of the cocaine import	0.54386
N3 - Principal lieutenant of Serero	0.525424
N31 - Non-trafficker	0.508197
N85 - Takes care of financial affairs.	0.469697

Degree centrality

N1 - Daniel Serero, the mastermind	29
N12 - Principal organizer of the cocaine import	10
N3 - Principal lieutenant of Serero	7
N31 - Non-trafficker	5
N85 - Takes care of financial affairs.	4

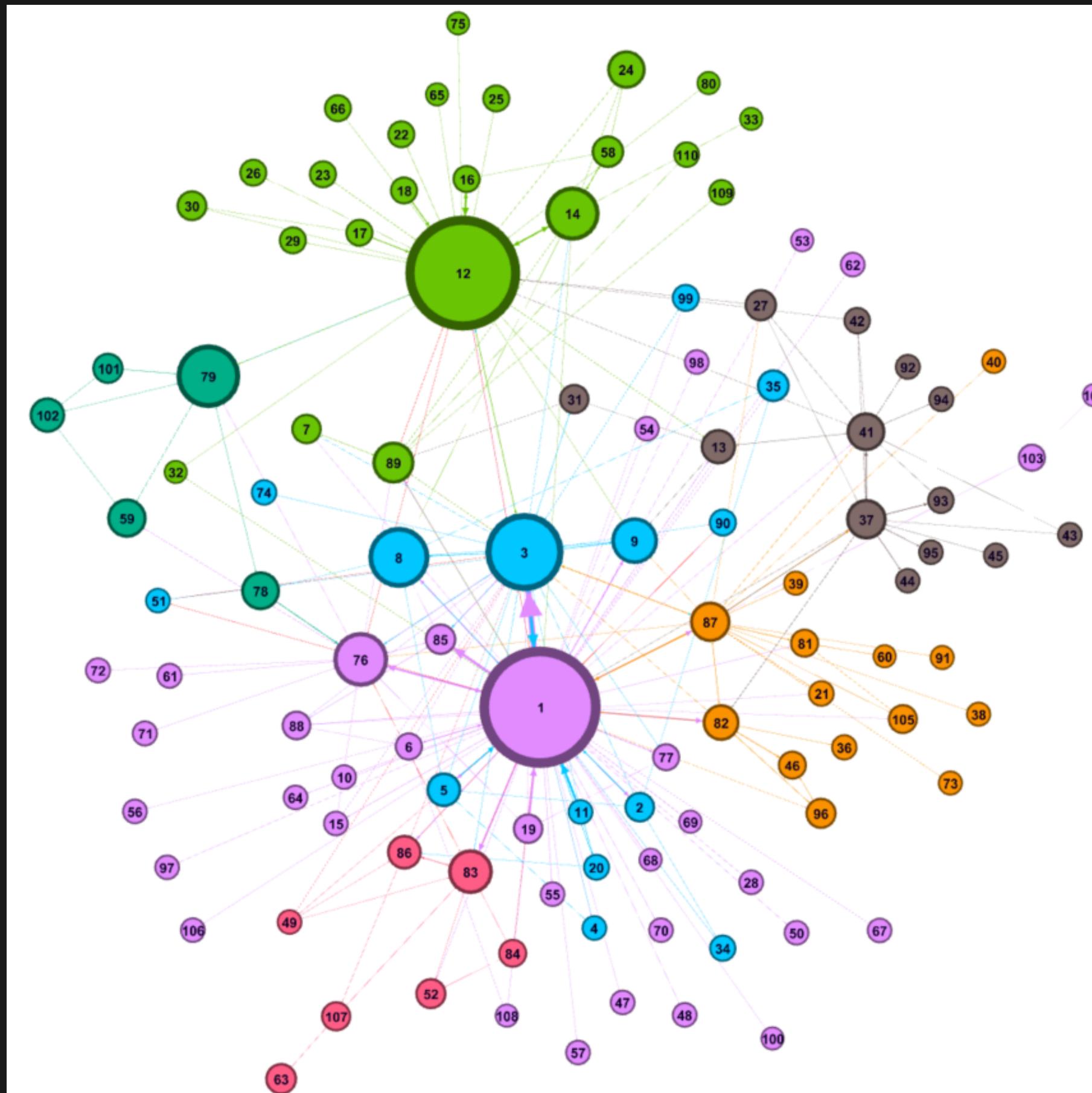
COMMUNITY DETECTION

Detect small network group in large network. In order to gain better understanding of criminal network

Calculated by using "Modularity" in Gephi to detect sub network. Here are the result



FULL DATASET



Daniel Serero Gang (N1)

Marijuana dealer (Core)

Daniel Serero Gang

Investors and transporters of money

Daniel Serero Gang

Supplier (In charge by N3)

Daniel Serero Gang

Transport Management

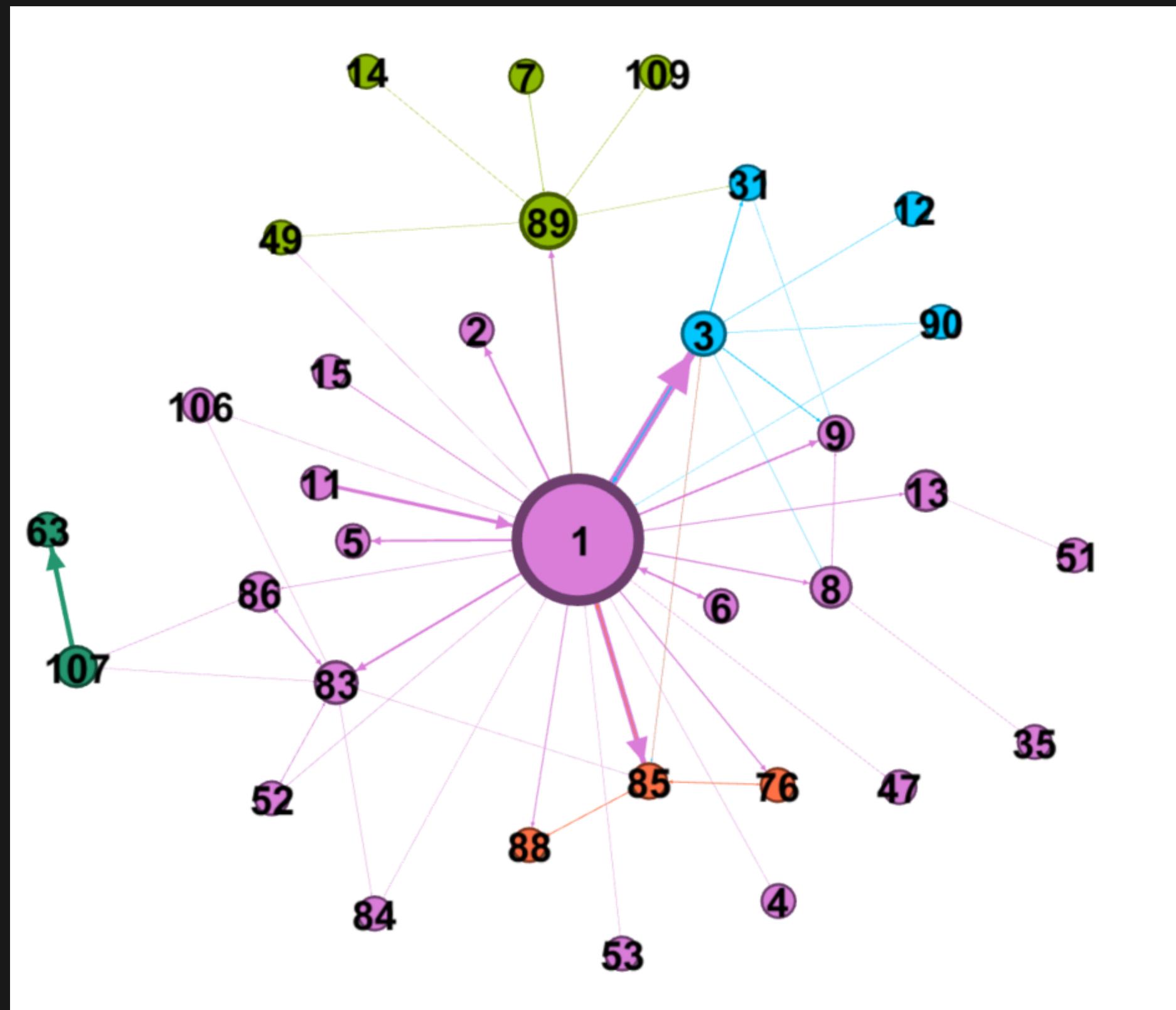
Ernesto Morales Gang (N12)

Cocaine Dealer (Core)

Non-traffickers

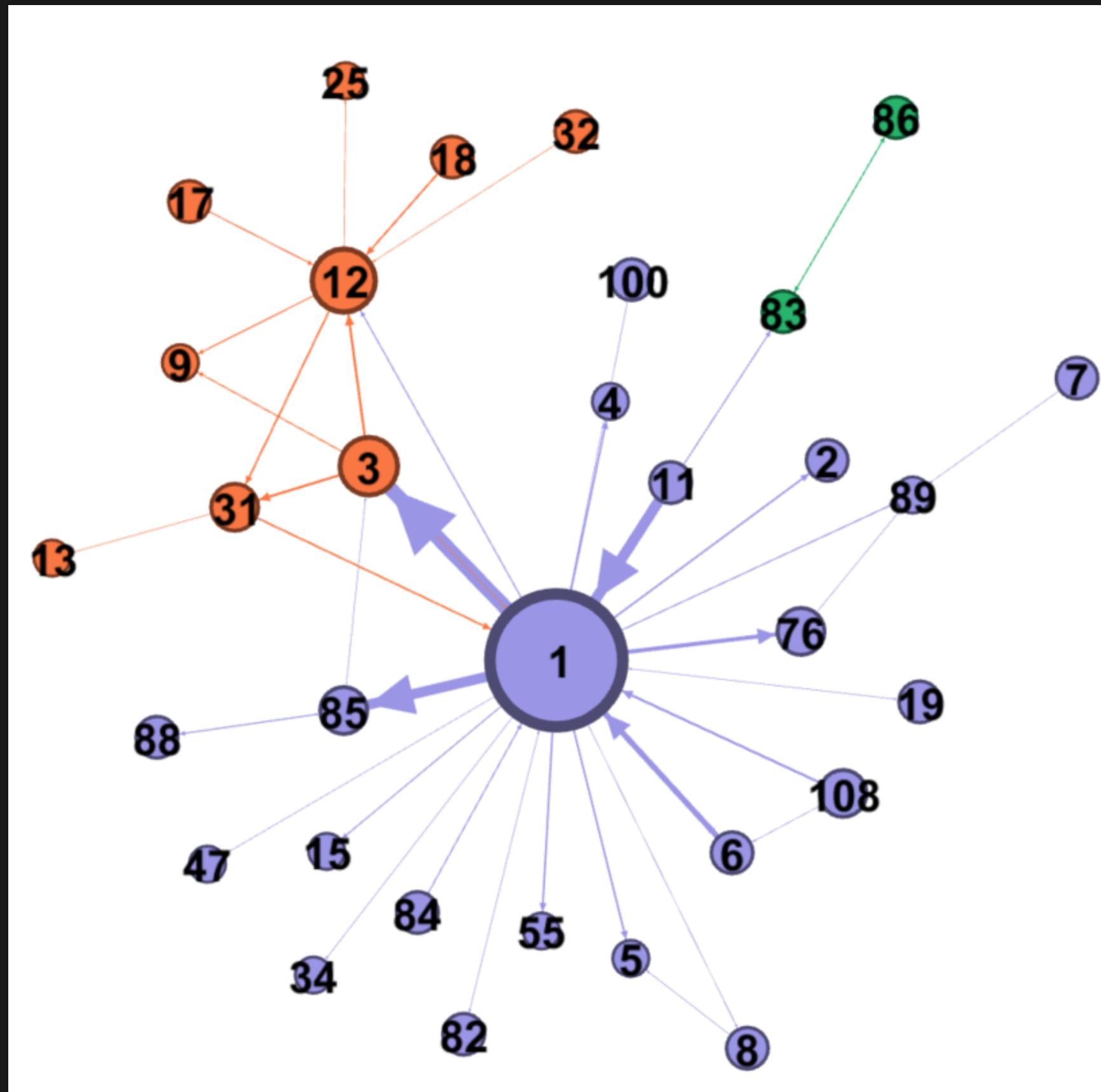
financial investors; accountants;
owners of various importation
businesses, etc.

PHASE 4

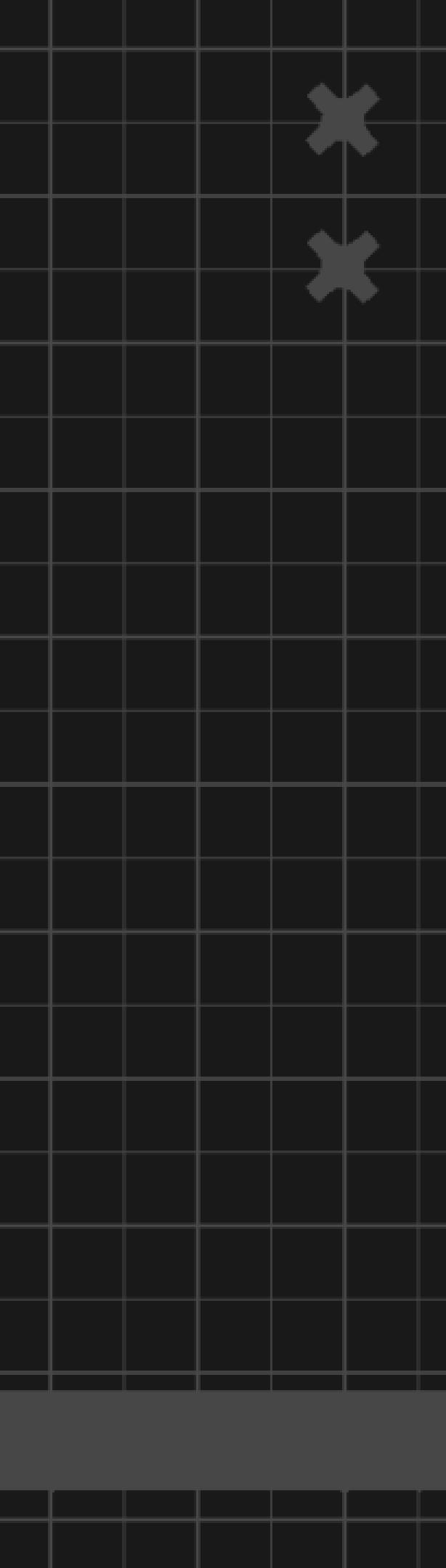


-  **Daniel Serero Gang (N1)**
Marijuana dealer (Core)
 -  **Daniel Serero Gang**
Investors and transporters of money
 -  **Daniel Serero Gang**
Finance and Accountant
 -  **Import Cocaine from Ernesto Morales (N12)**
Operated by N3
 -  **Investors and transporters of money**

PHASE 5



- **Daniel Serero Gang (N1)**
Marijuana dealer (Core)
- **Daniel Serero Gang**
Investors and transporters of money
- **Import Cocaine from Ernesto Morales (N12)**
Operated by N3



WHY LINK PREDICTION ??

We suppose a method which tries to identify the hidden edges
such as a collaboration using undetectable means to carry
covert activities

LINK PREDICTION

Using networkx library.

Dataset

Combine all 11 phases into 1 large network.

Link Prediction method

Randomly remove 20% of the edges then try to predict back using:

- Jaccard Coefficient
- Adamic Adar
- Preferential Attachment

Evaluation

ROC AUC

ALGORITHM

Jaccard Coefficient

เป็นวิธีการหนึ่งในการหาค่า similarity หรือเป็นการหาว่า set 2 set มีความคล้ายคลึงขนาดไหน โดยจะดูจากค่า intersection ส่วนด้วย union

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Adamic Adar

เป็นค่าที่ใช้ในการหาความเชื่อมโยงใน social network โดยวัดจากการเชื่อมโยงกันระหว่าง node 2 node ซึ่งจะมีวิธี การคือหาผลรวมของ 1 ส่วนด้วย logarithmic ขนาดการ เชื่อมโยง node neighbourhoods

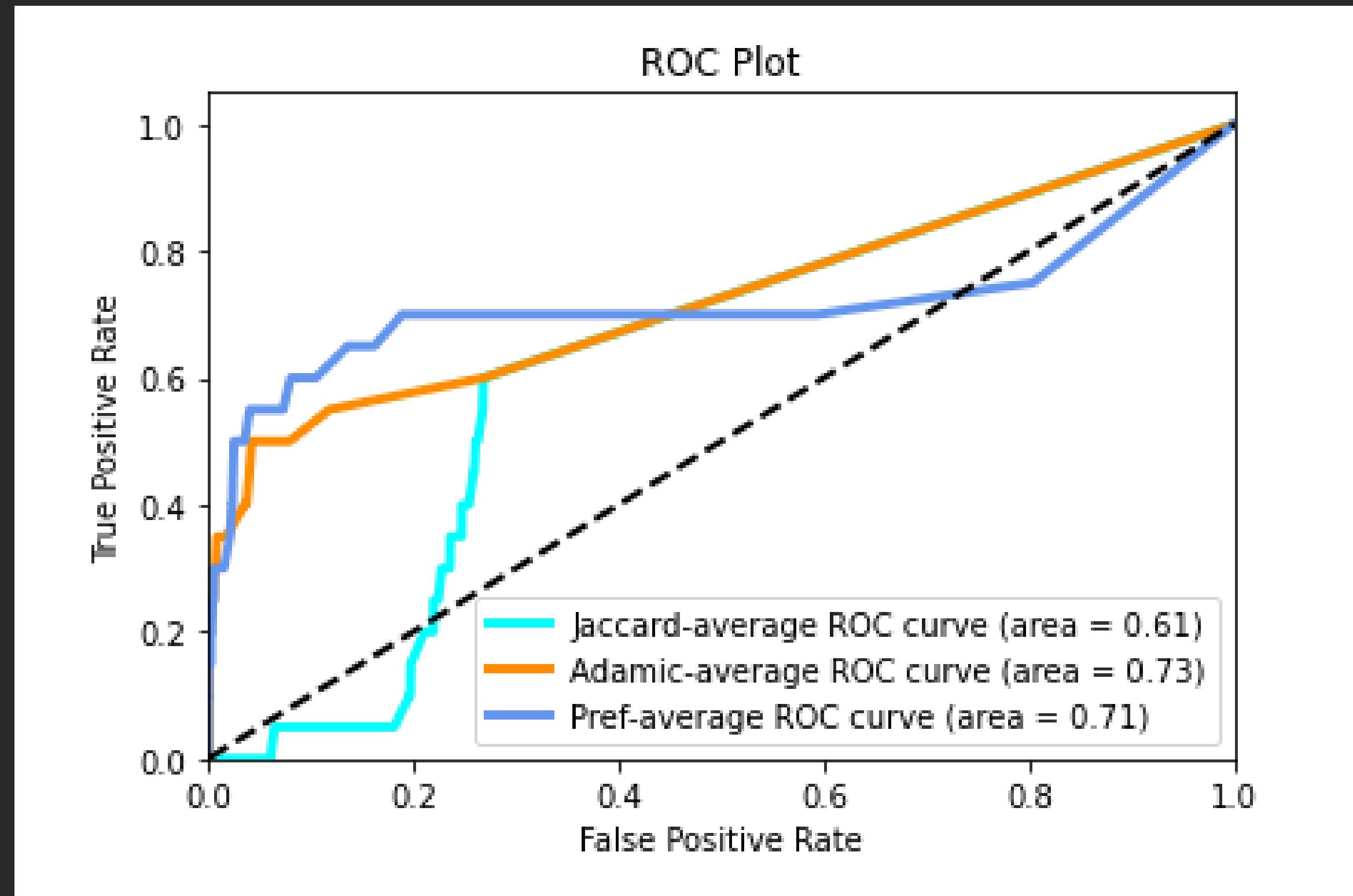
$$A(x, y) = \sum_{u \in N(x) \cap N(y)} \frac{1}{\log |N(u)|}$$

Preferential Attachment

เป็น Algorithm ที่ใช้ในการสร้าง link โดยจะอาศัยจากการ ดูความสัมพันธ์ระหว่าง degree กับ node

$$d_x^* \leq \lceil \delta d_z \rceil$$

EVALUATE



LINK PREDICTION

OUTPUT: ADAMIC ADAR

```
[('n76', 'n1', 3.3994269827266423),  
 ('n41', 'n37', 2.9819339427569274),  
 ('n87', 'n12', 2.600232160088021),  
 ('n58', 'n12', 2.352934267515801),  
 ('n8', 'n2', 2.2059750296877154),
```



Thank you

Group 7

Voramate Plodprong
6310422028

Suphanat Thaiprasit
6310422034

Rakchanok Thongkumpan
6310422039

Varattaya Rojanarachneekorn
6310422044

Piyaboon Kunakornjittirak
6310422047