

# Supervisors' Influence within the Chicago Police Department

When first starting our project with the Invisible Institute's Citizen's Police Data Project and learning about the Chicago Police Department, a story that caught our attention was the career of Eddie Johnson, the former Chicago Police Department Superintendent. He had a history of misconduct within the department, but was continuously promoted. [Use of Force](#), Part 7 of The Intercept\_'s The Chicago Police Files, highlights a few cases Johnson and his team handled, and says "Johnson and the officers under his direct command proceeded to make a number of troubling decisions." This caused us to wonder what sort of influence was Johnson having in the department, particularly over those he managed. Because he personally has a history of misconduct and has continued this behavior, was he influencing officers that he was in charge of to also engage in this behavior?

Looking at these questions with a more general perspective, we were able to create our theme: Analyzing the influence that supervisors have on the officers they manage. This theme is important because it allows us to measure and potentially predict the likelihood that a supervisor will have a negative effect on the officers they manage, and how much supervisors end up costing the department monetarily.

This second point is important to consider because, as mentioned in The Intercept\_'s [Use of Force](#), one officer's actions can end up costing the department millions of dollars in settlement payments. We thought this was an interesting aspect to study, especially if there's a relationship between supervisors and settlement payments. We were interested to see if exploring these relationships could potentially help to prevent misconduct within the department from spreading, and if it could identify patterns that would save the department money in the future.

Our approach to exploring our theme comes in four parts. In Part I we'll define who supervisors are and what their units look like. Part II will analyze their influence within the department based on co-accusal data. Using settlements data in Part III, we will take a look at how much money problematic supervisors cost the department. Finally in Part IV, we'll use the data from the previous sections to predict potential future outcomes relating to misconduct within the department and settlements costs.

Throughout our project, we used many tools including Tableau and D3 for data visualization, sql and DataGrips for querying the database, and DataBricks and Python.

## Part I: Who are supervisors?

Supervisors for each unit are not specified in our data, so we needed to define this ourselves. To find a supervisor, we looked at the salaries for each unit and chose the officers with the highest salary to be named the supervisor. If there were multiple officers with the same salary tied as the highest, we then looked at their rank and date appointed in order to choose the most senior officer.

We decided to use each officer's complaint percentile as an indicator for misconduct. We chose complaint percentile because it shows an officer's complaint amount in relation to the rest

of the department. Complaint percentile includes all complaints both from the community and from within the department. We chose to look at supervisors above the 75th percentile because we thought this may be an indicator of officers that are frequently committing misconduct.

Supervisors above the 75th complaint percentiles

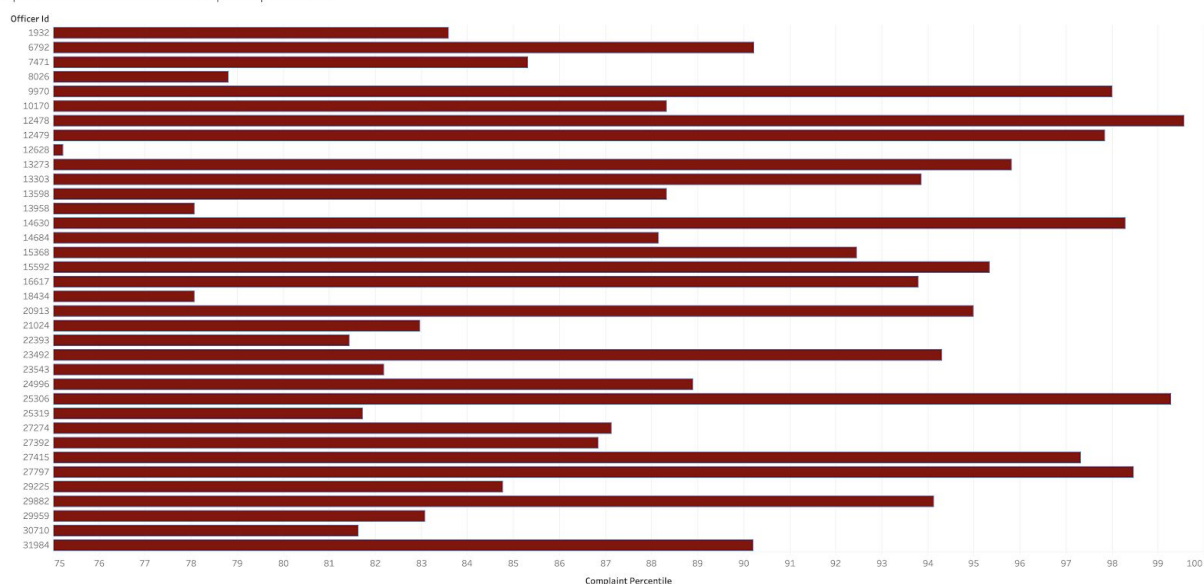


Figure 1: Supervisors with complaint percentiles above 75. The y axis is the supervisor's ID and the x axis is the complaint percentile.

Figure 1 shows all supervisors that have a complaint percentile above 75. There are 36 supervisors with complaint percentiles above 75 out of 98 supervisors total. Within these supervisors, you can see that their complaint percentiles vary.

Our hypothesis was that these 36 supervisors would be in charge of units that also had high complaint percentiles. To calculate a complaint percentile for an entire unit, we took the average of the officers' complaint percentiles within the unit.

The complaint percentiles of units whose supervisors are above the 75th complaint percentile

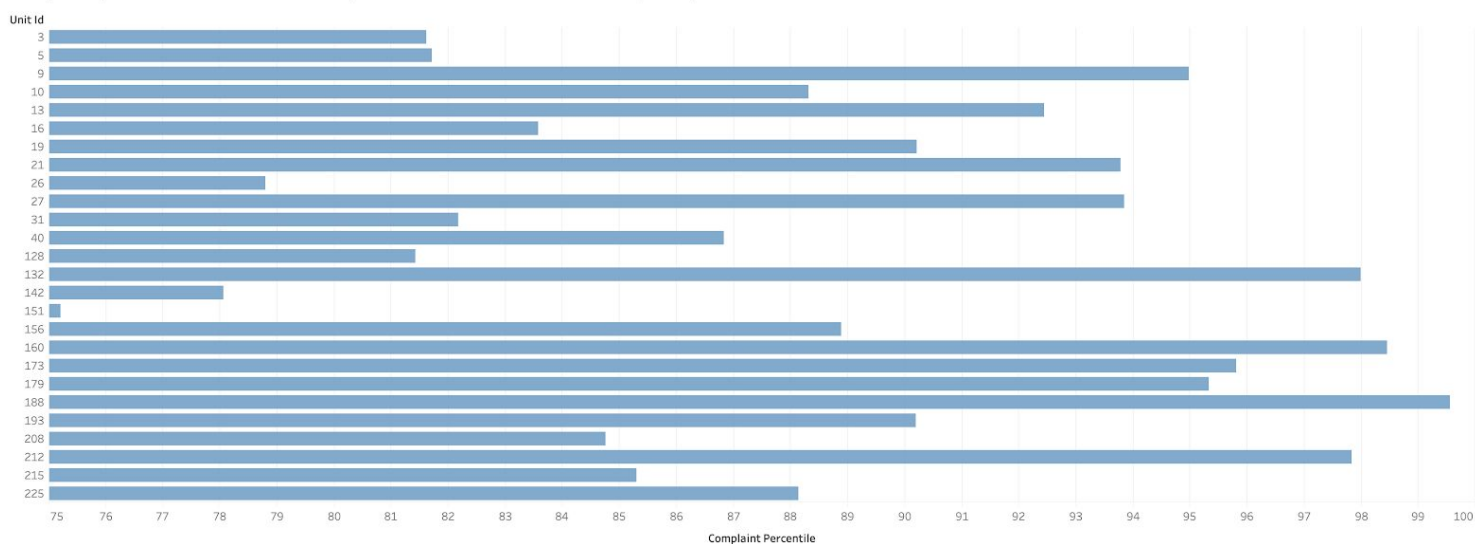


Figure 2: Units whose supervisors are above the 75th complaint percentile. The y axis is unit ID and the x axis is complaint percentile.

In figure 2, these units are displayed and our hypothesis was indeed correct. Every unit managed by the supervisors in figure 1 have a complaint percentile above 75. Looking at these units we wondered how many officers made up each unit. Could a unit's size have a relationship with its complaint percentile? We explored this question by creating an [interactive bubble chart](#). Figure 3 shows a static image of this visualization. We did not find any evidence of a relationship between unit size and complaint percentile.

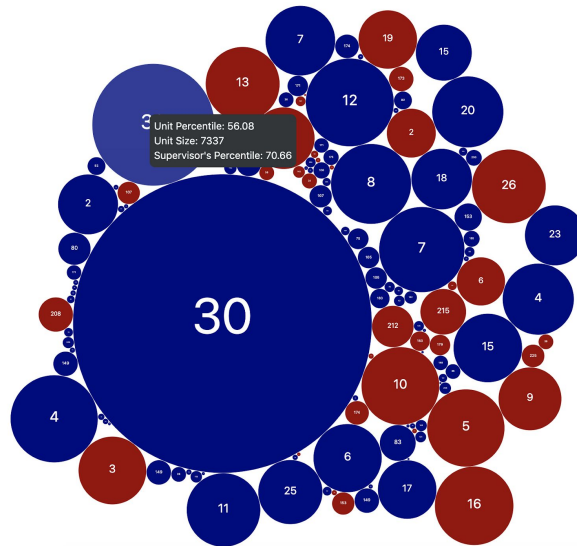


Figure 3: A still image of the bubble chart showing unit size and complaint percentile. The size of the bubble is an indication of how many officers are in that unit. The color red indicates if a unit's supervisor has a complaint percentile above 75. Upon hovering over the bubbles, you can see the unit's average complaint percentile.

## Part II: What is their influence?

To measure a supervisor's influence within the police department, we analyzed accusal data. The officers that are listed together on accusals provide links that we were able to use to create a network. We used degree analysis and data visualization to identify clusters of officers who frequently collude.

Our hypothesis was that supervisors with high complaint percentiles would have high degrees of connections and therefore look like the hubs of clusters of officers listed on accusals together. As you can see in our [network visualization](#) (figure 4 is a still image of this visualization), our hypothesis was proven correct. Through our degree analysis, we found that the officers with the most connections to other officers (the most coaccusals) were all supervisors with high complaint percentiles (above 70). We were not surprised that supervisors seem to be the nucleus of almost all of the clusters of officers who are named on allegations.

However, we did not expect that supervisors would be connected to officers outside of their unit. Our theme is to analyze the influence that supervisors have over those they manage, but our visualization illustrated that a supervisor's reach is not limited to their unit. This may indicate units that supervisors have been previously apart of well as their current unit, or it may

show that the supervisor extends their network outside of their current unit. Whichever the case, it increased the range of units a supervisor was able to influence.

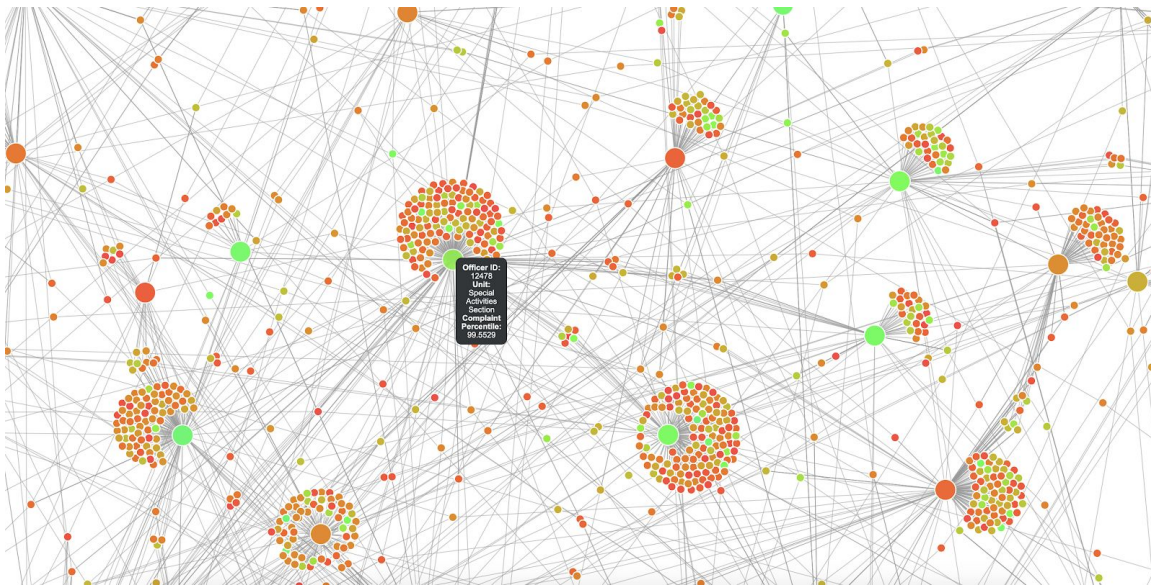


Figure 4: A still image of our degree analysis on the accusals data. Large nodes denote supervisors. Colors denote different units. Upon hovering, the officer's ID, unit, and complaint percentile are shown.

We also used PageRank to identify the most influential supervisors in the accusals network. Our PageRank results surprised us; the top 10 supervisors had complaint percentiles much lower than expected. We expected the most influential supervisors (based on allegations) to have complaint percentiles above 75. However, there's only one supervisor above 75 within our pagerank results, and most of the others are below 30 as seen below in figure 5.

id	complaint_percentile	unit	is_sup	pagerank
11332	18.7116	Juvenile Intervention Support Center (Jisc)	1	6.281494876
16127	16.2164	Technology And Records Group	1	6.281494876
23764	21.2946	Medical Section	1	6.281494876
14595	18.2301	Oemc - Detail Section	1	6.281494876
23543	82.1711	District Reinstatement Unit	1	3.577648432
30568	60.7428	Transit Security Unit	1	2.98695629
23836	51.4692	Bureau Of Organized Crime	1	2.98695629
7611	34.5917	Reproduction And Graphic Arts Section	1	2.812028556
13403	26.8274	Traffic Court Unit	1	2.812028556
7665	30.9809	Information Services Division	1	2.718745561

Figure 5: PageRank results from the accusals network. These are the 10 most influential supervisors according to PageRank. Notice the complaint percentiles are not in line with our other findings.

In our findings thus far, supervisors with complaint percentiles above 75 have been associated with other officers, supervisors, and units with complaint percentiles above 75. This PageRank data does not support the idea that these problematic supervisors (supervisors with complaint percentiles above 75) are a major influence on other officers in the department. This is also contradictory to our degree analysis findings. This requires further investigation as to how the PageRank algorithm weighs its value for each supervisor in relation to our original

findings.

## Part III: How much are problematic supervisors costing the department?

To understand how much problematic supervisors (supervisors with complaint percentiles above 75) end up costing the department, we analyzed settlements data. We used the total two fields, fees\_cost and payment, because they indicate how much money these officers are costing the department. Our hypothesis was that supervisors above the 75th percentile would make up a significant amount of the supervisors named in settlements. When we ran our queries, we found 45 total supervisors named in settlements. Out of these 45, 22 have complaint percentiles greater than 75. This means that 49% of the supervisors named settlements have a complaint percentile that is above 75. We thought this was quite high, but is in alignment with our hypothesis and our findings throughout our project.

Of the 45 supervisors that are named in settlements, payment data were available for 30 supervisors; and we displayed this data in descending total\_cost order in figure 6. We found our results shocking, especially how much money the top officer alone, such as officer\_id 27415, cost the department in settlements, which was \$10,326,759. These results suggest that supervisors are more costly than the average unit of officers when it comes to settlement payouts.

	supervisor_id ÷	unit_id ÷ description	÷	total_cost ÷
1	27415	36 Airport Law Enforcement Section – South		10326759
2	27797	160 Freedom Of Information Section		6887130
3	23492	174 Gang Enforcement – Area South		1710000
4	20271	157 Bureau Of Patrol – Area South		1695999
5	16617	21 District 020		1005500
6	21024	183 Gang Enforcement Division		870000
7	25306	153 Gang Investigation Division		750000
8	8026	26 District 025		449500
9	4620	4 District 003		372000
10	30254	36 Airport Law Enforcement Section – South		332500
11	27274	204 Central Investigations Division		271000
12	1193	174 Gang Enforcement – Area South		258500
13	14630	6 District 005		235000
14	28348	82 Human Resources Division		235000
15	2185	7 District 006		191000
16	1932	16 District 015		175000
17	22077	169 Court Section		96000
18	12628	151 Intelligence Section		85000
19	12479	212 Detective Area – South		85000
20	32313	149 Narcotics Division		80000
21	14192	86 Research And Development Division		70000
22	13303	27 Executive Officers Unit		70000
23	10170	10 District 009		66500
24	25661	42 Marine Operations Unit		49750
25	4131	55 Special Investigations Unit		39000
26	16870	38 Mounted Unit		26250
27	24044	82 Human Resources Division		23000
28	30180	8 District 007		21000
29	11756	126 General Support Division		18000
30	13273	173 Gang Enforcement – Area Central		2000

Figure 6: Results of the total\_cost (the sum of fees\_cost and payment) in descending order for the 30 supervisors payment data is available for.



After reviewing the supervisors' settlements data, we investigated how it relates to the settlements data of their subordinates. We found both the average settlements cost for each unit as well as the total settlements cost of each unit, which can be seen in figure 7.

	unit_id	description	total_cost	avg_cost	total_settlements
1	133	Auto Pound Section	15025000	5008333.33333333333333	3
2	134	Police Documents Section	2000000	1000000	2
3	72	Legal Affairs Section	11868122	912932.461538461538	13
4	132	Evidence And Recovered Property Section	20949647	910854.217391304348	23
5	155	Asset Forfeiture Investigation Section	23536299	574056.073170731707	41
6	85	Inspection Division	6429424	401839	16
7	206	Major Accident Investigation Unit	10592083	320972.212121212121	33
8	208	Detective Area - Central	209321450	314769.097744360902	665
9	107	Bureau Of Patrol	24489198	260523.382978723404	94
10	35	Airport Law Enforcement Section - North	58495694	229394.878431372549	255

Figure 7: Units and their average settlements cost (ave\_cost) and total cost (total\_cost) in descending order by average cost. These are the top 10 most costly units when sorting by average cost.

We found that the highest costing supervisors are significantly higher than the average unit. The top supervisor cost \$10,326,759.00 for instance, whereas the top unit cost an average of \$5,998,333.33. That said, the total cost for each unit varied dramatically. For that same unit, for instance, the total was approximately \$15,000,000. However, the averages for all other units were \$1,000,000.00 and lower. It would be worthwhile in our next set of queries to directly compare the settlement costs of supervisors to the average settlement costs of the units they oversee. It would also be useful to run standard deviation calculations to see exactly how much variation there is within a unit. These results suggest that supervisors are more costly than the average unit of officers when it comes to settlement payouts.

Overall, it seems that problematic supervisors can cost the department a significant amount of money in settlements, and cost more than the average officer.

## Part IV: Predicting Future Misconduct and Cost

Throughout this project, supervisors with high complaint percentiles have been associated with managing units that also have high complaint percentiles. However, we have found that when a unit has a high complaint percentile, it does not guarantee that the supervisor has a high complaint percentile as well.

Our indicator of misconduct in our project has been a high complaint percentile. Keeping with this indicator, we sought to predict a unit's complaint percentile from the given supervisor's complaint percentile. If we are able to predict a unit's complaint percentile based on their supervisor's, there may be an opportunity to make recommendations on who to promote to supervisors in order to keep misconduct in the unit, and overall in the department, low.

Our hypothesis for this question was that supervisors with high complaint percentiles would be predicted to manage units that have high complaint percentiles, and supervisors with lower complaint percentiles would be predicted to manage units with lower complaint percentiles.

We used linear regression for our model because once we plotted the data using unit's complaint percentile on the y axis and supervisor's complaint percentile on the x axis, we saw our data had a linear trend with only a few outliers, as seen in figure 8.

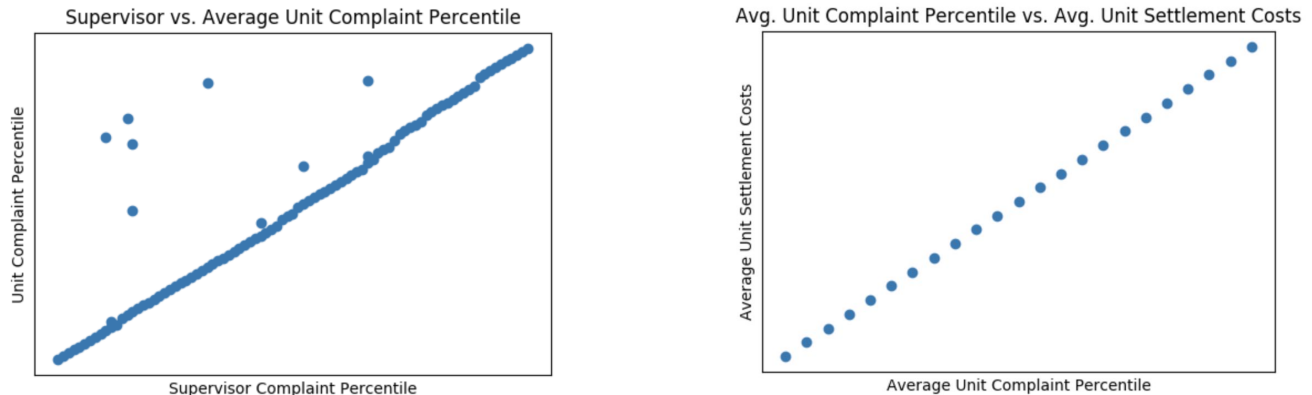


Figure 8 (left): A linear trend in complaint percentile data. Supervisor complaint percentile on the x axis, and unit complaint percentile on the y axis. Figure 9 (right): A linear trend in settlements data. Average unit settlements cost on the y axis and average unit complaint percentile on the x axis.

Unfortunately, the results from MSE and R2 metrics showed that our model did not score well, with high error and a low coefficient of determination ( $MSE=298.65$ ;  $R^2=0.132$ ). We attribute this to not having enough data to train and test on. There are 98 units total, so once we split that up into training and testing, it's not very much for our model to learn from; the previous metrics were reported using a 70/30 train/test split. Other train/test splits did not fare much better. However, based on the appearance of the plotted data, we decided to run a Pearson's correlation test. Our data has a highly statistically significant positive correlation ( $r=0.323$ ;  $p < 0.001$ ) of supervisor complaint percentile to their average unit complaint percentile. Despite not being able to run predictive statistics due to a lack of training data, our correlational results still support our hypothesis.

Our second prediction was regarding settlements. We wanted to predict how much a unit could cost the department in settlement payments by looking at the unit's complaint percentile. This information is important because if we can predict that a unit will cost the department a lot of money in settlements, there can possibly be intervention to prevent that from happening.

Our hypothesis was that units with high complaint percentiles will cost the department more money than units with lower complaint percentiles. We decided to look at cost in terms of the average cost for each unit.

We used linear regression for our model because when we plotted the data, it had a surprisingly strong linear trend, as seen in figure 9 above. We want to note that we were very suspicious about how linear this data was. We checked to make sure our data was mapped correctly throughout our process, and regardless of data conversions we confirmed that column relationships were preserved. We found no indication that any new mapping had occurred.

Again, the results from MSE and R2 metrics showed that our model did not score well, with extremely high error and a low coefficient of determination ( $MSE > 5 \times 10^9$ ;  $R^2 = -0.081$ ). We had to do a 50/50 train/test split, or else the error was much worse. We attribute this to not having enough data to train and test on. While there are 98 units total, we only have settlements data for 23 units. Once we split the data into training and testing, it's very little for our model to learn from. We found our model's scores interesting because there is such a clear correlation in this data. Thus, we tried running a Pearson's r test again. However this time the correlation was not significant ( $r=0.139$ ,  $p=0.527$ ). Some resources suggest calculating Pearson's correlation coefficient is not advised for data under 25 samples, so this could explain the inconclusive p value. Thus, no meaningful relationship could be discerned between average unit complaint percentiles and average unit settlement costs.

Our predictions seemed like they would compliment each other; If a supervisor with a high complaint percentile is more likely to lead a unit with a high complaint percentile, and a unit with a high complaint percentile is likely to cost a high amount in settlements, promoting an officer with a history of misconduct to a supervisor role may end up costing the department a significant amount monetarily. However, without enough data, our models are not able to make these predictions.

## Conclusion

Our initiative for this project was to analyze the influence that supervisors have on the officers they manage. Our hypothesis was that supervisors with a history of misconduct would influence those they manage to be more likely to commit misconduct, as seen through Eddie Johnson's career in the Chicago Police Department.

In Part I, we found was that supervisors above the 75th complaint percentile manage units with an average complaint percentile above 75. This supports our hypothesis that supervisors with a history of misconduct, as identified through a high complaint percentile, influence the officers they manage, proven through the units' high complaint percentiles.

Our network analysis in Part II confirmed our hypothesis that supervisors are hubs for misconduct, as proven by having the highest degrees among all officers when analyzing accusals data. We also found that supervisors' influence goes beyond just their unit. We were surprised to see the amount of units represented in a single supervisor's network. This may show an influence the supervisor has had in past units as well as their current, or it may show that the supervisor extends their network outside of their current unit. Either way, it increased the range of units a supervisor was shown to influence in a negative way.

A key finding from Part III was that supervisors above the 75th complaint percentile cost the department a significant amount in settlement payments, which is more than the average officer. While in line with our hypothesis that problematic supervisors cost the department a large amount in settlement payments, we were not expecting how high the actual dollar amounts were.

In our final section, we sought to predict the outcomes of a unit's complaint percentile based on the supervisor's, and the amount of settlements money a unit will cost the department



based on the unit's complaint percentile. We did not have enough samples for either of our models to make these predictions. However, there is still strong correlational evidence that a high supervisor complaint percentile is related to a high average unit complaint percentile; we laid the groundwork for future data scientists to answer these questions with our models when there is sufficient data.

## Future Research

While we were able to draw conclusions from our research, there is definitely more research that could be done to continue gaining insights on our theme. We would like to compare settlement costs of supervisors to the average settlement costs of the units they oversee. We also think there is more analysis to be done of the pagerank results as they were very different from our cluster analysis. Questions that remain include: can we predict how an officer's promotion to a managerial position will affect allegations within the department based off of their history? And can lenient punishments within the police department for misconduct lead to further misconduct? Exploring these concepts and answering these questions would be extremely useful to glean more information about supervisors and their influence on those they manage.