Olamide OLADEJI | 6.439: Statistics Computation and Applications | Fall 2017

Question 1. Collaboration: None

1a.

If A is the adjacency matrix of the directed citation network, then $A_{ij} \in [0, 1]$ indicating if a paper i cites paper j.i.e. we define A_{ij} to be 1 if there is an edge that goes from i to j.

For the co-citation matrix, if a third paper k cites both paper i and j,

then $C_{ij} > 0 \; if \; A_{ki} = 1 \; and \; A_{kj} = 1$ for some k.

For a given pair of i and j, the total co-citation is the number of nodes k, that have $A_{ki} = 1 \; and \; A_{kj} = 1$

Thus $C_{ij} = \sum_{k=1}^{n} A_{ki} A_{kj} = \sum_{k=1}^{n} A'_{ik} \, A_{kj}$

$$C = A'A$$

NB: Note that if we define our convention for adjacency matrix entries as $A_{ij} = 1$ if there is an edge from j to i, then it can easily be seen that C would be equal to AA' instead.

We can also by convention set the diagonal of C to be zero. Since the matrix computation may yield diagonal elements greater than zero and in reality we want that the co-citation of a paper to itself to be 0.

1b.

For the bibliographic coupling matrix B, if both paper i and j cite other papers k = 1,2,3…

then $B_{ij} > 0 \; if \; A_{ik} = 1 \; and \; A_{jk} = 1 \; for \; some \; k$

Thus $B_{ij} = \sum_{k=1}^{n} A_{ik} A_{jk} = \sum_{k=1}^{n} A_{ik} \, A_{kj}'$

$$B = AA'$$

NB: Note that if we define our convention for adjacency matrix entries as $A_{ij} = 1$ if there is an edge from j to i, then it can easily be seen that B would be equal to A'A instead.

1c.

Both similarity measures can lead to different results partly because they depend strongly and varyingly on the number of incoming and outgoing edges. For two papers to have a strong co-citation, then they must have a lot of citations (or incoming edges) each. For bibliographic coupling this is not necessary; rather, for high bibliographic coupling between two papers, one would expect both papers to both have large bibliographies in the first place i.e. outgoing edges from the nodes.

While the bibliographic coupling has some advantages in measuring similarity such as :

The fact that that many papers have little to none citations but almost all papers have bibliographies (i.e. on average bibliography size > no. of citation) means that bibliographic coupling may be provide more useful similarity insights in comparing more papers (which are largely uncited). It is also quick to compute immediately after publication.

I would still argue the co-citation network a better indicator of similarity between papers in the long run, for the following reasons.

i. The co-citation matrix shows if authors reading one paper are probably also reading and citing another paper for their research, which would mean the papers are probably in the same field/similar. The bibliographic coupling matrix may not necessarily imply this as bibiliographies of papers often

draw on papers from a wide range of fields. It is possible that two papers are seminal papers of two entirely different fields but readers of a third field that draws upon both fields would read and often have both papers in their bibliography, even though both papers have different content. I imagine that this scenario would be rather common.

ii. Co-citation matrix is more dynamic than the bibliographic coupling. Unlike bibliographic coupling, the co-citation network matrix values would change with time as more and more new papers cite either or both papers which are nodes of the matrix. Thus after a long period of time, high values of matrix entries will be more definitive as a measure of similarity between to papers. For bibliographic coupling, the matrix entries are static and depend on the bibiliography of the both papers at the time of publication.

Overall, The argument for co-citation matrix as a better measure of similarity is stronger if we are comparing well established/old papers. However for new papers, the bibliographic coupling would be a stronger measure.

Question 2: Collaboration None

2a.

i. Degree: The degree can be divided into the in-degree and the out-degree.

a. In-degree: The in-degree which is the number of incoming edges to each node was computed with MATLAB. The table below provides the in-degree for the nodes under investigation:

In- Degree over 11 phases

Node Under Investigation	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8	Phase 9	Phase 10	Phase 11
1	5	8	10	11	11	17	15	10	5	8	6
3	2	1	6	4	3	9	2	8	5	1	0
83	0	2	3	4	2	2	3	1	1	1	0
86	NaN	0	3	2	1	NaN	NaN	1	NaN	1	1
85	2	0	4	5	2	3	3	2	3	2	3
6	3	0	3	1	2	0	1	0	1	NaN	NaN
11	NaN	2	1	1	0	0	1	1	1	NaN	1
88	4	2	0	2	1	NaN	2	NaN	0	NaN	1
106	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
89	2	2	1	2	3	NaN	NaN	NaN	1	NaN	NaN
84	NaN	NaN	3	1	0	0	NaN	1	NaN	1	0
5	1	1	2	1	2	1	1	NaN	NaN	NaN	NaN
8	1	2	2	2	1	4	1	3	2	1	NaN
76	NaN	2	1	2	1	6	3	2	3	3	3
77	NaN	NaN	NaN	NaN	NaN	2	2	1	NaN	NaN	NaN
87	NaN	NaN	NaN	NaN	NaN	0	1	4	5	3	3
82	NaN	NaN	NaN	NaN	1	1	NaN	2	4	5	3
96	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3	2	1
12	NaN	0	2	1	5	7	5	8	5	3	9
17	NaN	NaN	NaN	NaN	0	1	0	0	1	1	2
80	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0	NaN	NaN	NaN
33	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0	NaN	NaN	NaN
16	NaN	NaN	NaN	NaN	NaN	NaN	0	0	1	1	1

b. Out-degree: As with the in-degree, the out-degree was also computed for all the nodes under investigation and is presented below:

Out-Degree for 11 phases.

Node Under Investigation	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8	Phase 9	Phase 10	Phase 11
1	11	17	26	21	18	13	18	19	9	13	6
3	2	2	6	6	4	9	9	9	11	1	1
83	2	0	7	6	1	1	1	0	0	2	1
86	NaN	1	4	3	1	NaN	NaN	1	NaN	0	0
85	3	2	0	4	2	3	2	3	2	2	2
6	1	1	2	1	1	1	1	1	0	NaN	NaN
11	NaN	0	2	1	1	2	2	2	0	NaN	1
88	4	2	2	1	1	NaN	0	NaN	1	NaN	0
106	NaN	NaN	NaN	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
89	2	2	1	6	0	NaN	NaN	NaN	0	NaN	NaN
84	NaN	NaN	2	2	1	2	NaN	0	NaN	1	2
5	0	0	0	1	0	3	0	NaN	NaN	NaN	NaN
8	0	2	0	2	1	1	0	0	2	2	NaN
76	NaN	0	2	1	2	4	5	5	3	1	7
77	NaN	NaN	NaN	NaN	NaN	2	2	0	NaN	NaN	NaN
87	NaN	NaN	NaN	NaN	NaN	2	3	9	7	11	4
82	NaN	NaN	NaN	NaN	1	1	NaN	0	5	3	4
96	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2	1	2
12	NaN	1	0	1	5	10	1	4	7	6	9
17	NaN	NaN	NaN	NaN	1	1	1	1	1	0	1
80	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1	NaN	NaN	NaN
33	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1	NaN	NaN	NaN
16	NaN	NaN	NaN	NaN	NaN	NaN	1	1	1	1	1

Overall Degree: We can combine both the in-degree and out-degree to obtain the degree. The table below shows the degree of the nodes under investigation over the 11 phases. The degree was computed by summing the in-degree + out-degree for each node at a particular phase. Note that some nodes have “NaN” values for particular phases. This shows that they were not active during these phases.

Degree across all phases.

Node Under Investigation	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8	Phase 9	Phase 10	Phase 11
1	16	25	36	32	29	30	33	29	14	21	12
3	4	3	12	10	7	18	11	17	16	2	1
83	2	2	10	10	3	3	4	1	1	3	1
86	NaN	1	7	5	2	NaN	NaN	2	NaN	1	1
85	5	2	4	9	4	6	5	5	5	4	5
6	4	1	5	2	3	1	2	1	1	NaN	NaN
11	NaN	2	3	2	1	2	3	3	1	NaN	2
88	8	4	2	3	2	NaN	2	NaN	1	NaN	1
106	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
89	4	4	2	8	3	NaN	NaN	NaN	1	NaN	NaN
84	NaN	NaN	5	3	1	2	NaN	1	NaN	2	2
5	1	1	2	2	2	4	1	NaN	NaN	NaN	NaN
8	1	4	2	4	2	5	1	3	4	3	NaN
76	NaN	2	3	3	3	10	8	7	6	4	10
77	NaN	NaN	NaN	NaN	NaN	4	4	1	NaN	NaN	NaN
87	NaN	NaN	NaN	NaN	NaN	2	4	13	12	14	7
82	NaN	NaN	NaN	NaN	2	2	NaN	2	9	8	7
96	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	5	3	3
12	NaN	1	2	2	10	17	6	12	12	9	18
17	NaN	NaN	NaN	NaN	1	2	1	1	2	1	3
80	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1	NaN	NaN	NaN
33	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1	NaN	NaN	NaN
16	NaN	NaN	NaN	NaN	NaN	NaN	1	1	2	2	2

ii. Betweenness Centrality.

The betweenness centrality was also computed using the MATLAB and noting the weights of edges. Below is a table for betweenness centrality values for the investigated nodes across all 11 phases.

Node ID	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8	Phase 9	Phase 10	Phase 11
1	71	175	315.166667	413.5	365	257.166667	368	425.5	249	370	151.666667
3	0	0	192.166667	126	120	175.833333	62	498	377.5	0	0
83	0	0	16	50.5	18	32	2	0	0	17	0
86	NaN	0	22	28	0	NaN	NaN	25	NaN	0	0
85	0	0	0	91.5	39	34	38	0	31	26	8
6	0	0	28	0	0	0	0	0	0	NaN	NaN
11	NaN	0	24	0	0	0	17	24	0	NaN	0
88	33	27	0	0	0	NaN	0	NaN	0	NaN	0
106	NaN	NaN	NaN	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
89	18	12	0	78	0	NaN	NaN	NaN	0	NaN	NaN
84	NaN	NaN	0	34	0	0	NaN	0	NaN	0	0
5	0	0	0	0	0	0	0	NaN	NaN	NaN	NaN
8	0	17	0	34	18	5	0	0	151	17	NaN
76	NaN	0	6	0	0	69.5	153	103	100	105	52.3333333
77	NaN	NaN	NaN	NaN	NaN	0	0	0	NaN	NaN	NaN
87	NaN	NaN	NaN	NaN	NaN	0	57	157.5	164	121	70
82	NaN	NaN	NaN	NaN	0	0	NaN	0	122	81	68.3333333
96	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	24	0	7
12	NaN	0	0	0	121	203.833333	4	280	197	25	211.333333
17	NaN	NaN	NaN	NaN	0	0	0	0	0	0	0
80	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0	NaN	NaN	NaN
33	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0	NaN	NaN	NaN
16	NaN	NaN	NaN	NaN	NaN	NaN	0	0	0	0	0

Again, recall NaN in the table denotes that the criminal was not active during the phase.

iii. Eigenvector Centrality.

Since the eigenvector centrality is typically computed on undirected graphs and this problem set involves around directed (communication) graph between criminals, I computed another measure of eigenvector centrality which works on directed graph called the ‘Pagerank’ centrality.

This is presented in the following table for the given criminals.

Node Under Investigation	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8	Phase 9	Phase 10	Phase 11
1	0.11790356	0.18417983	0.14029226	0.15094514	0.14029314	0.3168903	0.16826897	0.10883709	0.05915765	0.0843707	0.04012738
3	0.0587737	0.07174754	0.0634872	0.07314183	0.08028219	0.1071001	0.0398057	0.1151795	0.07692698	0.00995986	0.00688335
83	0.01982038	0.0487504	0.0574119	0.06511357	0.07533341	0.0308918	0.06885762	0.00975917	0.01230644	0.01087959	0.00688335
86	NaN	0.01899172	0.06084093	0.04346393	0.07377217	NaN	NaN	0.01158453	NaN	0.01460025	0.012735
85	0.12631053	0.01899172	0.05373773	0.1062826	0.08650916	0.05038297	0.01654059	0.02733483	0.03746574	0.01962234	0.02556785
6	0.12608127	0.01899172	0.02394825	0.01586233	0.01585256	0.00635627	0.01325796	0.00861759	0.01094745	NaN	NaN
11	NaN	0.06136429	0.01637125	0.0118545	0.0097267	0.00635627	0.00753676	0.01261312	0.01094745	NaN	0.00843395
88	0.17738989	0.02876566	0.01433952	0.04396736	0.058757	NaN	0.00928219	NaN	0.01026796	NaN	0.0079592
106	NaN	NaN	NaN	0.01480373	NaN	NaN	NaN	NaN	NaN	NaN	NaN
89	0.11030818	0.07092014	0.01870833	0.03071089	0.02668929	NaN	NaN	NaN	0.01157632	NaN	NaN
84	NaN	NaN	0.02599471	0.01430275	0.0097267	0.00635627	NaN	0.00918838	NaN	0.0081204	0.00688335
5	0.0267336	0.02750072	0.03090362	0.01836722	0.02373431	0.03387441	0.01039736	NaN	NaN	NaN	NaN
8	0.02327699	0.06140765	0.04480511	0.01758216	0.01121745	0.04342267	0.00753676	0.03014194	0.02533938	0.01363878	NaN
76	NaN	0.03387106	0.0174601	0.02991484	0.02016193	0.08525308	0.08915983	0.01651568	0.03494666	0.02267505	0.03163671
77	NaN	NaN	NaN	NaN	NaN	0.01123598	0.03544011	0.01204233	NaN	NaN	NaN
87	NaN	NaN	NaN	NaN	NaN	0.00635627	0.00896706	0.07699226	0.06553449	0.04772709	0.01885346
82	NaN	NaN	NaN	NaN	0.01121745	0.00832185	NaN	0.01615085	0.0931089	0.09498372	0.02633498
96	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.05436537	0.05923519	0.01549437
12	NaN	0.01899172	0.02233214	0.01479318	0.068519	0.09784543	0.11547683	0.08099753	0.0752173	0.04924435	0.11956703
17	NaN	NaN	NaN	NaN	0.0097267	0.02578117	0.00610646	0.00861759	0.01737428	0.00763647	0.01509372
80	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.00861759	NaN	NaN	NaN
33	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.00861759	NaN	NaN	NaN
16	NaN	NaN	NaN	NaN	NaN	NaN	0.00610646	0.00861759	0.02625718	0.02201781	0.01726209

2b.

The central actors are:

1. Serrero Daniel (n1): He is active all through the 11 phases and has relatively high centrality measures across all phases and when computed under betweeness and eigenvector centrality.
2. Pierre Perlini (n3): He is also similarly active all through 11 phases. Between phase 3 and phase 9, he records very high betweeness and eigenvector centrality measures.
3. Ernesto Morales (n12): While he is not active in the first phase and has 0 value for betweeness centrality in the 2nd to 4th phase, Ernesto betweeness centrality values become quite large as the phases progresses. His eigenvector centrality are also relatively high when he becomes active and until the 11th phase.
4. Gabrielle Casalle(n76): He or She has large betweeness values from Phase 6 to Phase 11.
5. N87: comes active from the 6th Phase. Has highest betweeness in the 9th phase and remains one of the most central players till the 11th phase.
6. N82: The transport manager had high centrality values as phases increased. This was probably due to the increase in seizures and the transport manager's need to provide feedback on seizure and restrategize transportation after every seizure.

Of the above, the three principal actors are n1, n3 and n12.

Peripheral actors: Using the centrality measures computed as well as a visualization of the network over the 11 phases (presented subsequently), we can identify some peripheral actors as the nodes below. These nodes had very low centrality measures consistently across the 11 networks as well as being inactive for many phases as can be seen in the tables and graphs.

1. N80
2. N33
3. N77
4. N16

2c.

Actors Playing Important Roles But not Investigated.

To do this I computed a matrix of centrality (for all nodes) over all phases and tried to identify nodes which are not under-investigation but have high centrality values in the different phases.

Using degree as one criterion, we can examine the nodes not under investigation which have relatively high degrees. I use a threshold degree of > 5 to filter out these "relatively important nodes" by degree measure that we may have missed.

A table of these nodes for each of the eleven phases is shown below:

Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8	Phase 9	Phase 10	Phase 11
None	None	9	None	None	None	19	2	None	14	41
		49					14		37	79

For the betweenness centrality, I examined the other nodes which were not under investigation but with betweeness centrality values greater than 0. A table showing this for all 11 phases shown below:

Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8	Phase 9	Phase 10	Phase 11
None	None	9	13	31	14	19	2	7	14	14
		13	31		15	20	14	46	24	27
		49	52		19	74	22	78	37	37
			90		20	78	34	79	41	41
			107			79			46	58
									71	79
									81	93
										102

Using the pagerank centrality, I filtered for pagerank centraility values greater than 0.03 that were not under investigation for all phases. I obtained the following nodes per phase as above this threshold.

Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8	Phase 9	Phase 10	Phase 11
7	10	9	9	9	18	9	14	14	14	14
		7		31		22	37	46	37	58
						78			58	59
										78
										79

Considering the above, I think the following nodes might be worth investigating,

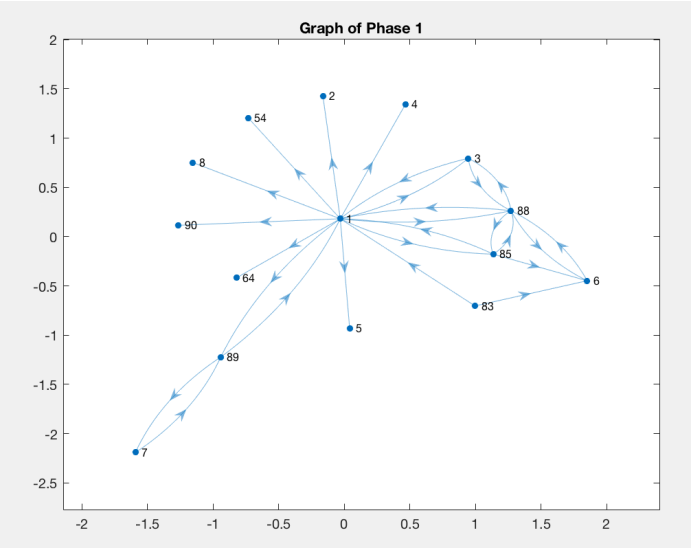
1. Node 9: This node can be seen to have relatively high pagerank centrality in many phases (phase 4 to 7).
2. Node 14. Node 14 has high centrality values for pagerank and betweeness and high degrees in the latter phases.
3. Node 79: From the tables above, we can see that this node has some relatively high centrality and degree values in some phases.

In general, if we had resources, all the nodes in the above tables could be worth investigating.

2d. Network Evolution: Coarse Patterns

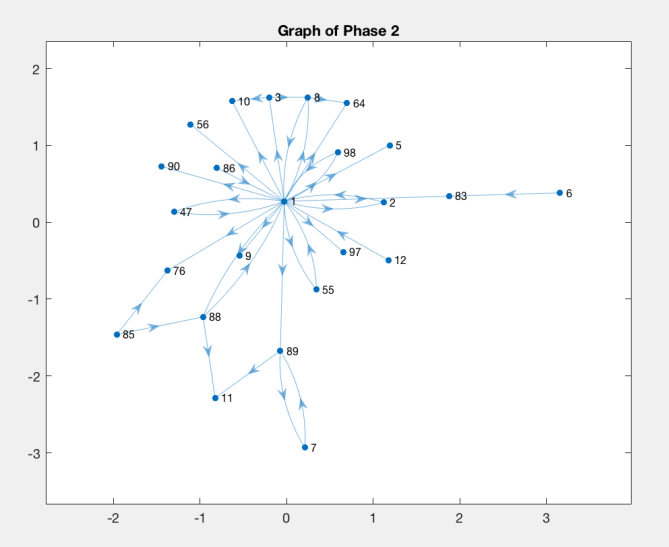
To understand the coarse patterns of the network as it evolves, I plotted the network for all 11 phases.

Phase 1.



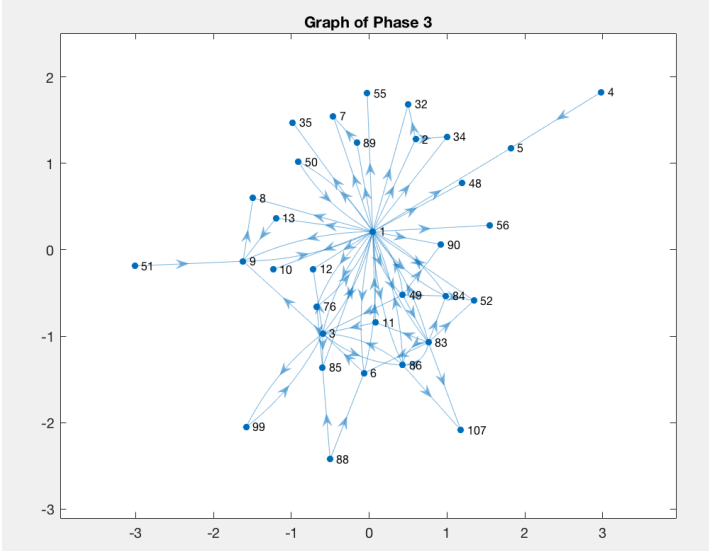
In phase 1, we see a radial graph with node 1 at the center and most of the communication from node is seen to be outwards.

Phase 2



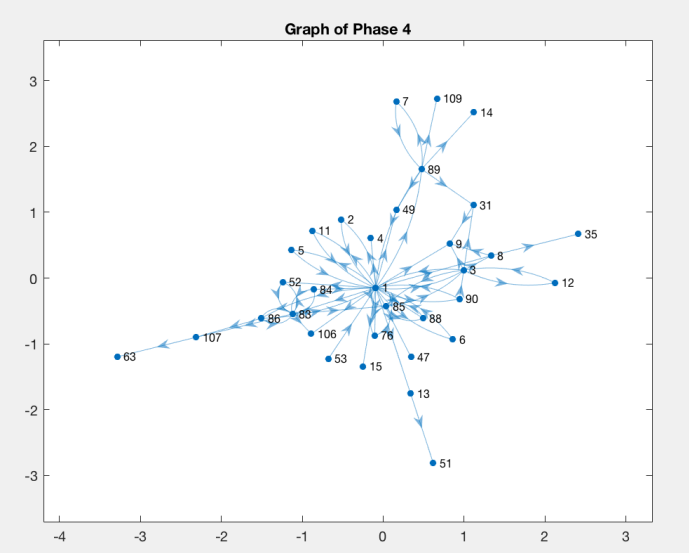
In phase two, node n1 still maintains a high centrality in the network but there is more bidirectional communication between n1 and other nodes and nodes like n89, n83 and n88 emerge as link nodes to link n1 to peripheral nodes at the outskirts of the network.

Phase 3



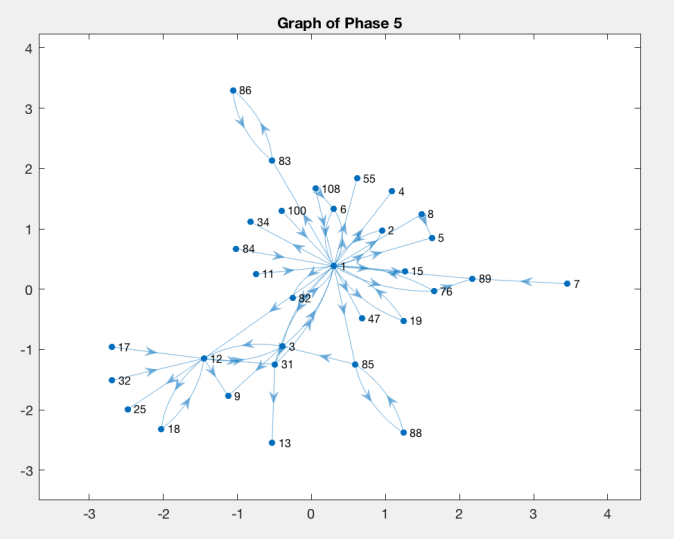
In phase 3, more nodes become active with many edges radiating from node 1 which is still by far the most central node. There are a lot of intermediary nodes now such as node 3, n6 and n83, n85, n86 which connect peripheral nodes to node 1.

Phase 4



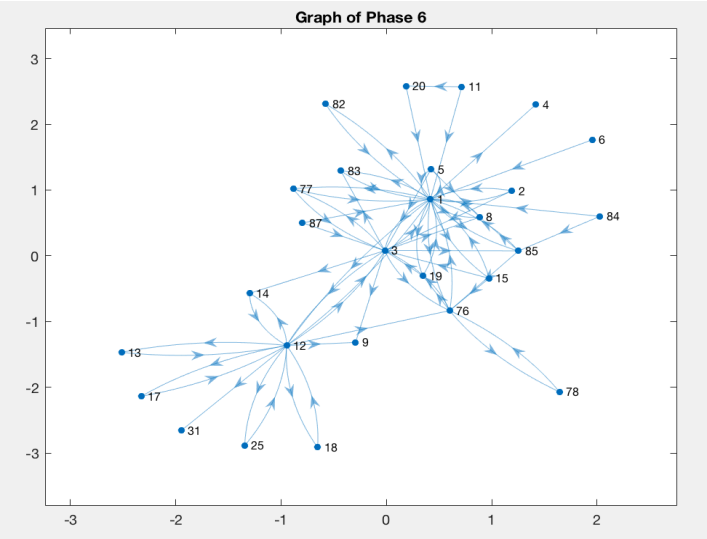
In phase 4, other highly central nodes start to appear as can be seen above. Node 1 still remains the most central with many degrees, but node 89 and node 3 also exhibit central star topologies in their parts of the network. There is a lot more activity in the network. This makes sense as this was when the first seizure occurred.

Phase 5



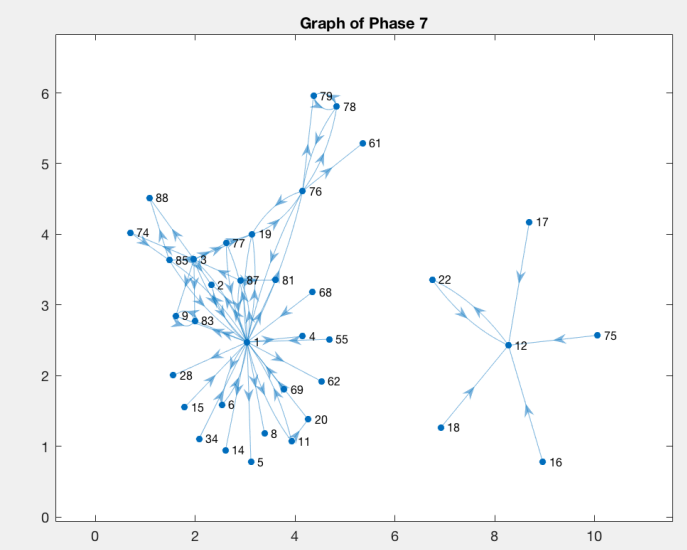
In phase 5, node 12 appears and there appears to be multiple communities in the network. A larger community in which node 1 is central and connected to many nodes, and subgraphs with node 12, node 31 and node 85 having high centralities.

Phase 6.



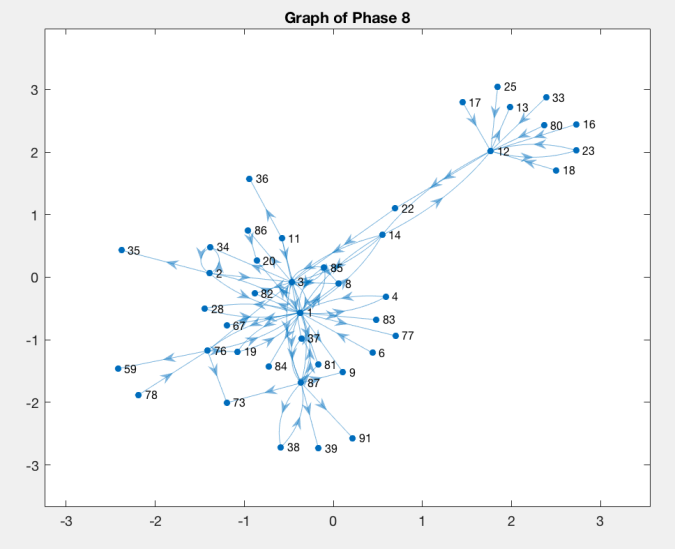
In phase 6, there appears to be three communities/clusters within the network. A large, dense cluster with node 1 as highly central, and two smaller linked communities with node 12 as central and then with node 3 as most central nodes.

Phase 7



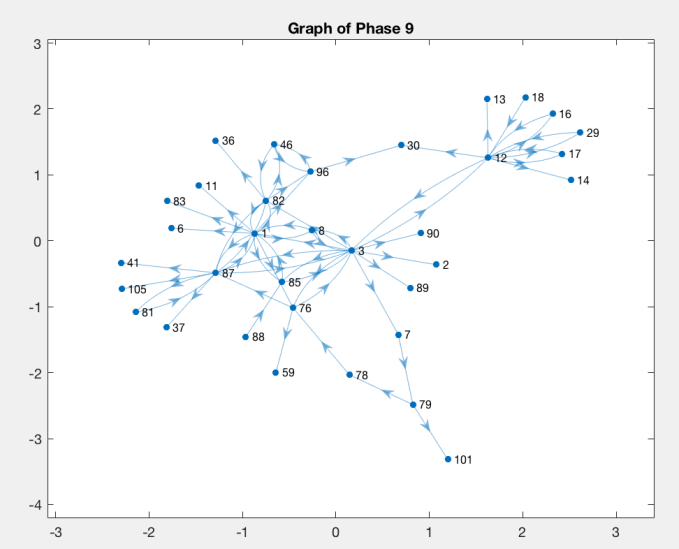
The are two disjoint subgraphs seen. The first is a larger network with node one as the most central and nodes 3 and 76 having high degrees as they provide paths from node 1 to other peripheral nodes. The second is a smaller network with node 12 at the center largely receiving information.

Phase 8



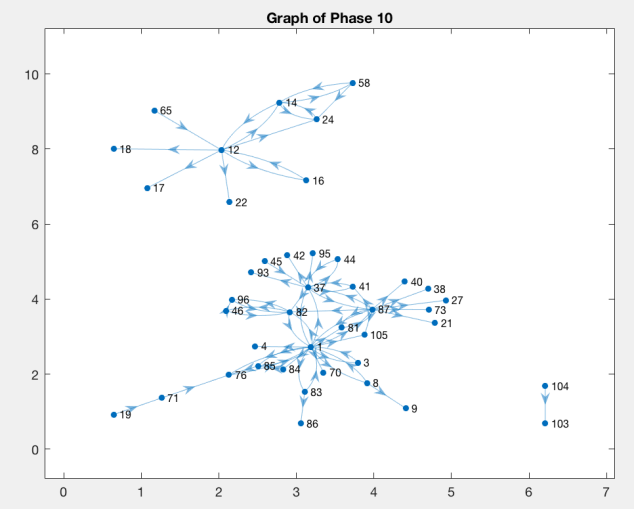
In phase 8, there is now a path from node 12 to 1, although two communities/clusters centered around node 12 and node 1 can still be seen. Node 87 and node 86, among others, continue to remain active in linking peripheral nodes to node 1.

Phase 9



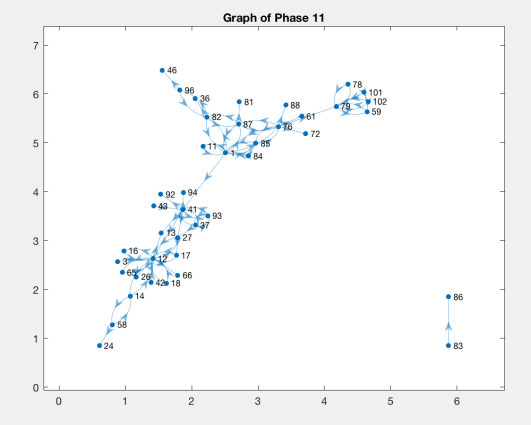
In phase 9, in addition to node 1 and 12 which are central in two parts of the network, we can see that node 3, 87, 76 and 82 gain more centrality as they link many nodes with node 1. The neighborhood of node 1 appears densely connected.

Phase 10



Here, there are three disjoint subnetworks. In the first subnetwork, node 12 is the most central with node 14 having a high degree and linking node 12 with a couple of nodes. In the second larger more densely connected subnetwork, we can see that nodes 82, 37, 87 and 1 are highly central and link with other peripheral nodes. There is a final unidirectional communication between node 104 and 103 which are both isolated from other nodes. Since the heaviest seizure occurred in this phase, we can explain this split in network as n12 continued coordinating the cocaine shipment in isolation with some parties while actors involved in the marijuana, from investors to transporters, investors, recuperators and coordinators, were concerned and discussing the heavy marijuana seizure.

Phase 11.



In the final phase, there appears to be about 3 clusters in the network. A large cluster can be seen in the network with node 1, 87, 82 and 76 as highly central and this cluster has an outgoing edge to another cluster with node 79 as its center and an outgoing edge to a cluster with node 41, 12 as highly central.
Finally, two nodes 83 and 86, which are isolated from the rest have unidirectional communication between them (from 83 to 86).

Correlation of Network Evolution with Background Story.

Yes there is a correlation with network pattern evolution and the background story. The following observations validate this:

1. The network evolution presented previously corroborates the background story that node 1 is the most important figure or mastermind of the network.
2. Node 6 (the Spanish Broker) was a central figure in phase 3 but no longer from phase 4 corroborating the story that there was a switch from a Morroco-Spain route as transition to Colombia-USA route.

3. Node 12, Ernesto, becomes only active from phase 4 as we see in the network and becomes highly important/central until end of phase 11. This corroborates the fact that there was a switch to the Colombia cocaine import and Ernesto's storyline as the principal organizer of this part.
4. The seizures also correlate well with activities of different actors. For instance, node 12 is active and has high centrality especially cocaine seizures occur in phases 8, 9 and 11. Marijuana recuperators also increased communication frequency and number of directions particularly during marijuana seizures.
5. At all phases node 3 is present. Also node 3 can be seen to be quite active (having high degrees) and in communication with node 1 in many phases, further confirming his storyline as the right hand man of node 1.
6. Node 83 and node 86 communicate frequently as seen in phase 10 and 11 corroborating the story that they are both into investing and transporting the money.
7. Node 82 is also quite central and active especially at the latter phases, justifying his storyline as the transports arrangement manager.

2e.

Recall that the central actors are:

1. Serrero Daniel (n1): Mastermind of network.
2. Pierre Perlini (n3): Principal Lieutenant of n1.
3. Ernesto Morales (n12): Cocaine Import Organizer.
4. Gabrielle Casalle(n76): In charge of recuperating marijuana.
5. N87 Patrick Lee. Investor
6. N82: Panetta. Transports arrangements manager.

Nodees n1, n3 and n12 are the top three central actors.

Activity: As we see in the previous tables provided for centrality measures across phases, Serrero(n1) and Pierre(n3) are active across all 11 phases. The indices that reflect their description would thus be 1,2,…11.

Ernesto was inactive in the first phase. In the second to third phase, Ernesto (n12) was active but with very low centrality. As soon as the transition to Colombia occurred, Ernesto(n12) had high centrality (from phase 4 upwards) till the end of the investigation. Thus the network evolution indices fitting n12's description are 4 to 11.

Patrick Lee (n87) the investor had high betweeness centrality from phase 8 to end. He was inactive prior to phase 6. The network evolution indices fitting n87's description would be 8 - 11. As an investor, he could have been increasingly more concerned about returns and losses towards at the latter phases, when the drugs are repeatedly seized.

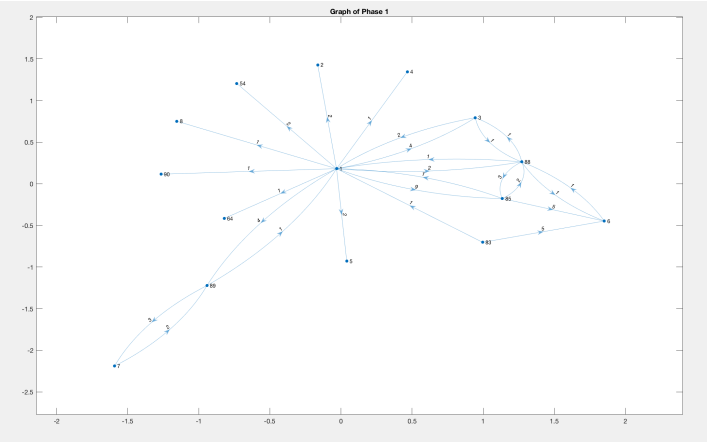
Gabrielle Casalle (n76): became active in phase 2 but became relatively more central as from phase 6 to 11. He or she was active and with more communication in the phases when there were seizures (this is observable in the network), which matches well the job description of "recuperating marijuana".

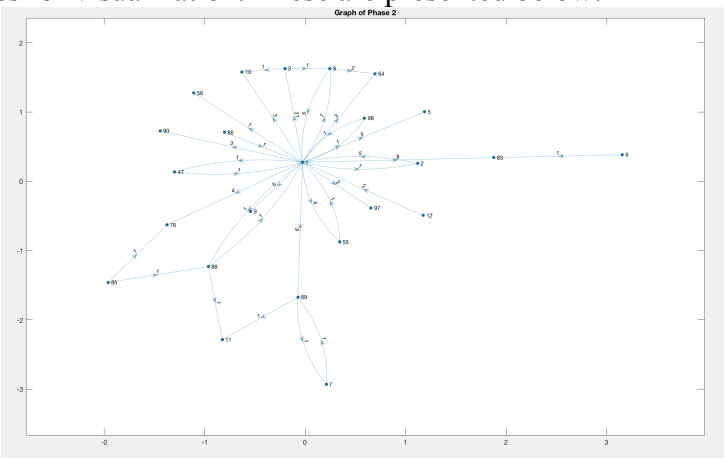
Finally, Panetta(n82) was inactive for the first four phases, and phase 7, was relatively peripheral in communication during the active phases except phases 9 to 11 when she or he became relatively more central in the network. As the transport manager, Panetta's description can be observable with network indices of 9,10,11. N82 has increased communication particularly with n1 when there are seizures in these latter phases (phase 9, 10, 11).

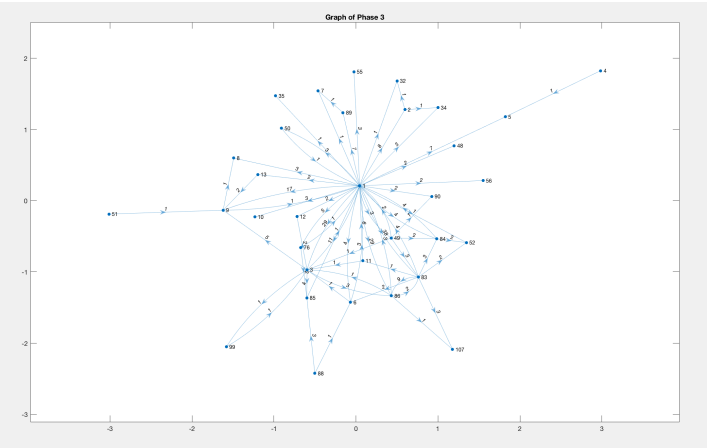
2f.

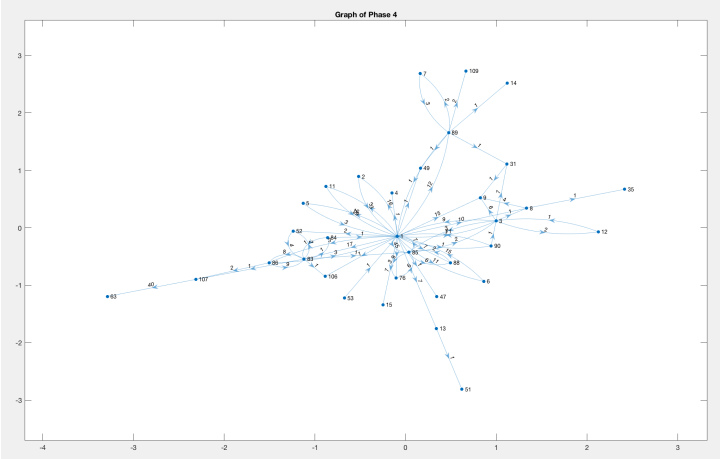
Frequency and direction of communication of Node 1.

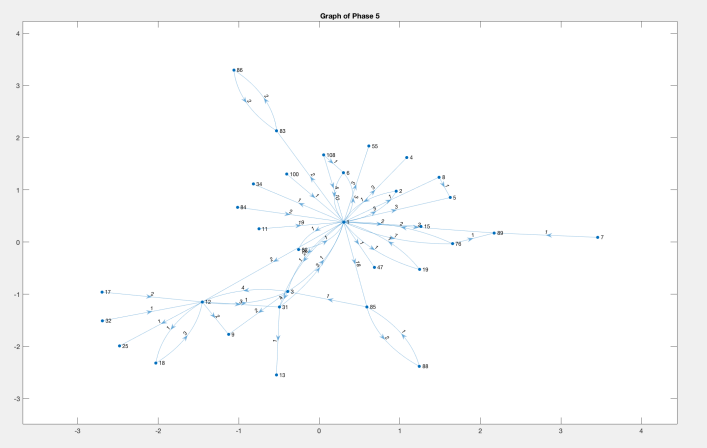
To study the frequency and direction of communications of Serrero (n1), I replotted the network evolution graphs but included the value of the weights of the edges for visualization. These are presented below:

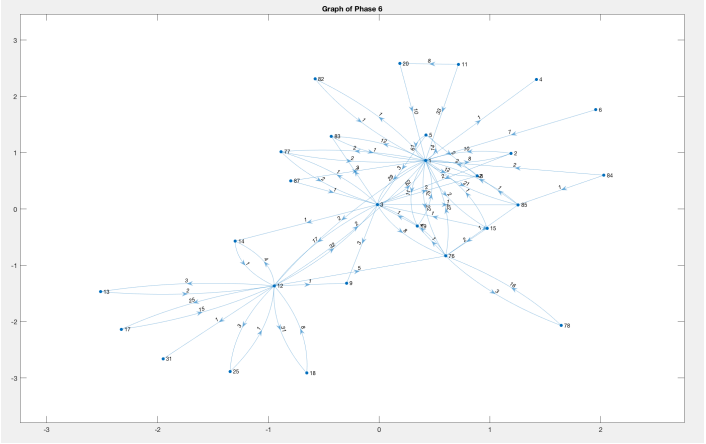


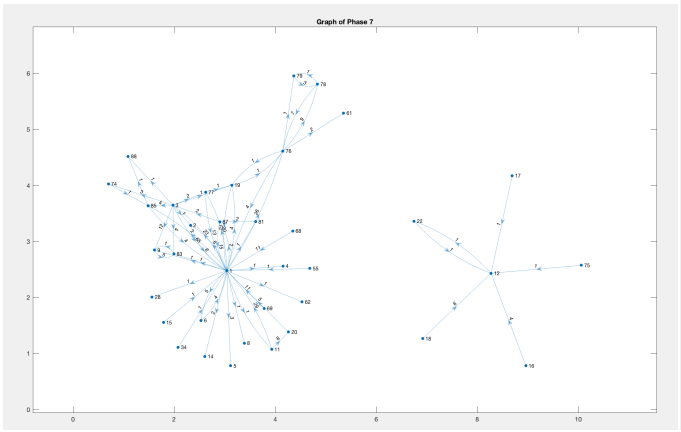


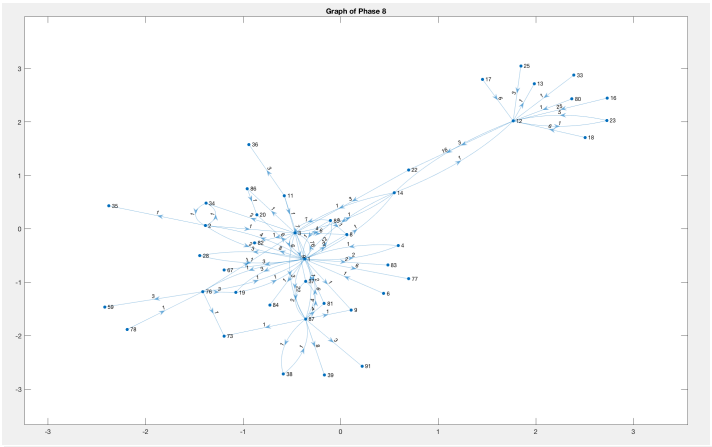


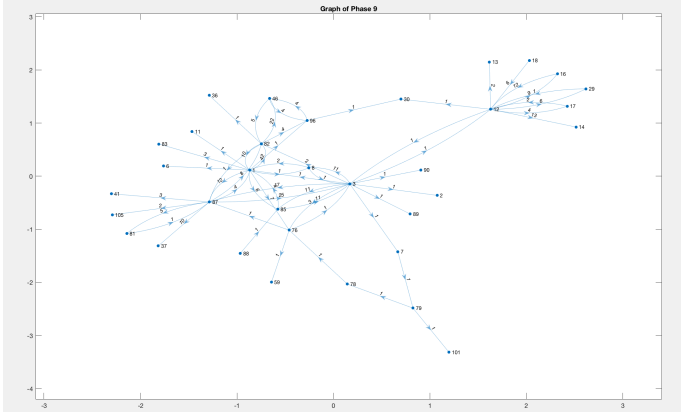


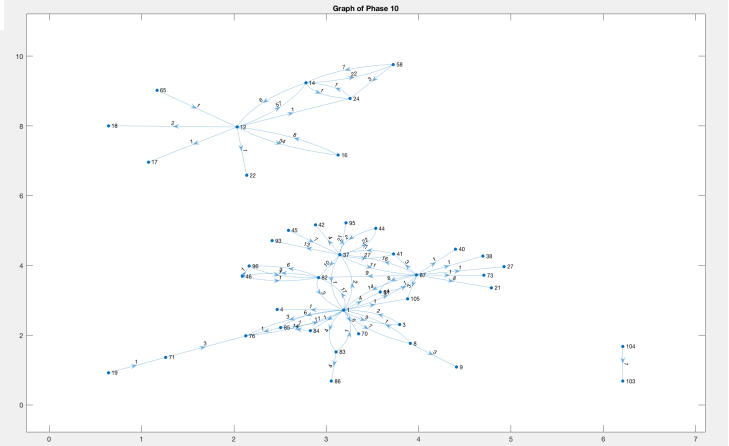


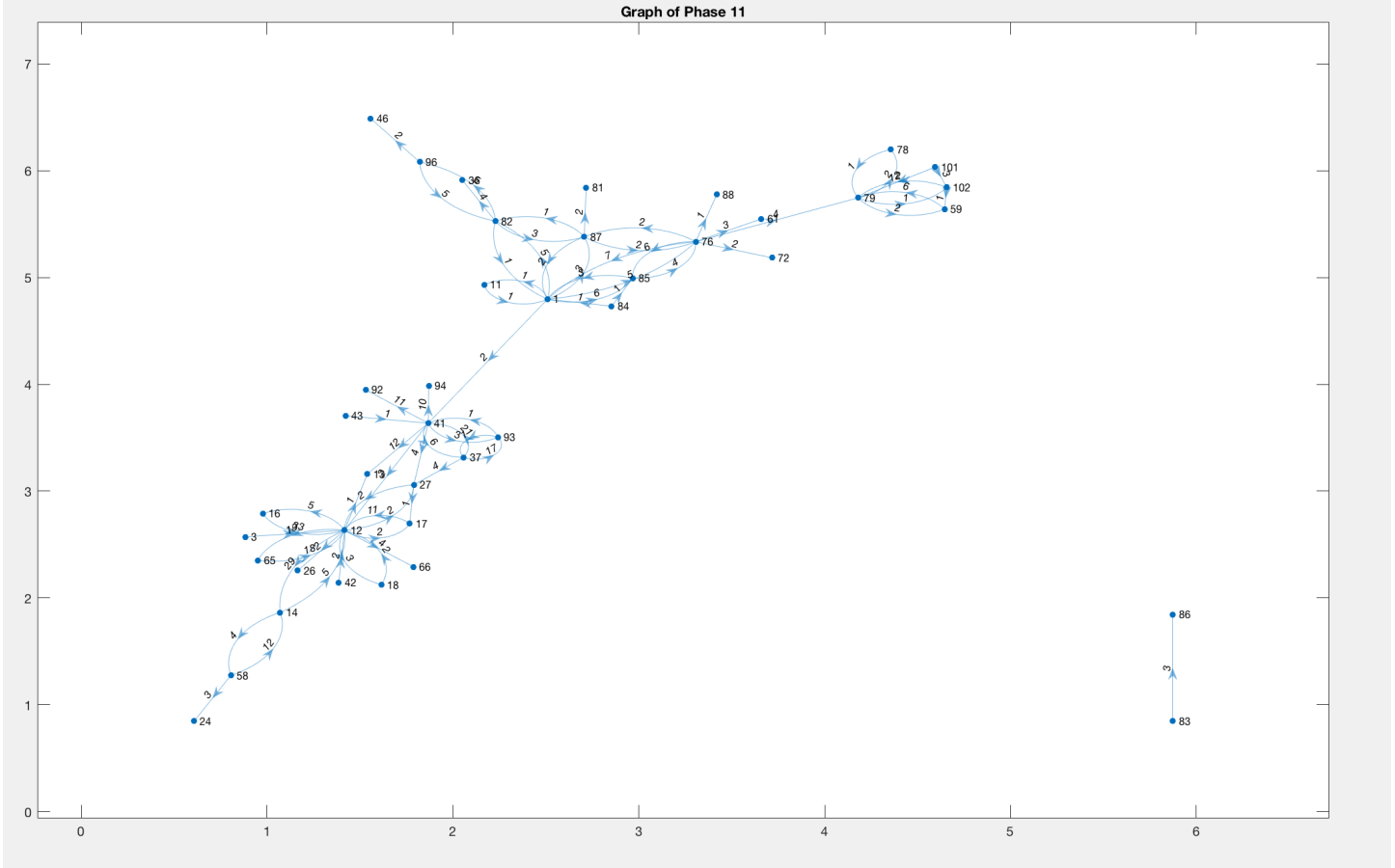



Graph of Phase 7


Graph of Phase 8


Graph of Phase 9


Graph of Phase 10


Graph of Phase 11

I summarize below, my insights from studying the direction and frequency(weights of edges) of n1 from the above graphs.

Phase	Frequency and Communication summary for (n1)
1.	Most of the communication is outward and low frequency (once), except to n3 (his right hand man), n85 & n88, financial affairs men where there is bidirectional communication. For the bidirectional case, the outward weights are more than inward. Overall, we can interprete this phase as n1 giving instructions (hence outgoing links) to the different parties as the smuggling activity is about to start.
2.	There is a lot more bidirectional communication indicating feedback from parties to n1 as he gives instructions. Frequency is low except for individuals such as n3 (his right hand man) where there is outward link with weight of 31. He is probably still setting up the whole criminal activity and instructing his right-hand man to undertake many tasks. Nodes 88, 76 and 5 (marijuana recuperators) also receive relatively more instructions than others.
3.	In terms of direction this is quite similar to previous phase as lots of communications going outward at low frequencies. However n1 communicates more frequently to node 83 (outgoing frequency is '38' times), node 86 (outgoing 29 times) and node 3 (his right hand man; 29 times). N83 and n86 are the investors and transporters of money so he's trying to ensure that the monetary transportation and investment aspects of the criminal activity are in place.
4.	He speaks to more people more frequently. N11 calls in 16 times while he communicates out to n5 29 times. Other notable newly frequent communications include 15 times out to n9 and to n89 (12 times). Also, he calls out to n83 17 times and n85 at 45 times. The inward communications are relatively less. This phase seems to be heavily focused on financial communication as n85, n83, n89 are finance and investment personnel. This makes sense because there is a seizure of $2.5million worth of marijuana this period, the first seizure.
5.	Most of the conversations here are bidirectional and infrequent (once) except for conversations with n6 and n11. He hears from n11 (the Mexican provider) 19 times, while n6 (Spanish Broker) calls in 10 times and he also calls n6 as well. Overall, as this is just after phase 4 during which there was a seizure, we can interprete this as the Mexican provider calling to inform him about the seizure and the Spanish Broker also heavily involved in the seizure conversation as well.
6.	There is heavy bidirectional communication between n1 and n3 (29 times out and 62 times in), n76(42 in, 22 out) and n5 (61 out, 14in). There is also heavy inflow from n19(75 times), n11(33 times) and n20(10 times). Communication with others are largely infrequent. The right-hand n3 communicates a lot with n12 during this period so we can interpret that n1 tries to explore and set up the Colombian-US route for cocaine. There is also still heavy communications with the the marijuana recuperators n76 and n5 as we see perhaps to discuss recuperating strategy since we know that there were three seizures of marijuana in this phase.
7.	More frequent inflow ensues from the marijuana recuperators (n76 and n5) and n68. This makes sense as there was a heavy seizure of marijuana in this phase. The number of participants increased. There is a roughly equal mix of inflow and outflow edges (in-out degrees) for n1 in this phase. Frequent bidirectional communication from and to n1 exists with n77 and n3 and communications to other nodes are relatively infrequent. N77 has airport contacts and so n1 is probably trying to explore air transportation smuggling strategy for this new route.
8.	Lots of bidirectional communication with other nodes, infrequently. Frequent communication with node 3, the right hand man, with more instructions going out than in. I believe the new operation of cocaine is in full effect here and the communications with n3 are to ensure smooth running of the business.
9.	Bidirectional frequent communication to n82 n87 and n85. Other communication is largely unidirectional and outward. The most frequent conversation is with n82 who is the transport arrangement manager. This makes sense as there were two seizures, for cocaine and marijuana,

	again during this phase and so the drug transportation discussion ensue. N87 is probably the investor who lost the most money during this phases' seizure and so that explains the frequent bidirectional communication with one.
10.	There is bidirectional communication with 6 nodes and unidirectional with 4 nodes. Nodes n 85, n82 and n87 are communicated with the most. As there is the heaviest seizure this period, we can interprete this as n87, the marijuana investor, was quite concerned hence frequent communication. It also makes sense that N82 as the transport manager would be frequently communicated with during this heavy seizure phase. It is also not surprising that n85 who is the financial person engages with frequent bidirectional communication with n1 given the heavy financial losses involved with the seizure in this phase.
11.	Bidirectional communication with n85, n82, n87, n78 and n84 ensue. There is also outgoing communication to one. The frequency of communication is relatively now in this phase compared to other phases (perhaps because he has become used to the seizures). But the actors communicated with, n82 (transport), n85(accountant) and n87(investor) suggest conversations focus still focus on the seizure and logistics of transportation as in phase 11.

2g

I think that the police's strategy of 'seize but don't arrest' definitely impacted the network over the phases. I base this conclusion on the following observations:

1. After the first seizure in phase 4, we see that the network grows in number of actors as the importation of cocaine through the Colombia-US route is begun and actors like n12 become active.
2. Node n1's influence/centrality relative to other nodes wanes as the phases increase. Other nodes such as 12 begin to gain more centrality. This could be interpreted as less dependency on a central mastermind as seizures happened and more people began to adapt and coordinate activities at different parts of the network rather than route all communication through one central node n1.
3. We can see that the network splits into disjoint subnetworks at some phases. This is likely due to the seizure activity of the police. For instance in phase 10 when the heaviest seizure occurred, we can we see a split and in one subnetwork, n12 continued coordinating the cocaine shipment in isolation with some parties while in the other subnetwork, actors involved in the marijuana, from investors to transporters, investors, recuperators and coordinators, were concerned and discussing the heavy marijuana seizure.
   Also in phase 7, n12 coordinates the cocaine smuggling with some other nodes in a separate, isolated subgraph while the n1 and some nodes are connected in the other graph. This is because the seizure of marijuana did not really affect the cocaine part of the business.
4. By the 11^{th} phase unlike the first phase where node 1 was at the center of most communication paths/communications radiated from n1 to everyone and there is at most 3 nodes separating two nodes, we see that the network has evolved to have multiple clusters/communities of facilitators and the number of nodes in the path between pairs of nodes is on average significantly higher. Some nodes such as 83 and 86 do not even communicate with other nodes. This can be interpreted to mean that the seizures over time has decentralized power and communication as actors have adapted to solving problems locally rather than escalating every issue or decision to a single central overall top node.