**Written Report – 6.419x Module 3**

**Name:** (Xiao Nan)

* **Problem 1**

***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:**

Using matrix multiplication, cocitation is C = A^T \* A (and bibliographic coupling is B = A \* A^T). With the standard algorithm, both cost O(n^3) in the worst case—same as your friend’s pair-enumeration (Σ k\_i^2 ≤ n \* n^2). In practice, the matrix form is faster due to sparsity and optimized BLAS: runtime ≈ Σ k\_i^2 (often written O(n \* d^2) for average out‑degree d). With fast multiplication the asymptotic cost is O(n^ω) with ω ≈ 2.37.

***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 (B = A \* A^T) links papers that cite the same works (outgoing edges). Cocitation (C = A^T \* A) links papers that are cited together by others (incoming edges). They can differ because coupling reflects authors’ reference lists—often including generic/foundational citations—so it may connect papers across fields; it is fixed at publication time and works well for new papers with few citations. Cocitation is community-driven and evolves over time, often highlighting established or highly cited/review papers but disadvantaging new ones. Use cocitation when you want topic similarity as recognized by the field in mature areas; use bibliographic coupling for very recent papers or when focusing on shared references.

* **Problem 2**

***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 sharp rise in nodes across the early phases is consistent with ramp‑up effects: (i) investigators progressively widen surveillance (more phones tapped, more associates captured), and (ii) the network itself expands/reshuffles after early seizures (e.g., switching from marijuana to cocaine logistics), temporarily exposing additional participants. After several phases, both the organization and enforcement settle into a steady regime, so the discovered node set plateaus. For Part (b) Question 5, this means we should avoid giving late‑appearing actors an artificial advantage. Using zeros for non‑appearance (as specified) is appropriate; additionally, it’s good practice to check robustness by also reporting appearance‑conditioned means or time‑decayed means.

***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: measures how many direct contacts an actor has. In criminal networks it proxies activity/visibility (handlers, dispatchers). Limitation: it ignores who those contacts are; a node tied to many peripheral actors can look “important” but be operationally minor; it is also sensitive to observation bias.

- Betweenness centrality: measures brokerage—how often an actor lies on shortest paths. It flags couriers/coordinators who control information or commodity flow between subgroups. Limitations: unstable to small edge changes, high in tree/star structures even for nonleaders, and more costly to compute.

- Eigenvector centrality: measures influence via connections to well‑connected others (core proximity). It highlights leadership cores, financiers, or accountants embedded among other central actors. Limitations: can over‑emphasize tightly knit cliques and under‑value hub‑to‑leaf structures; component‑specific.

Most relevant to identify who runs the operation: eigenvector centrality (to locate the leadership core), supported by betweenness to surface operational coordinators. Degree alone is the least discriminative for leadership.

***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.*

**Solution:**

Define “importance” as persistent ability to influence and coordinate across time: actors with high eigenvector and betweenness centrality in many phases, not just momentary spikes. Compute per‑phase scores, average across all 11 phases with zeros for absences, and prioritize those with (a) high mean eigenvector (embedded in the core), (b) high mean betweenness (brokers between subgroups), and (c) stability across phases. Interpreting the outputs, the mastermind profile is a node that consistently ranks near the top in both eigenvector and betweenness; lieutenant/coordinator profiles rank high in betweenness but moderately in eigenvector; peripheral suppliers/investors show low betweenness and low eigenvector even if their degree spikes briefly.

***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 two panels correspond to consecutive phases 4 -> 5. Around this time, the first major seizure occurred and the group pivoted from a marijuana route to a cocaine pipeline. The network contracts and rebalances: the number of edges drops and several peripheral ties disappear, while a new sub‑hub emerges. Centrality shows a marked rise of n12 across all measures (degree, betweenness, eigenvector), consistent with taking over import coordination. n1 remains the dominant broker, with very high betweenness before and after, but its neighborhood becomes more focused on links to the emerging sub‑hub rather than many small spokes. n3 and n83 lose relative prominence as authority shifts toward actors tied to the new logistics. Visually, the left plot shows a star with multiple secondary spokes; the right plot tightens into a hub–subhub shape, with paths funneled through n12. This structural change matches the event narrative: the seizure disrupts the original channel, the organization reassigns roles, and communications reroute through the new import manager.

***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?*

**Solution:**

At a global level the network expands quickly in the first phases, then, after the initial seizure, shrinks and reorganizes, and finally grows to a stable size with intermittent bursts. A persistent core is visible across time: n1 is consistently near the top in both betweenness and eigenvector centrality, indicating long‑run brokerage and core embedding. After the pivot, n12 rises from the periphery to a central coordinator, reflected in a jump in betweenness and eigenvector. n3 oscillates: strong authority in early phases, brief hub behavior mid‑course, then mixed roles. The aggregate trend is: discovery and mapping of the group -> shock and reconfiguration -> consolidation of a smaller but more efficient core with clearer division of labor. This temporal story is consistent with the background: pressures from enforcement cause the organization to switch commodity and route, which temporarily diffuses ties but then reconcentrates around a few coordinators.

***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.*

**Solution:**

Several actors outside the 23 suspects emerge as operationally important in the directed view. Mid‑course hubs include n19 and n20; in later phases n46 and n65 appear as strong broadcasters, and n16 becomes a leading hub while n12 dominates authority. Their prominence suggests control of transportation legs or local distribution nodes, which became critical only after the pivot—hence they were not initially targeted but are influential in the reconfigured network.

***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:**

Direction adds role semantics that undirected graphs lose. Out‑degree highlights broadcasters and dispatchers; in‑degree highlights information or resource sinks (accountants, financiers, receivers). Left‑/right‑eigenvector (or hubs/authorities) separate “who points to important nodes” from “who is pointed to by important nodes.” In our data, undirected measures always rate n1 highly, but direction shows phase‑dependent role shifts: early on, n1 acts as a strong hub pushing to authorities; after the pivot, n1 temporarily becomes a major authority (flows point to him); later he resumes hub behavior. n3, by contrast, is largely an authority early on, then briefly turns into a hub during reorganization. These distinctions are essential for inferring command vs. execution roles.

***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:**

HITS over time shows clear alternation between issuing and receiving roles. Early phases are hub‑centric around n1; principal authorities include n83, n86 and n3, consistent with a leadership core surrounded by accountants and lieutenants. After the seizure, n1’s hub score stays high while authority mass concentrates on n3 and n85—authority shifting to actors tied to the new channel. In the mid‑course pivot, the pattern flips: hubs include n19, n3, n5 and n76, while n1 becomes the dominant authority, indicating information and coordination flowing toward him. Soon after, n1 again acts as a strong hub, but authority mass disperses to logistics nodes such as n82 and later consolidates around n12 when the cocaine route stabilizes. Tracking n1 and n3 across all phases confirms this: n1 alternates hub -> authority -> hub, whereas n3 alternates authority -> hub -> authority. The final phases show new hubs (n16, n65) and a dominant authority (n12), signaling that the organization’s communications end at import control rather than diffuse financial handlers. These role transitions align with the qualitative evolution described earlier and explain why undirected centralities alone cannot capture the operational leadership shifts.

* **Problem 3 - Project**

Research question

Did police disruptions produce durable shifts in leadership and structure in the CAVIAR network? I test whether the first major seizure and subsequent pivot (around phases 4->5) led to: (i) a redistribution of leadership from the initial organizer(s) to new actors; (ii) persistent changes in the global structure (clustering, assortativity, core–periphery). Sub‑questions: How do hub/authority roles (directed) and betweenness/eigenvector (undirected) evolve across phases? Are the rank changes brief turbulence or regime shifts that persist beyond one phase?

Methodology

Data are 11 two‑month phases of the CAVIAR wiretap network. I analyze both undirected binary graphs (ties if any communication in either direction) and directed weighted graphs (call counts). For roles I compute, per phase: degree, betweenness and eigenvector centralities; and HITS hubs/authorities. I summarize leadership using top‑k rank series (k=10) and quantify stability as Kendall’s τ between consecutive phases. I conduct an event study around 4->5 and run a permutation test of phase labels for actor n12 (pre 1–4 vs post 5–11; 5,000 shuffles) to assess whether its post‑event mean betweenness exceeds the pre‑event mean beyond chance. For structure I compute per‑phase transitivity, average clustering, degree assortativity, and maximum k‑core index and size. These measures are within the module’s toolkit; they jointly capture local cohesion, degree mixing, and core–periphery.

Results

Leadership reconfigures at 4->5 and then partially stabilizes. n12’s betweenness jumps by ≈0.26 and remains high post‑event; the permutation test yields p≈0.013, indicating a statistically meaningful increase. n1 stays extremely central, but its directed role flips: strong hub early, dominant authority mid‑course, then hub again—consistent with temporary inward consolidation after the seizure and later outward coordination. Rank stability is modest: Kendall τ for top hubs shows several weak or negative values (e.g., 5->6 ≈ −0.59; 9->10 ≈ −0.88), quantifying churn during reconfiguration windows. Structural series show persistent disassortativity (≈ −0.65 to −0.38), moderate clustering with peaks during reorganization (avg clustering ≈ 0.53 at phase 6), and a mostly stable max k‑core index of 3 whose membership size varies; the final phase drops to index 2 with a larger core, compatible with diffusion under pressure. Overall, shocks triggered a pivot and leadership redistribution anchored by n12 while retaining a small operational core.

Discussion

The evidence supports a resilience‑through‑reassignment narrative: enforcement shocks do not decapitate the network; instead, authority migrates toward logistics actors (n12) while the long‑standing organizer (n1) alternates between outward broadcasting and inward aggregation. Disassortative mixing and a stable core index imply a centralized, role‑differentiated structure that can rewire spokes quickly. Practical implication: focusing solely on high‑degree hubs risks missing emerging authorities; combining directed (hub/authority) with undirected (betweenness/eigenvector) metrics better identifies succession paths. Limitations include partial observation (wiretap coverage), phase granularity (two‑month bins), and reliance on shortest‑path centrality. As robustness, one could add time‑decayed centralities or bootstrap the rank trajectories; nonetheless, the permutation evidence for n12 and the repeated role flips of n1/n3 make the reorganization claim credible within the course toolkit.

Figures:

* Structural trends over phases (transitivity, clustering, assortativity, max k‑core).

A graph of different types of lines

AI-generated content may be incorrect.

* HITS hub/authority trajectories for n1 and n3.

A graph of different colored lines

AI-generated content may be incorrect.

* Event study (betweenness Δ, phases 4->5 for n1/n3/n12).

A graph with blue squares

AI-generated content may be incorrect.