The Glorious Gorillas
DSS 396

Checkpoint 4: Graph Analytics

As an introduction, we wanted to give an outline of what our research project is to get a better understanding of our research goals. We are interested in investigating the trends between the seniority of an investigator in terms of their age and rank, and their history of complaints. The focus of this investigation is twofold: First, how does the history of complaints correlate with each investigator over time? For example, as an investigator rises in the ranks, does the rate of complaints increase for this officer? Similarly, is there a trend in the type of complaint that the investigator receives? Second, we can understand potential biases that the officers might hold against their victims.

Additionally, investigating this topic may uncover related trends that may not necessarily be confined to the scope of this topic, but are interesting and important, nonetheless. For example, trends between officers that rise in rank and their complaint volume might reveal some information about how the Chicago PD operates as an institution. That is, an officer may receive complaints from the public after an incident but may be awarded or promoted internally but the institution.

The purpose of this checkpoint was to help answer 2 questions through graph analytics:

1. Explore the relationship between co-accusal networks of officer's as grouped by most commonly accused demographic group per officer.
2. Understand network analytics of co-accused officers based on how long they have served for and who they are co-accused with. Basically, understand co-accused networks based on duration of an officer's career.

Question 1

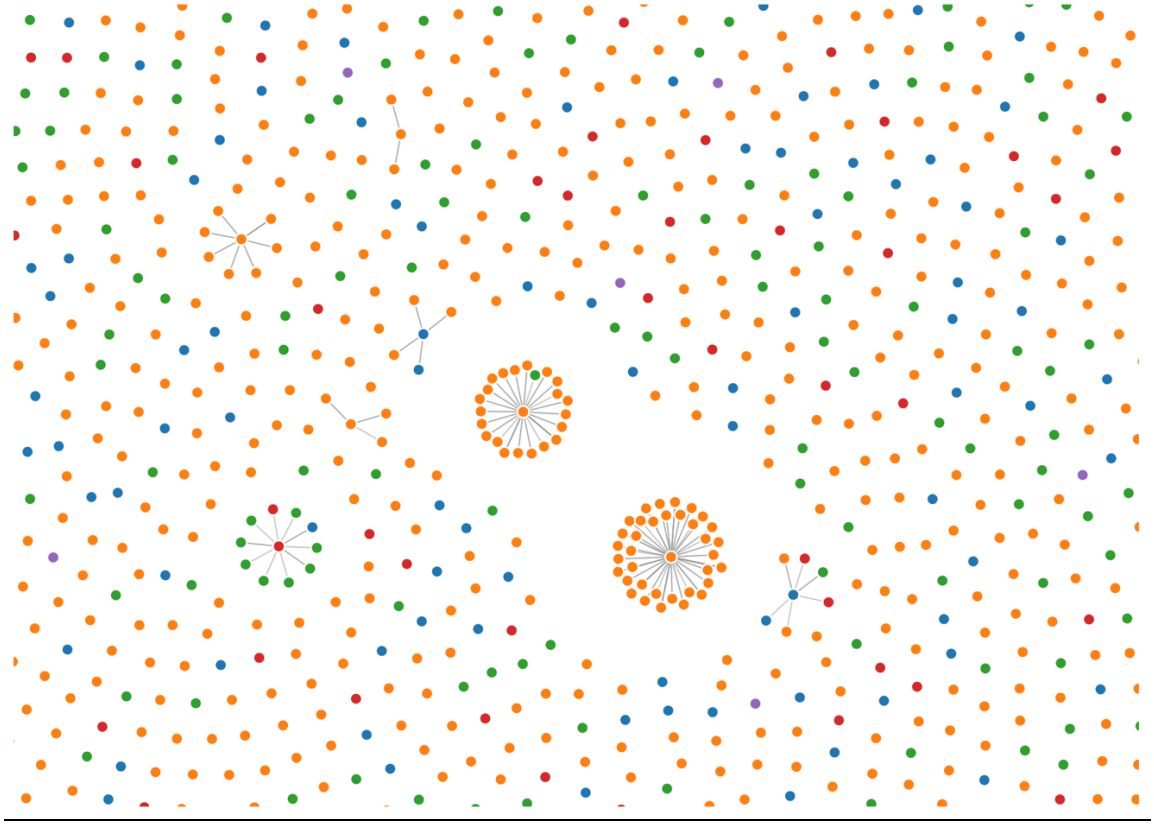
To place it again, our inquiry for question 1 is as follows: to explore the relationship between co-accusal networks of officer's as grouped by most commonly accused demographic group per officer. What this means is we want to understand if groups of accused officers have any relationship on the basis of an officer's most accused demographic group.

We decided the best way to represent this was through network graphics utilizing D3. We utilized a template from D3 themselves and adjusted it for our code. This helps create graph analytics to represent co-accusal networks based on accusation complainant demographic and group involvement of co-accused offers for that accusation. We prepared the data in a way that grouped the nodes (officer_id) based on the most common demographic group of accused per officer (e.g. 'Black', 'White', 'Hispanic', etc.).

To make the graphics work out nicely, we had to limit links down to 200. A researcher cannot be expected to understand graphical representations of all data resulting from our queries.

A link to visualize this can be found here: https://observablehq.com/@ilan2020/network-graph-question-1

With the above link, you can actually hover over each node to see the officer_id and interact with it.



What we learned from this analysis is that an officer does seem to share accusations with other officers who have the same grouping category (with of course, exceptions). This matched our expectations that officer groups would be co-accused with other officers of the same grouping label. Another notable observation from the visualization is that there is a larger proportion of nodes labeled as 'Black' than other label types. This means as there is a much higher proportion of complainants that are of one racial group over all others as shown by the nodes (orange label). It also can be said that there is a strong correlation between officers to be co-accused with other officers of a similar label. Even if you look at the blue label, which is the 'Unknown' label, you can see that is fairly randomly spread out and random in terms of its connections to other nodes and groups. This question certainly also raises further future lines of inquiry. As it relates to our main research question, it would be great to uncover more about networks of demographic grouping for an officer over their career and based on their seniority/years worked. To understand if these groups last over the duration of a career and how it could potentially translate department wide. The labels for the color of nodes (and groups) are shown below.

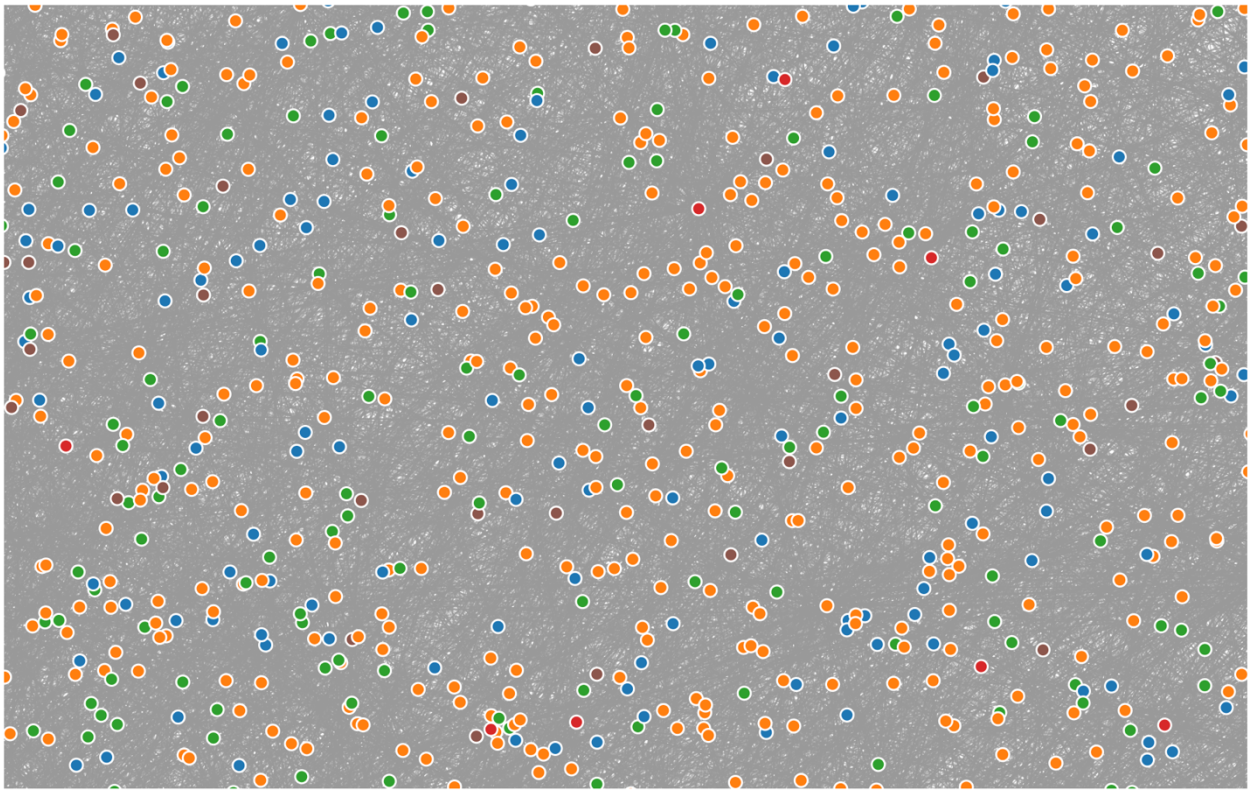
Orange: Black
Red: Hispanic
Green: White
Blue: Unknown

Question 2

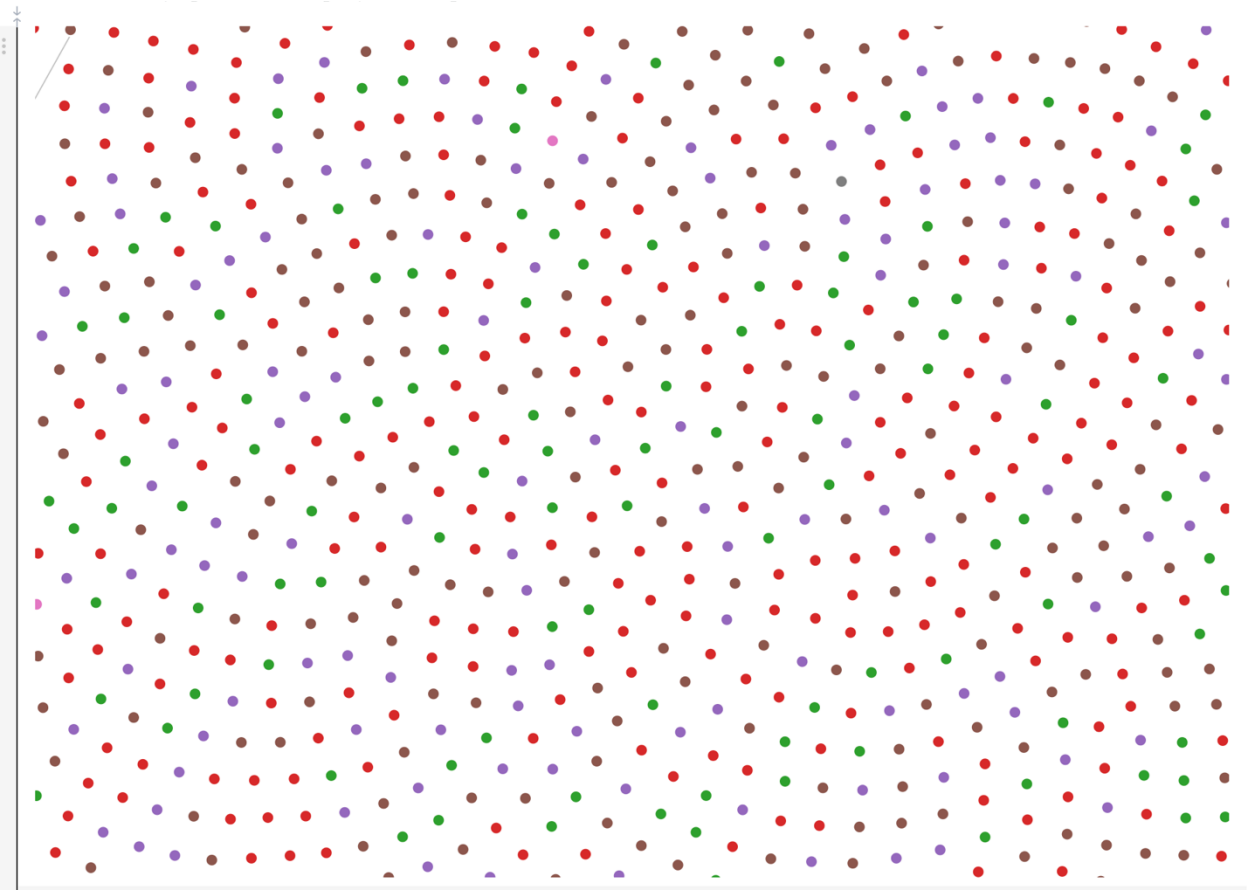
Again, our group's aim was to understand network analytics of co-accused officers based on how long they have served for and who they are co-accused with. It was important to create a question that moved our research forward in helping understand officer seniority from officer's with complaints against them.

To do this, decided on a network graph using D3. We used a really great D3 template from D3 itself. This would help create visuals to analyze co-accusal networks for seniority and group involvement We prepared the data in a way that would group the nodes (officer_id) based on years worked based on groupings of 10. For example, all officers who have worked up to 10 years are labeled as 'green on the graph. From there, we made edges that represent co-accusal between nodes.

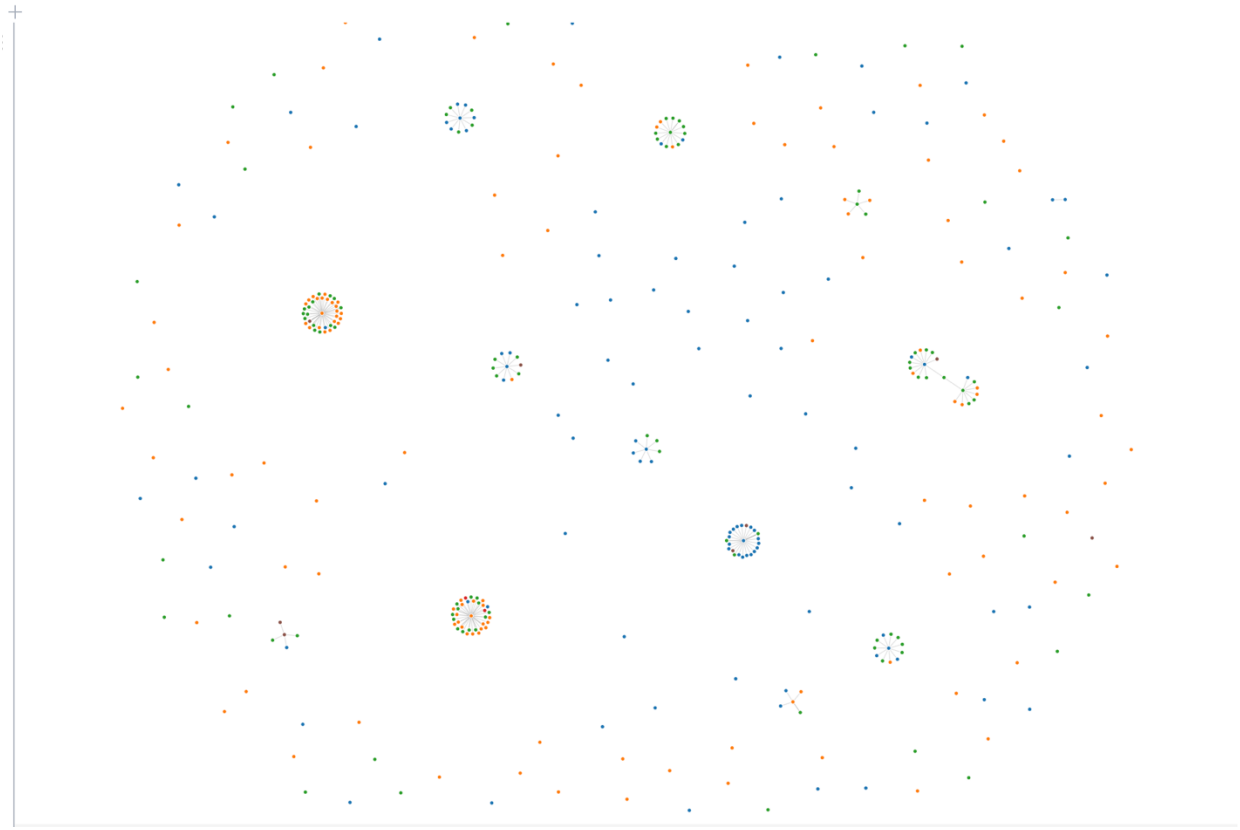
Our first iteration can be seen below. We tried using the entire breadth of data to input into D3. There was obviously way too much data to process as a graphical analytics tool, let alone fit on one screen. Different colored nodes can be seen to represent different officer's and the color for each to represent how long they have been in the force, categorically. The grey lines in-between represent all of the co-accusations between officers.

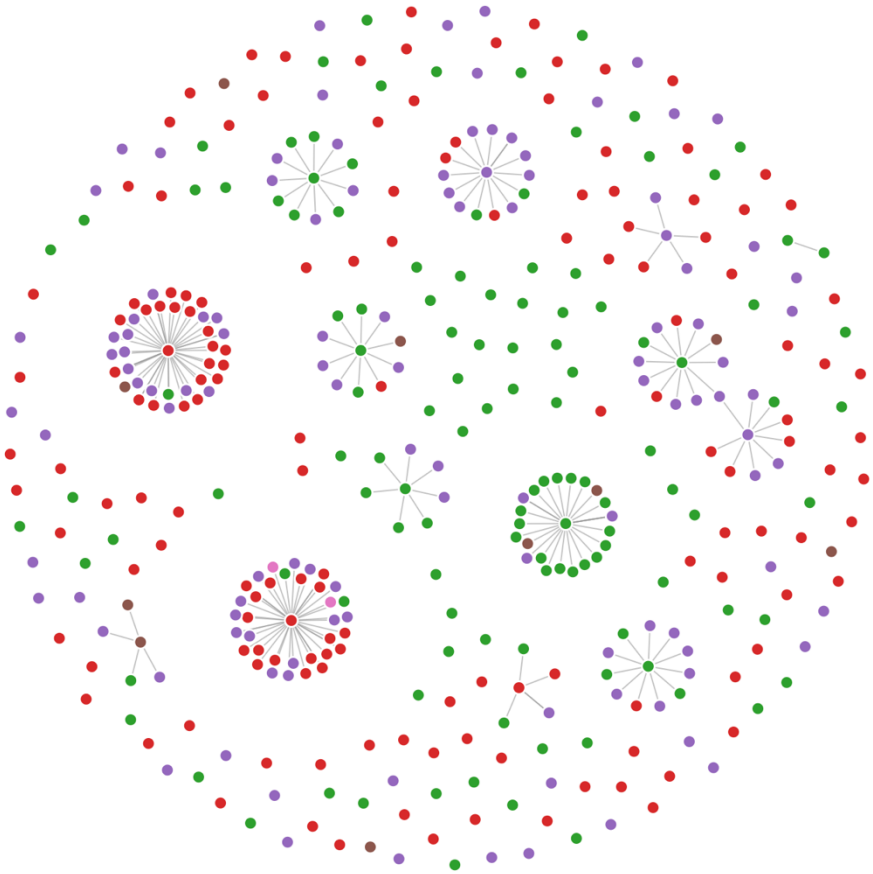


Our group fixed this by reducing the amount of links between officers to process a smaller sample of the available nodes (as shown below). This again proved unsuccessful as the resulting graph showed the nodes with barely any linking connections. It also didn't show a representational graph of nodes either.



The solution our group had, finally, was to limit down nodes and links down to 200. However, this posed more problems as the links would at time reference officer_id's that were not included in the nodes dataset. We had to ensure that all of the available links were mapped to nodes that we actually did have information for in our very small subset. This small subset makes it much easier to do a quick overview analysis on the network graph (as shown below in two forms).





A link to visualize this can be found here: https://observablehq.com/@ilan2020/network-graph-question-2

With the above link, you can actually hover over each node to see the officer_id and interact with it.

What we learned from this analysis is that an officer seems to share accusations with other officers within the same category. This is of course a generalization but holds true overarchingly, nonetheless. This came as a surprise as the graph tells us that officers seem to be accused with other officers of the same ranking. This could make sense logically as officers of the same rank might assigned to work together and patrol together. Moreover, we see aggregation of some higher-ranking individuals in many of the networks. This could be due to the fact that sometimes senior officials get listed in the reports as a co-accuser even if they were not directly involved. This actually is probably some of the best evidence we have collected yet in terms of understanding how seniority can be represented and shown between groups and over time. This certainly raises future lines of inquiry. We would certainly like to now understand a way to differentiate the officers of 30+ years of seniority and understand how to distinguish accusals that were peripheral versus direct involvement and how we can rectify these distinctions. The labels for the color of nodes (and groups) are shown below.

Green:	10 years worked
Purple:	20 years worked
Red:	30 years worked
Brown:	Unknown years worked