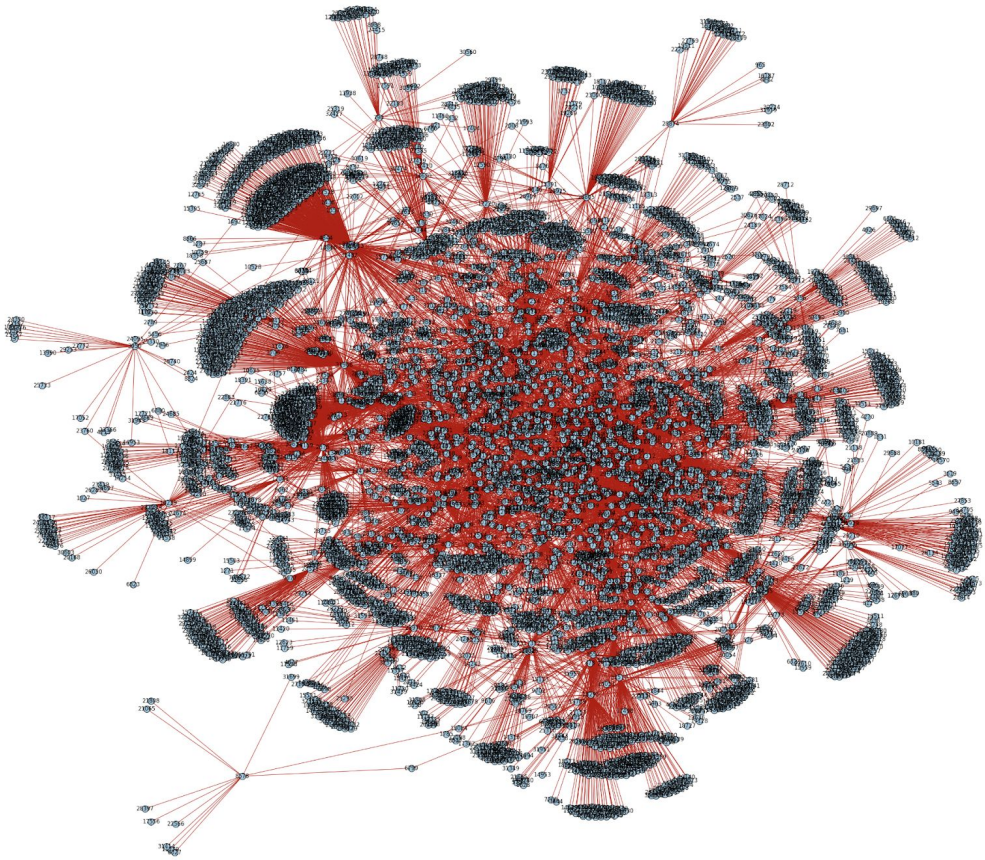
Theme: The theme of the project is to analyse the trends before and after the cpdp went public. For example: to find changes in general behaviour of the officer, settlement amounts etc. and to predict for upcoming years. The questions are focussed towards finding out the things that have been tried to be hidden in the official record. For example, finding the relationship between the arrests and the allegation, because there are chances that if no arrest took place then the action taken by the officer might be illegal or looking for cases where things common between investigating officer and allegated officer impacted the case. So, for every question mentioned below, findings will be based on before and after cpdb release.

1. **Can officers who have been co-accused for misconduct be clustered into groups? Are there any common attributes for the most important nodes from the graph? (Using page rank to find the most import nodes)**

   Graph structure:
   Each officer is a node. Nodes have an edge between them if they are co-accused in any allegations.

   For data before 2015:



For data after 2015:

There are several clusters formed for data before 2015. We have taken a smaller sample of the entire data for better visualization. For the data after 2015, there is not enough data to conclude anything but the expected trend would also be to have fewer number of allegations since the data has gone public.

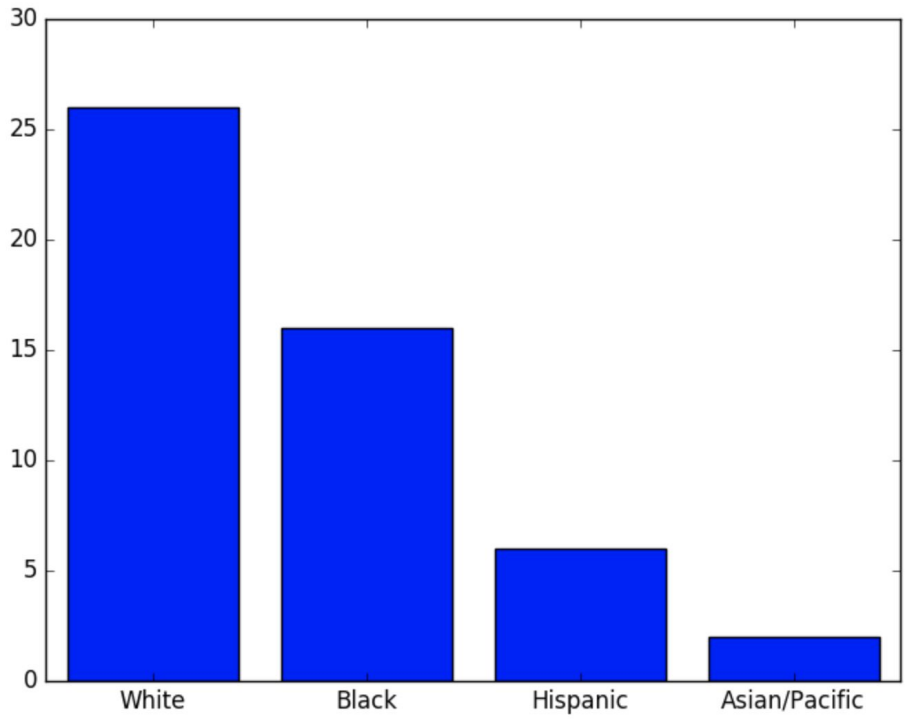
To analyse the most important nodes for both these graphs, we have performed page rank. Page rank will help us find the most important nodes; in this case, the officers who are major players in the co-accused network.

Page rank for data before 2015:

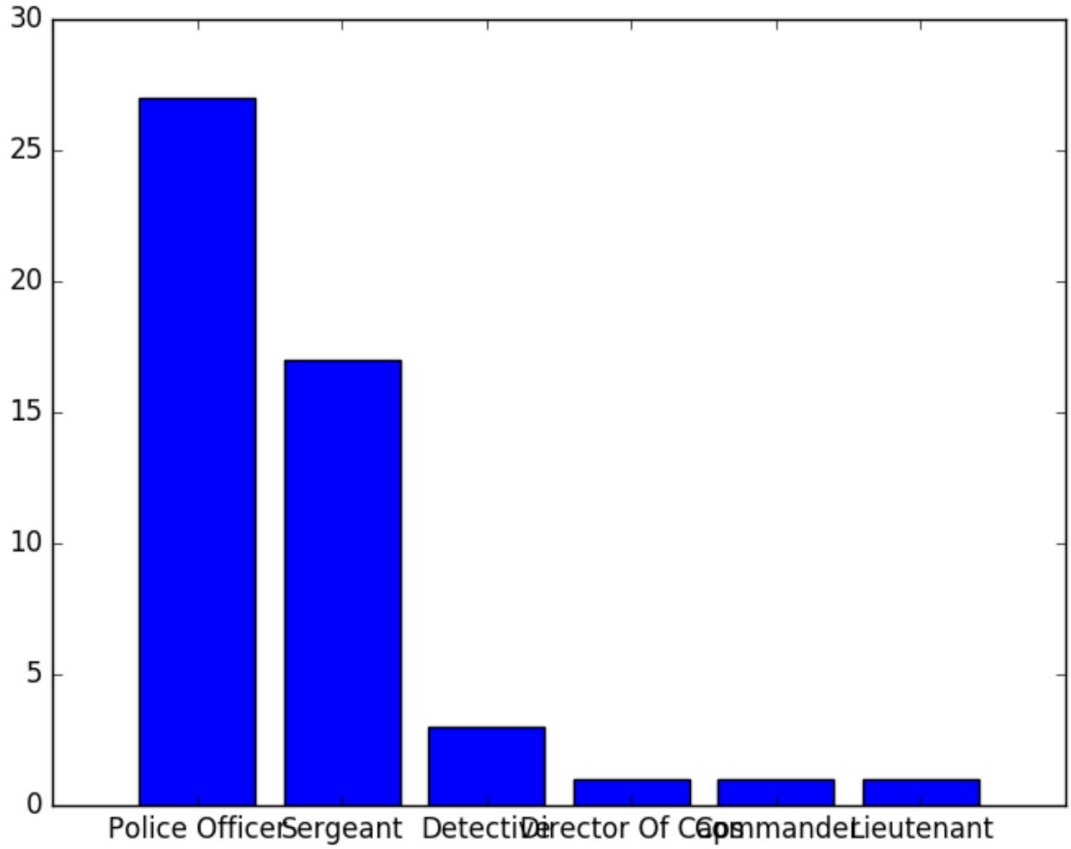
id	pagerank
8138	88.971293107427
31906	73.85933693369522
2375	59.55497965626626
4844	50.93177291514222
17816	47.85829685642143
16567	45.85660812921399
1553	45.5083712258918
8562	42.970205695668405
32255	39.121435071993155

As we can see, the officer with id 8138 has the highest page rank value of 88.97. This officer is Glenn Evans. So, we took the top 50 officers with highest page rank value and tried to see their characteristics.

Bar chart for top 50 officers by race:



By rank:



But these results are **not normalized.**
So, we perform these normalizations manually. Our sample consists of 100 officers and all their links. The distributions for these 100 nodes in terms of race and rank are:

	count bigint	race character varying (50)
1	3	Asian/Pacific
2	34	Black
3	16	Hispanic
4	47	White

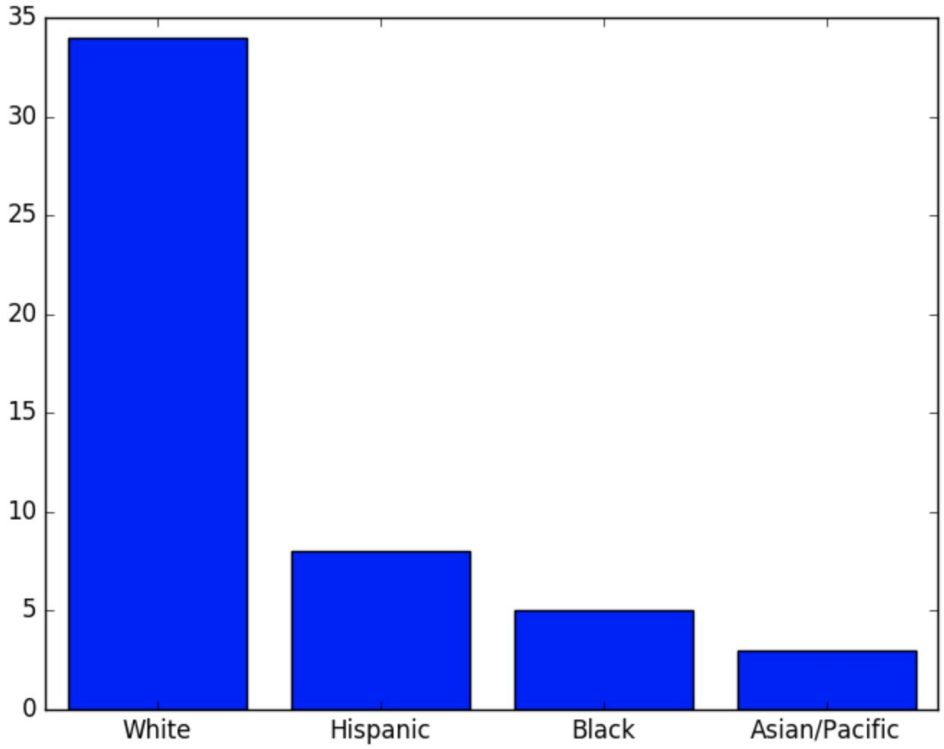
	count bigint	rank character varying (100)
1	2	Commander
2	8	Detective
3	1	Director Of Caps
4	3	Field Training Officer
5	3	Lieutenant
6	61	Police Officer
7	22	Sergeant

After analysing the top 50 most important nodes we see that more than half of those were whites, followed by blacks. **So, out of the 3 total Asian officers, 2 of them were part of the top 50 officers. So, studying them more could provide significant information.**

In case of ranks we have: `[('Police Officer', 27), ('Sergeant', 17), ('Detective', 3), ('Director Of Caps', 1), ('Commander', 1), ('Lieutenant', 1)]`

The Director of Caps and Sergeant have the highest normalized count. Director of Caps has (1/1 = 1) and Sergeant has (17/22 = 0.77). So, these ranks should be studied more.

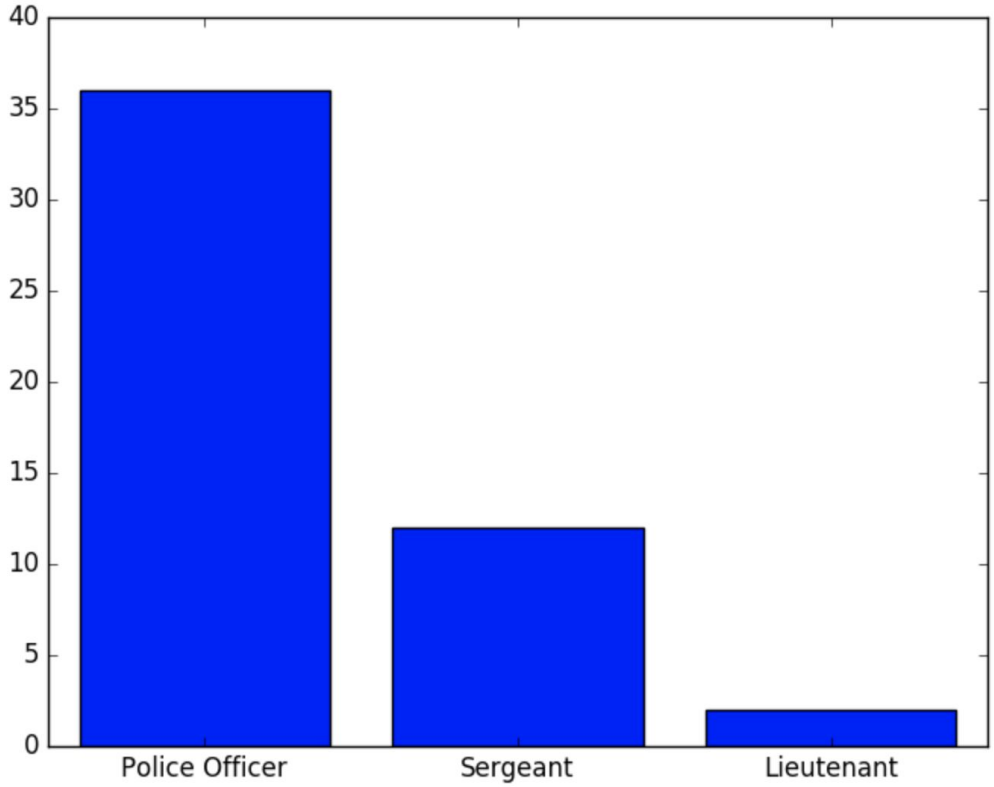
Page rank analysis for data after 2015:



[('White', 34), ('Hispanic', 8), ('Black', 5), ('Asian/Pacific', 3)]

Normalized count:
White = 34/47 = 0.72
Hispanic = 8/16 = 0.5
Black = 5/34 = 0.147
Asian = 3/3 = 1

Yet again, there is a high number of normalized Asian count. The three Asians from our original data are present in the list of top 50 officers.



[('Police Officer', 36), ('Sergeant', 12), ('Lieutenant', 2)]

Normalized count:

Police Officer = 36/61 = 0.59

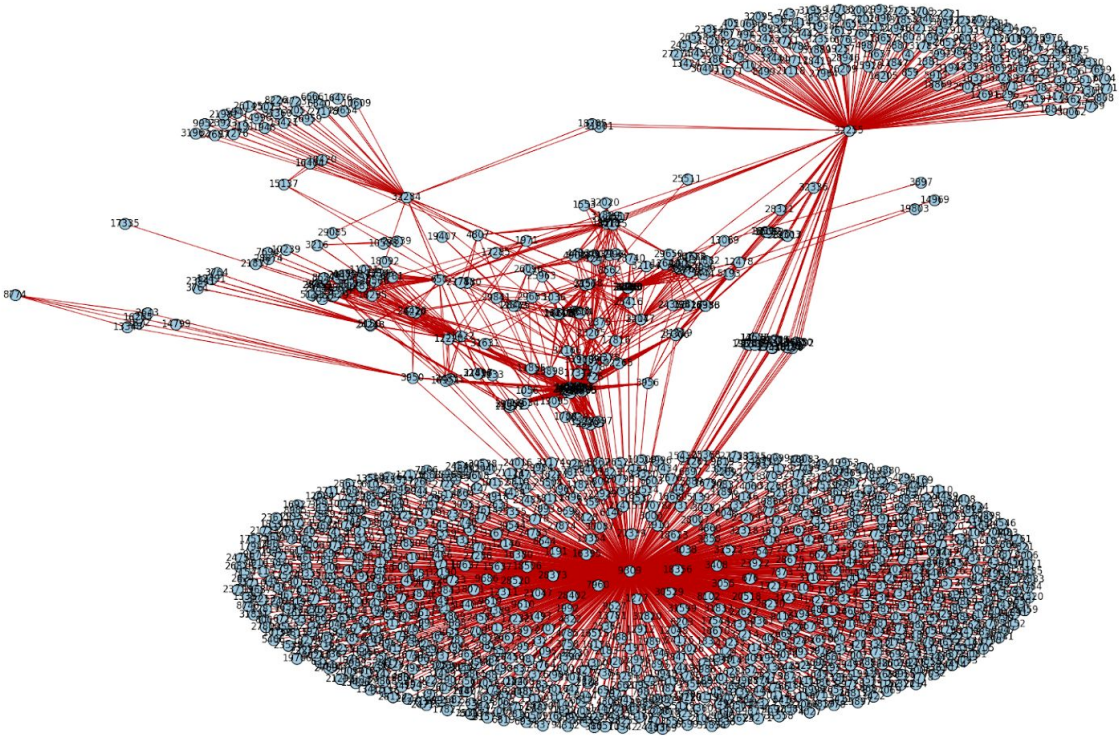
Sergeant = 12/22 = 0.54

Lieutenant = 2/3 = 0.6

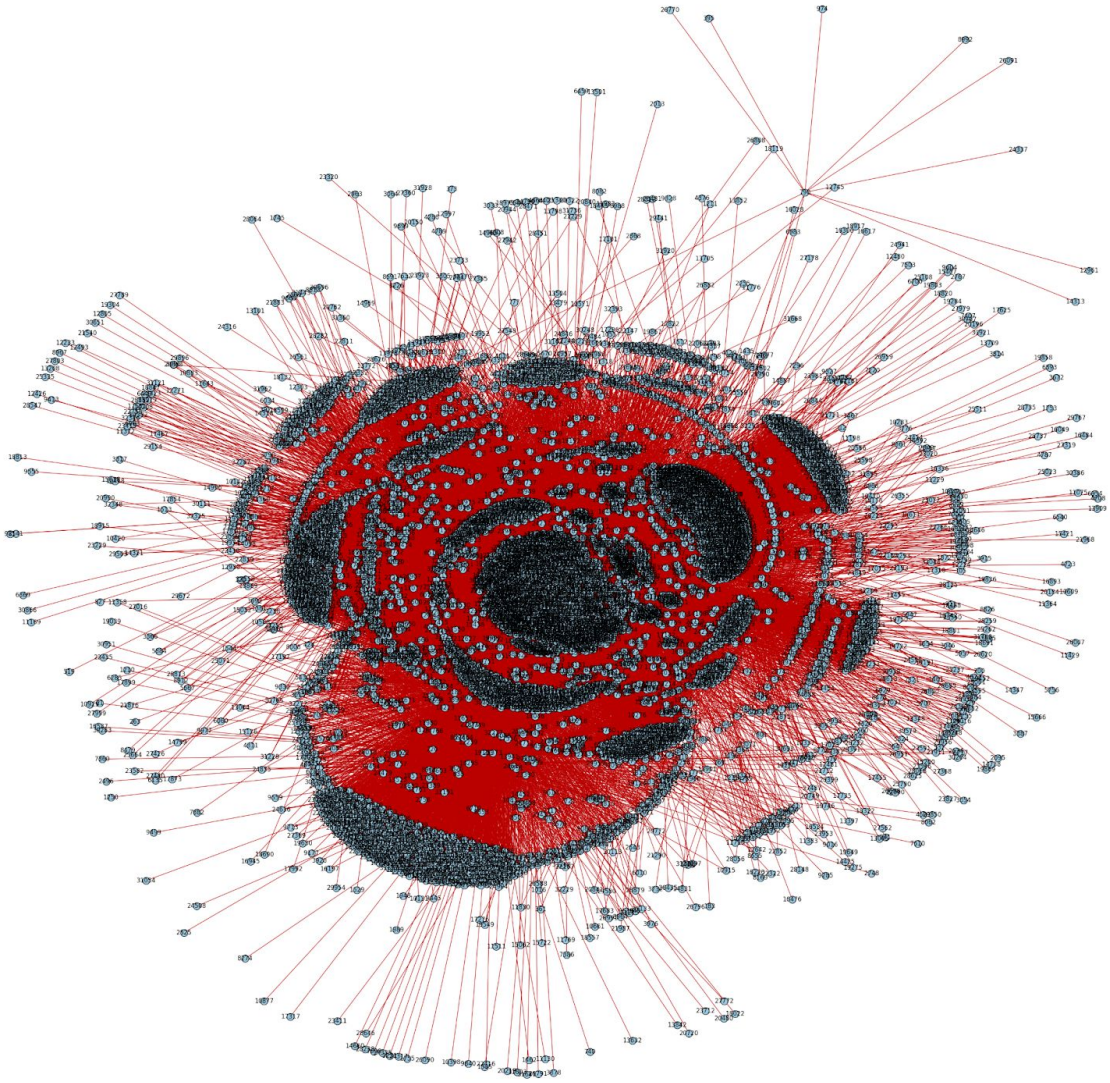
After 2015, the relevant normalized officer ranks are different from the ones before 2015. Also, only 3 ranks out of 7 are present in the top 50 officer list.

2. **Can the officers who are charged for the same allegation category be clustered into groups? Are there any common attributes for the most important nodes from the graph? (Using page rank to find the most import nodes)**
   Graph structure:
   Each officer would be node. Nodes are connected if they have the same allegation **category.** This will help us find if officers with similar characteristics are performing similar kind of offenses.

   Graph for data after 2015:



Graph before 2015:



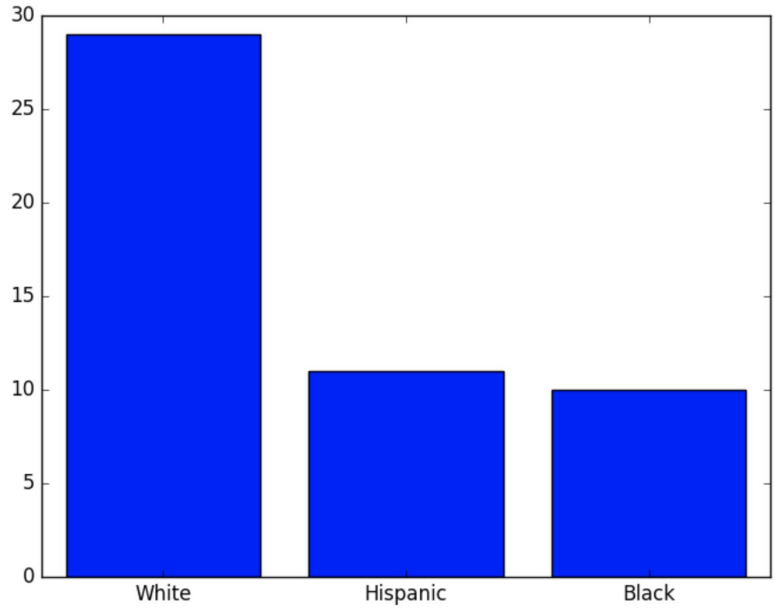
For data after 2015, we have two big clusters and several other small clusters. Both the big clusters have prominent centers. For data before 2015 there is a huge cluster and may have several clusters but has too much data to visualize it. Therefore we will use page rank to analyse the most important nodes.

We have considered a random sample of size 3% of the original data for better visualization in the graph.

Page rank analysis:

	count bigint	race character varying (50)
1	3	Asian/Pacific
2	34	Black
3	16	Hispanic
4	47	White

	count bigint	rank character varying (100)
1	2	Commander
2	8	Detective
3	1	Director Of Caps
4	3	Field Training Officer
5	3	Lieutenant
6	61	Police Officer
7	22	Sergeant

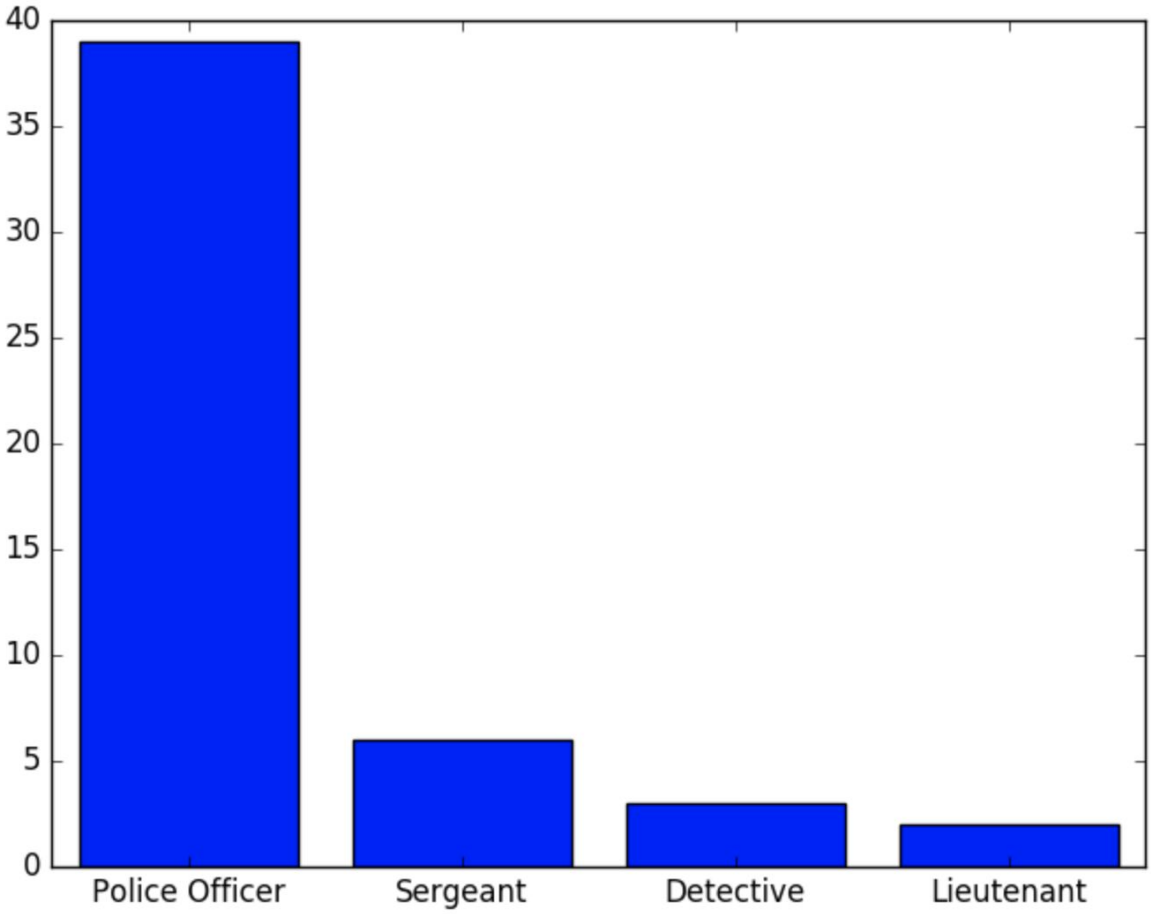


[('White', 29), ('Hispanic', 11), ('Black', 10)]

Normalized count:

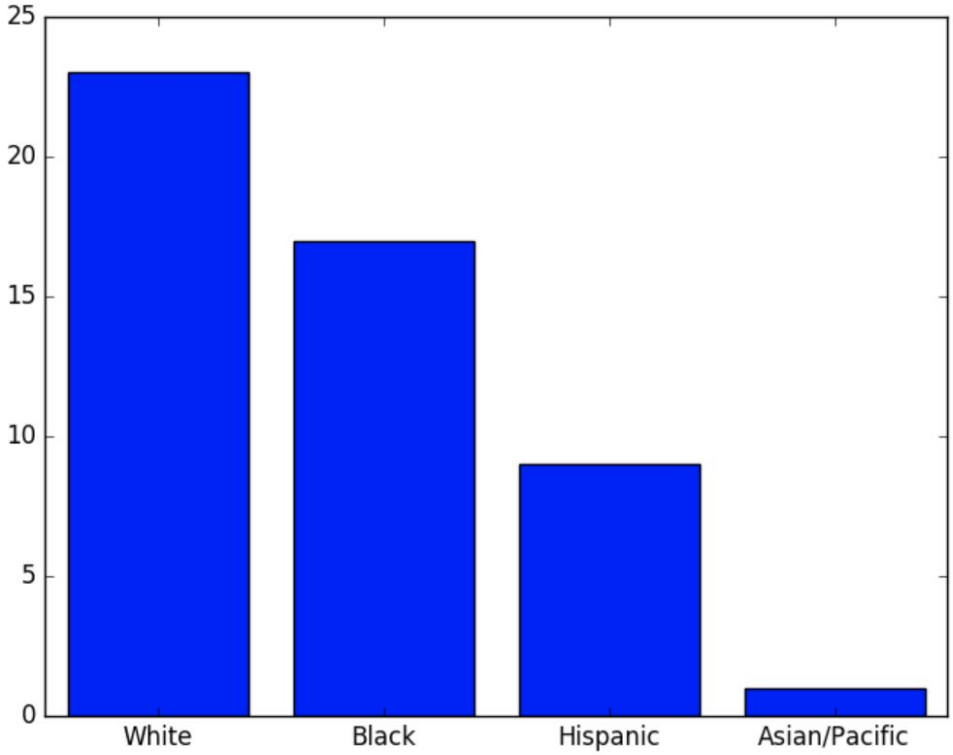
White: 29/47 = 0.617
Hispanic = 11/16 = 0.687
Black: 10/34 = 0.294

So, in this case, White's and Hispanic officers have significant normalized counts. So we can say that these officers are connected to many other officers with same or different allegation category. So, the impact of allegation category can be studied by studying these officers further.



[('Police Officer', 39), ('Sergeant', 6), ('Detective', 3), ('Lieutenant', 2)]
Normalized count:
Police officer = 39/61 = 0.639
Sergeant =6/22 = 0.2727
Detective = 3/8 = 0.375
Lieutenant = 2/3 = 0.66
So, police officer and lieutenant are the significant ranks. So, these ranks should be studied further.

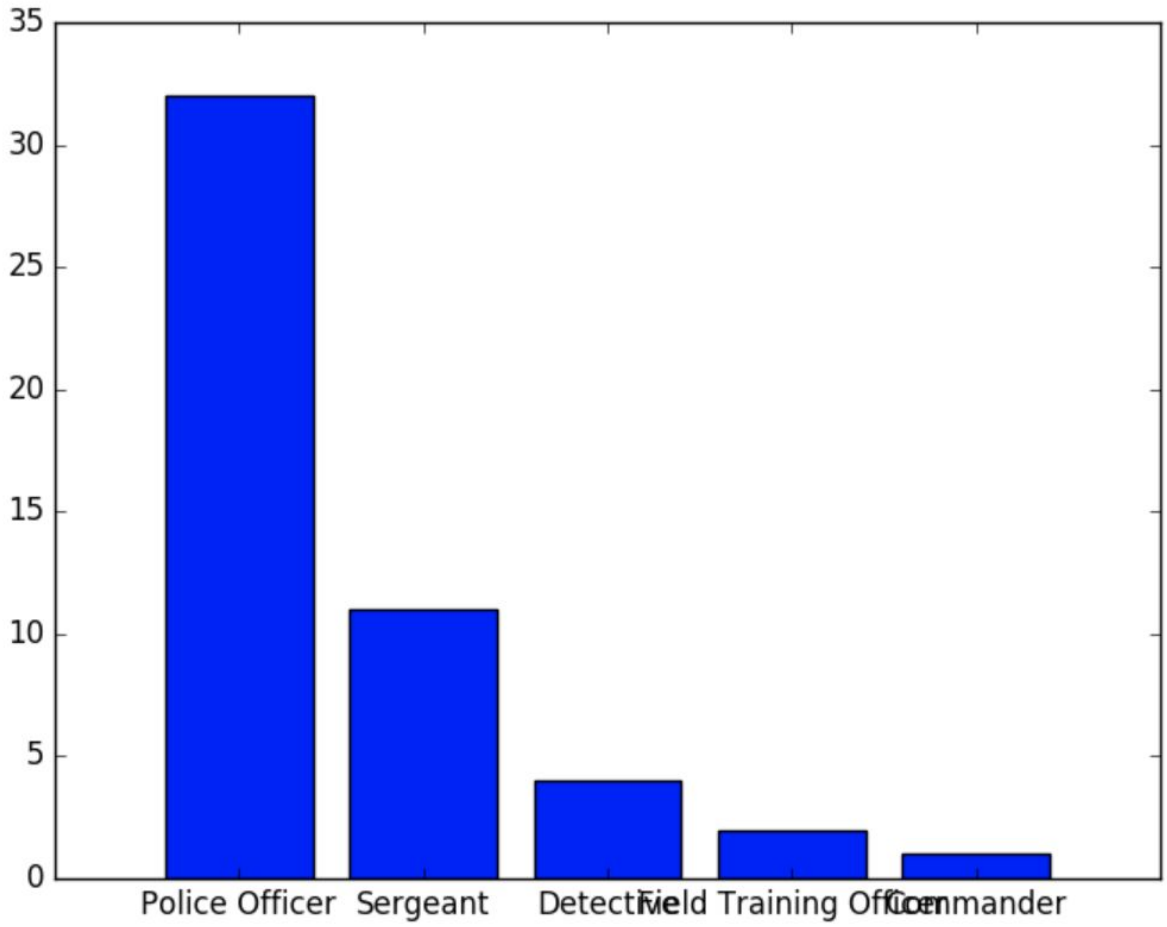
Page rank analysis:




```
[('White', 23), ('Black', 17), ('Hispanic', 9), ('Asian/Pacific', 1)]
```

Normalized count:
White: 23/47 = 0.489
Black: 17/34 = 0.5
Hispanic: 9/16 = 0.562
Asian: 1/3 = 0.33

In this case, the Hispanic officers have the highest normalized count followed by Black and then by white officers.




```
[('Police Officer', 32), ('Sergeant', 11), ('Detective', 4), ('Field
Training Officer', 2), ('Commander', 1)]
```

Normalized count:

Police officer: 32/61 = 0.524
Sergeant = 11/22 = 0.5
Detective = 4/8 = 0.5
Field Training Officer = 2/3 = 0.66
Commander = 1/2 = 0.5
Here, the field training officer and police officer have the highest normalized count.

Bonus: Here is the link for interactive visualization of the growing officer network by allegation categories. A slider has been provided to see the growth over years
https://observablehq.com/@rautnikita77/officers-network-by-allegation-category