Checkpoint 4: Graph Analytics

The Wise Lobsters

Theme: _Identifying allegations early in a repeater officer's career and exploring patterns that lead to increasing allegations over time._

Question 1

How often are repeaters co-accused with an officer's first allegation of misconduct? Does this pattern form a network?

Overview

Assumptions:
● We have considered only those officers appointed between 2000-01-01 and 2007-12-31.
● Officers having 10 allegations count are considered as Repeater Officer. The goal of this question is to determine whether first time offender is co-accused with a repeater and how many other officers does he co-accused going forward. We need to trace the network where one repeater corrupts X first-timer officers who can then go forward to get accused with Y other first timer officers. In this case, we have focused specifically on violent allegations and removed allegation categories Lockup Procedures, Traffic, Supervisory Responsibilities, Unknown, and Medical.

Step 1:
● Firstly, we have queried and created officer subset (active officers with appointed date between 1-1-2000 and 12-31-2007, this limits to officers with full allegation data and at least a 10 year experience) where we determined repeaters officers on the basis of the criteria that is, repeater officer is an office having 10 allegations count.
● There allegation types does not come under the categories ('Operation/Personnel Violations', 'Lockup Procedures', 'Traffic', 'Supervisory Responsibilities', 'Unknown', 'Medical').
● We have excluded the above categories because this category does not come under violent allegations criteria. So we have selected categories other than the above mentioned.
● We have added one more field called repeater_start_date which is when the officer became a repeater (date of the 10th allegation). And finally, we have created a temporary view termed as "officer_subset" from above created data frame.

Step 2:
● For the network we want to include all repeaters and not just officers in our subset. These officers will have no out-degree in the network, but can have in-degrees when they co-accuse with a first time allegation of an officer in our subset.

Step 3:
- On the next step we have created the nodes which are:
  - Nodes (Vertices): Officers in subset and Repeaters from the entire dataset along with id, appointed date and repeater's start date.
  - Some officers in our subset are also repeaters

Step 4:
- Created a view for first allegations of officer in our subset.

Step 6:
- We have determined different officers having common allegation id. These are co-accused officers. Along with this, we have matched the co-accused officers with repeater to identify relation between them which is identified on the basis of allegation id.
- We also tried to identify the relationship between officers having relation with non-repeater officer. There might be a scenario where the first allegation officer can later on have same allegation id with another officer's having first allegation.
- These first allegation links create the edges of the network.

Analysis

Now as we have nodes and edges, then we have used graph frames for the displaying the out degrees for the nodes. Looking at in degrees and out degrees of the vertices, we are able to determine the maximum number of allegation id along with their allegation category.

• We have displayed allegation categories (Violent only as described above). From the graph below we can observe that allegation category "illegal search" is done by maximum number of officers.
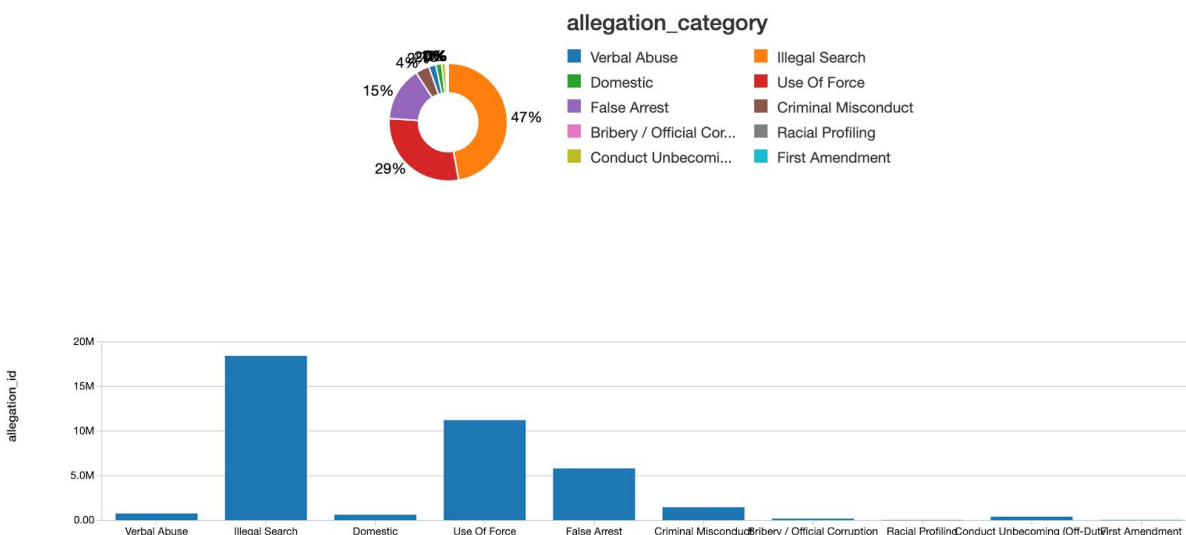
• Edges shows id, appointed_date and repeater_start_date for the officers.

| id	appointed_date	repeater_start_date
28119	1991-01-02	2003-03-30T00:00:00.000+0000
29935	1982-06-14	2010-02-19T00:00:00.000+0000
29066	1994-10-03	2001-12-10T00:00:00.000+0000
29256	1999-03-08	2003-02-08T00:00:00.000+0000
29894	1997-07-07	2005-11-21T00:00:00.000+0000
24180	1996-07-08	2014-02-25T00:00:00.000+0000
30351	1990-03-26	2001-02-21T00:00:00.000+0000
3284	1998-04-13	2007-08-17T00:00:00.000+0000
10530	1994-10-03	2013-05-13T00:00:00.000+0000

Showing the first 1000 rows.

Plotting the Network graph
Note: We have plotted graph using networkx library and by converting Spark dataframes to pandas.

By selecting officers with first allegations and forming relation between repeaters having common allegation id (edges). We are able to determine co-accused officers along with the repeaters. The graph forms clusters of officerts and shows a pattern that how first time offender is co-accused with a repeater and there are many other officers get co-accused going forward.
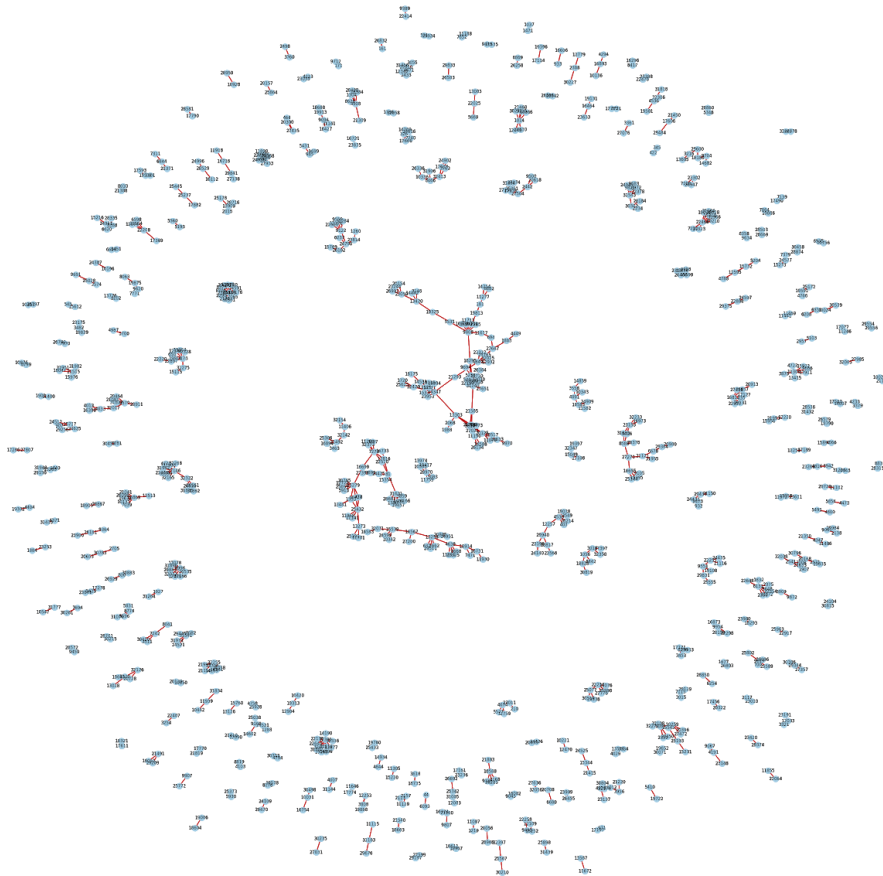
From the entire network, we took the officers the in-degree greater than 2, meaning they have been co-accursed with a first time allegation with at least 3 different officers. From there we look down 3 levels to see how these cluster from. The result is the visualization below. A few large clusters have started to form.
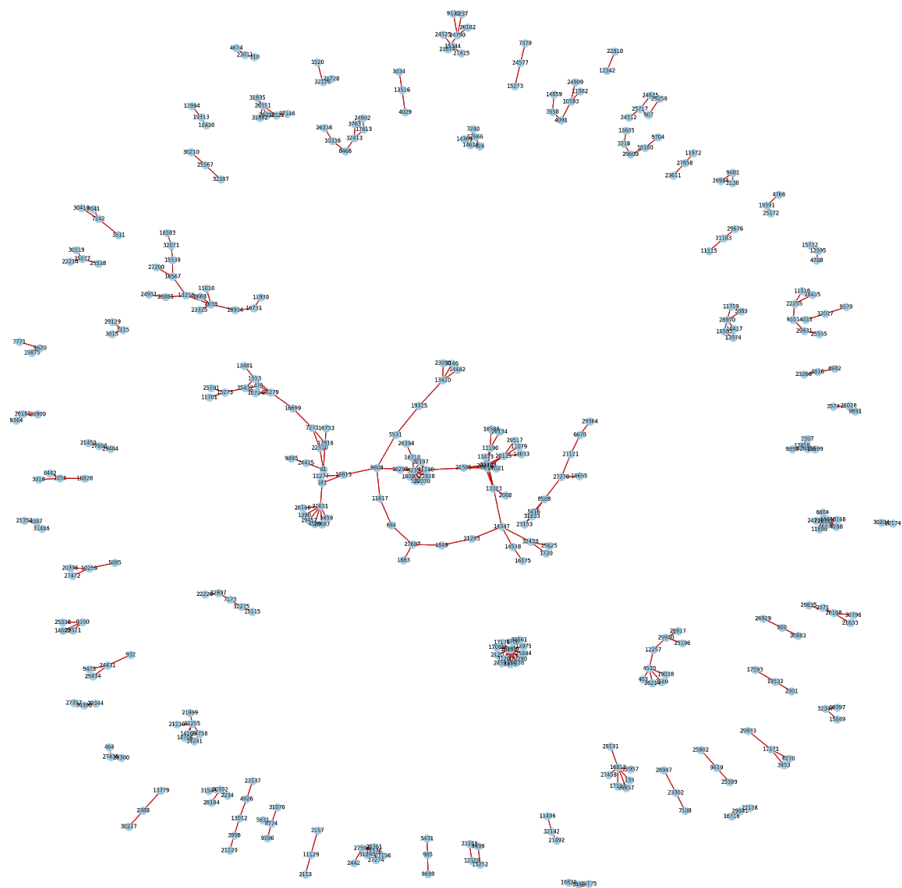
Thus this analysis was able to answer the goal of finding the network and we were able to relate it to our theme of Identifying allegations early in a repeater officer's career and exploring patterns that lead to increasing allegations over time.

Relating To Theme & Future Research

The formation of these networks can provide a point to our theme that an officer's first allegation, if with a repeater, can form a path towards repeated behavior. If the first time offending officer in the clusters are repeaters themselves and have allegation counts above the

norm, then this could be an indicator early in an officer's career of how likely they are to have repeat behavior. The following are some questions we can explore in future analysis.

- We can look deeper into the cluster being formed and find the average number allegations for officers in these clusters and identify whether they are higher than average?
- We can identify are there types of allegations that are most likely in these clusters?

Question 2

Do allegations of a certain type/name tend to draw a path to a certain award?

Overview

The goal of this question is to see how often awards are given in a time frame directly after an allegation. We focused specifically on violent allegations and removed allegation categories Lockup Procedures, Traffic, Supervisory Responsibilities, Unknown, and Medical. Due the amount of time it took to compute the edges of this graph and the number of failures we received running on the free databricks tier, we focused specifically on awards. Promotions could be a good analysis to look at in the future.

This graph analytics is less of a network and more about the weights (count) of edges between two vertices. There are two types of vertices, allegation type and award type. An edge of this graph is a record of data_officerallegation where an award was given within 60 days of an allegation incident date. With this there are many edges between vertices, but the amount of edges is what we are focusing on. The vertices are constructed with a type (allegation_category or award_type) and a value (the allegation category name or award type) Edges will only link from a source of type allegation_type and a destination of award.

Since this workflow cannot be run in a responsible amount of time, here is a step by step walkthrough of what we did to build the graph:

1. Queried for all awards received by officers in our subset (active officers with appointed date between 1-1-2000 and 12-31-2007, this limits to officers with full allegation data and at least 10 years experience). Queried only for awards that are in final status.
2. Queried for all allegations for officer in our subset with a violent category.
3. Created vertices with distinct allegation categories and award types
4. Created edges by iterating through each allegation and finding any awards that were received by that officer within 60 days of the incident date

Analysis

Our graph resulted in 25,599 edges. Looking at the in-degrees and out-degrees of vertices we can see what allegations categories are quickly followed by an award and what awards are preceded by allegations.

In-Degree

Honorable Mention is by far the most predominant award given and therefore has the highest in-degree. In our subset of data there are 155,977 Honorable Mention awards and nearly 20,000 examples of awards directly following an allegation Below is the top 5 awards measured by their in-degree, followed by a table of total number of awards version edges found for each of the top 3 award types:
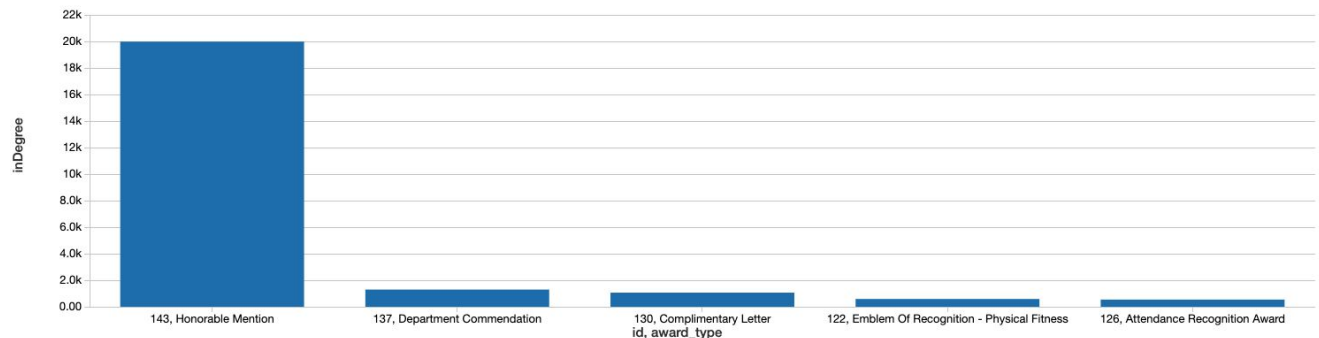

| Award Type	In-Degree	Total Number of Awards
Honorable Mention	19,986	115,997
Department Commendation	1,297	13,359
Complimentary Letter	1,065	13,836

There is a chance of awards being repeated in our edges, so you cannot directly extract a percent from the table above, but there is a strong number of allegations that result in a reward quickly after. In general, awards do not follow after an allegation which is thankfully what you would expect.
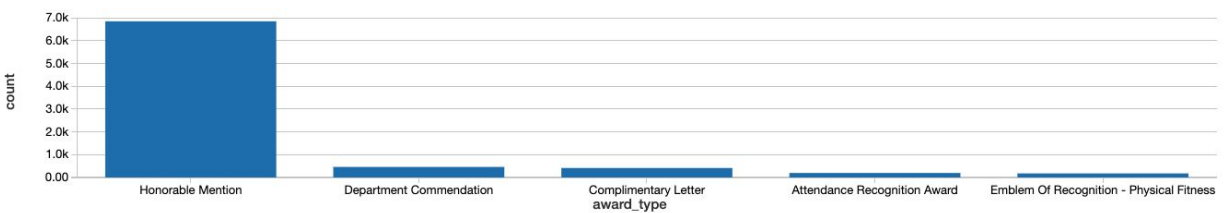
Out-Degree

Out-degree is more spread out, as we can see by the below graph of allegations measured by their out-degree (top 10)
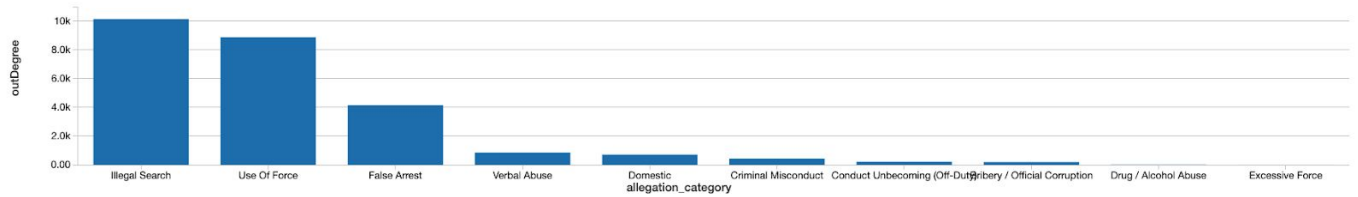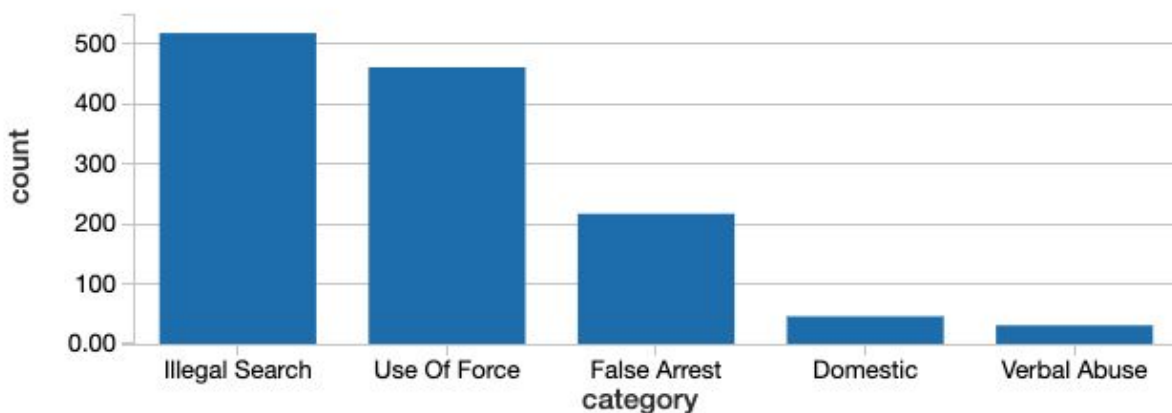
As one would expect, the vast majority of awards given after any allegation is Honorable Mention. Below is the count of awards given after Illegal Search:
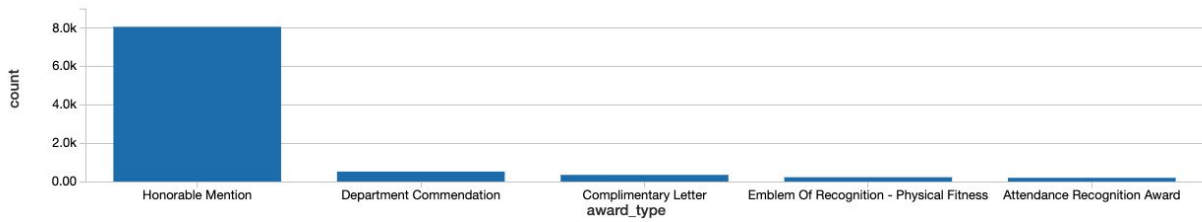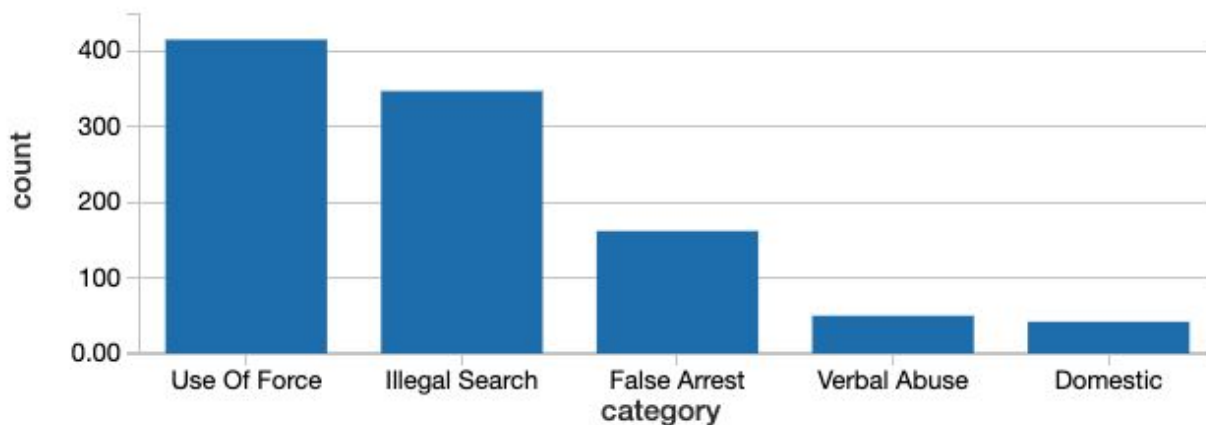

And Use of Force:

Looking at in-degrees, we explored two awards below Honorable Mention, Department Commendation and Complementary Letter. Below is the distribution for Department Commendation by in-degree (displaying top 5 categories)
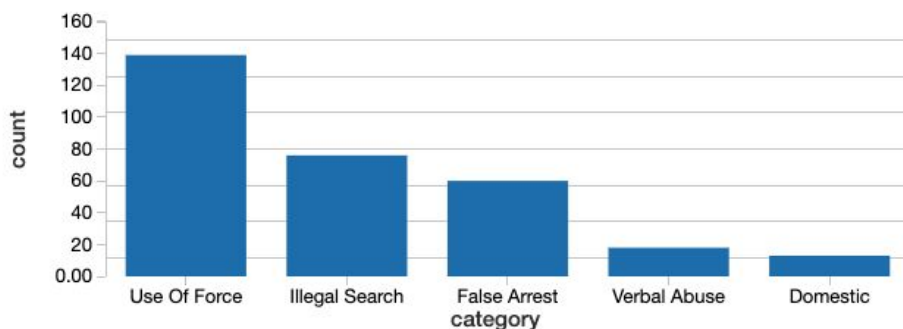

And the same visualization for Complementary Letter

count

400

300

200

100

0.00

Use Of Force   Illegal Search   False Arrest   Verbal Abuse   Domestic

category

Again, this is not a surprising finding since the top three categories remain consistent and the distribution remains similar. Given this analysis, we are concluding that awards may not reflect in any way allegations filed by the public, which can lead us to further exploration of this topic in the future (see below). The best representation of this conclusion we found is the number of the case of 139 "2009 Crime Reduction Awards" issued within 60 days of a Use of force allegations.

count

160
140
120
100
80
60
40
20
0.00

Use Of Force   Illegal Search   False Arrest   Verbal Abuse   Domestic

category

Relating To Theme & Future Research
Do answer our question, yes, there is a lot of allegations that often lead to award, but most do not. Since the vast majority of awards are honorable mentions there isn't much information to extract other than the event is not uncommon. We can draw a generalized conclusion about what awards mean related to officers behavior as reported by. Our analysis for this question leads to the assumption that awards cannot be used as a representation of behavior. Also note that the edges of hour grah do not indicate a direct link to an allegation and an award. Instead it indicated that bad and good behavior can be recognized in a relatively small period of time.

This analysis does lead to some new questions about the relationship of awards and allegations:
● Does receiving an award lead to an allegation?

-
  - This is essentially the flip of the analysis we did for this question, do any awards lead to allegations.
- What happens when we shorten the timeline? 30 days, 7 days, etc?
  - Do we see new patterns emerge?
- Is there a stronger association between awards and allegation names (allegation category sub type)
- (not data related) What is the award processes within the chicago police department?