Checkpoint 4: Graphic Analytics

The Silent Foxes

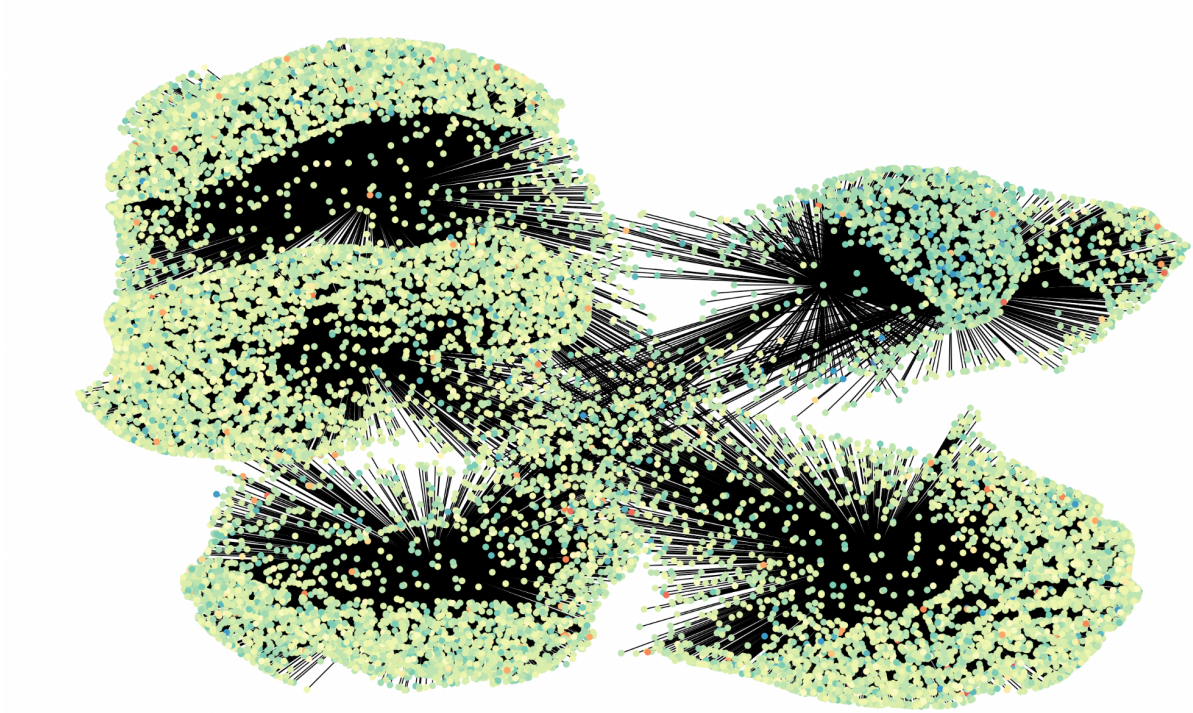
Our theme is to explore the relationship between the misconducts in police officers and their career development.

Questions List:

1. Is there a relationship between the clusters of officers with most complaints and their salary(and awards received)?
2. Can we identify clusters of officers who are likely to be co-accused by factors related to their career, like working environment, salary, rank?

Q1: Is there a relationship between the clusters of officers with the most complaints and their salary? What about working environment

(1) We collect officers who have received complaints and their salaries. We regard each officer as a node, and connect nodes in the same range of complaint percentile. We group officers with the complaint percentile by 0, 20, 40, 60, 80, 100. For example, if two officers have 0 complaints, we connect these two nodes. If one officer has 15.7 as complaint percentile and the other officer has 11.4 as complaint percentile, we connect these two nodes. Also, the color of nodes reflects the current salary of each officer, which a node with lighter color, such as green, means the officer has a lower current salary compared to other officers; a darker color means a higher salary.

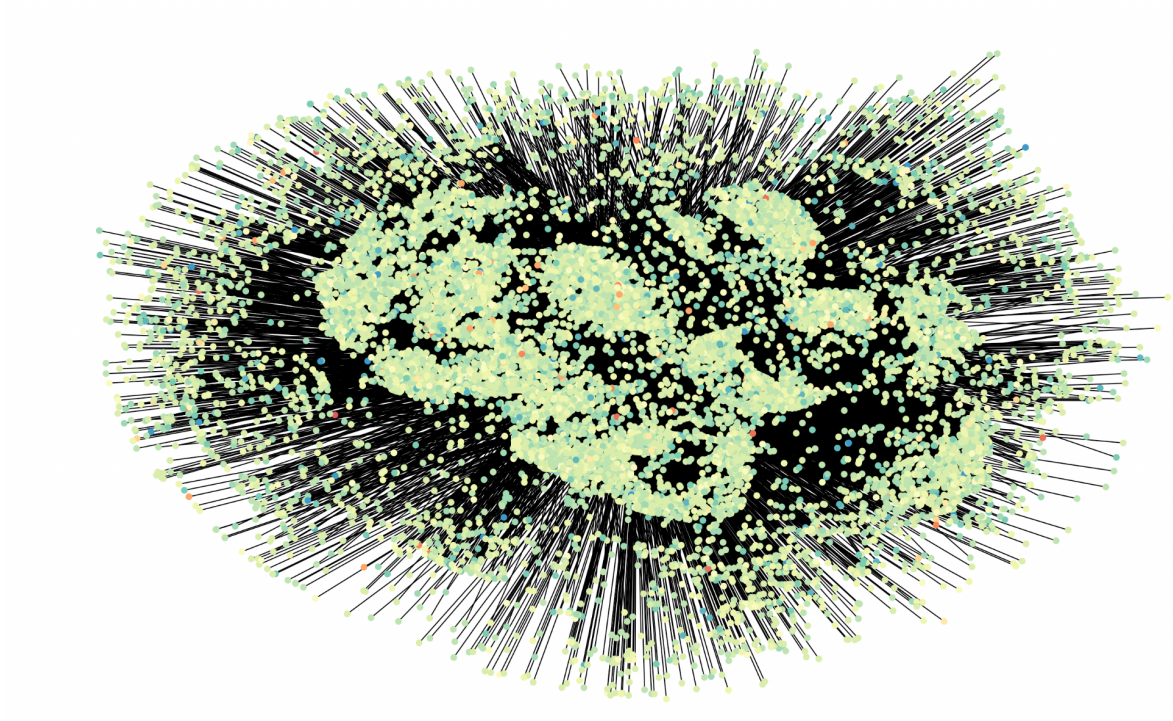


With the settings we describe above, we create this network in a spring layout with iterations = 10, and k = 0.1. In this network, we have a total of 17991 nodes, 17985 edges.

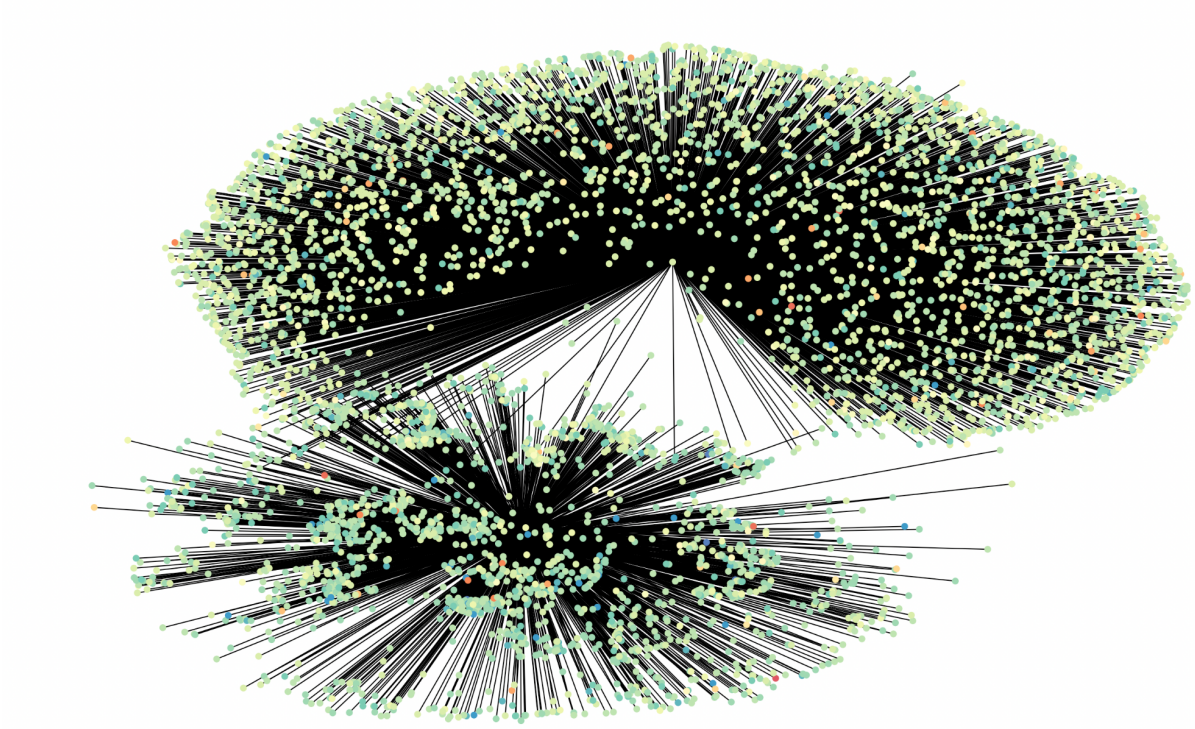
As shown by the network, there are five clusters, which represents the number of complaints received by officers in different ranges we mentioned above. By looking at the color of each cluster, we can see officers' salaries in each cluster are consistent, which is basically similar in color. And the overall color of the nodes in the network is light, which means most officers who receive complaints have a lower salary compared to overall salaries. Also, we notice that the cluster of nodes at the top right corner of the network seems to be darker than other clusters, so we would like to take a closer look to find out more information. The total number of nodes in a cluster cannot be shown clearly in the graph of this spring layout network, but we can roughly compare the density of them. This cluster seems to have the least number of nodes compared with other clusters, so we calculate the total number of officers in each range of complaint percentile. We find out that the cluster at the top right corner of this network is the group of officers who receive 0 complaints, and a darker color means a higher salary. This

finding is consistent with a conclusion we have in previous analysis, that officers who receive higher salary have less complaints.

(2) We add more edges by using last_unit_id to find out if there is a relationship between number of complaints received by officers and officers' working area. Based on the network we described above, we add edge if officers have the same last unit id, which means they work in the same area.



In this network, we have 17991 nodes, and 472157 edges
To make it clear, we create a network only adding edges for clusters with 0 complaint and complaint percentile below 20.

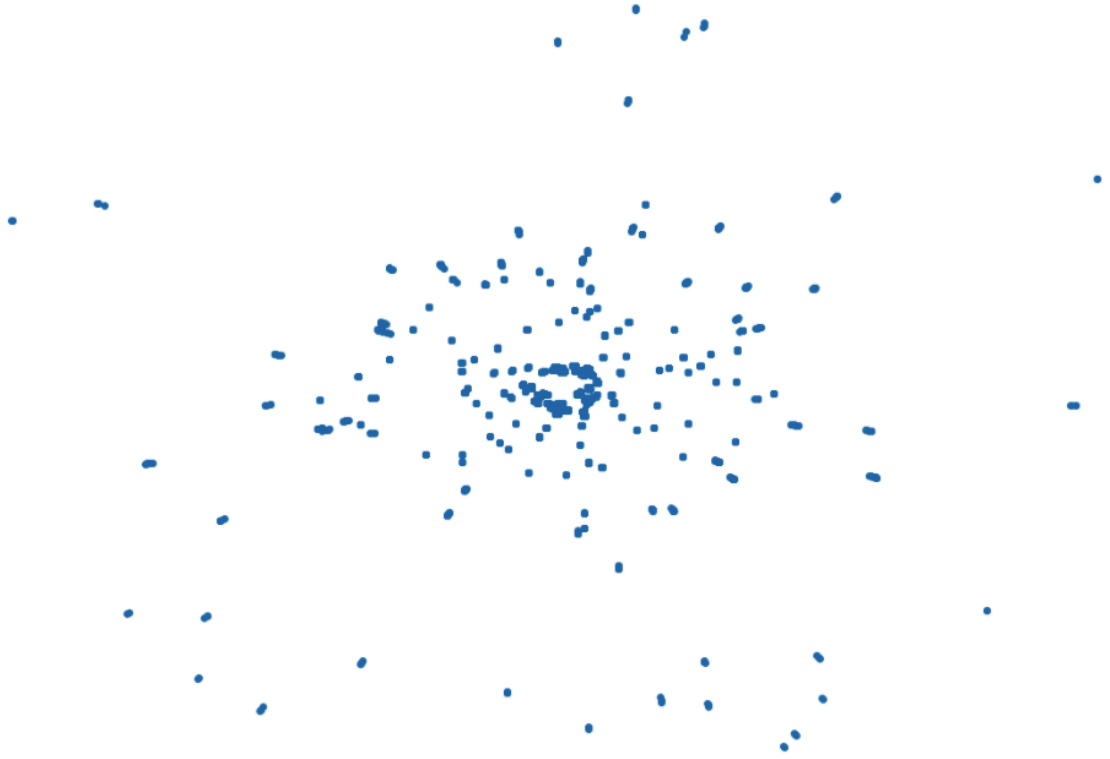


From this network, we can see there is no clear connection between two clusters by using the working area, which means officers who receive a similar number of complaints seem to come from the same working area. This provides us with a further analysis question: Does public security in the specific area have influence on the total number of complaints received by officers working in the same area? If there are more incidents in an area, the officers working in this area may receive more complaints, which we could take these factors into account in future analysis if dataset supports.

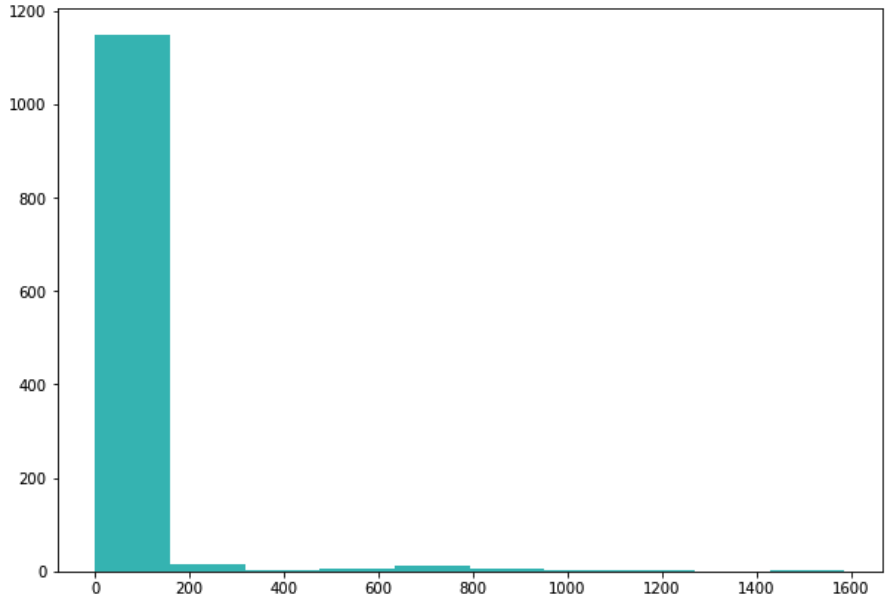
Q2: Can we identify clusters of officers who are likely to be co-accused by factors related to their career, like working environment, salary, rank?

(1) We collect officer data and believe officers who are in the same last unit have a connection. In this case, each officer is a node, and the connection between each other is an edge.

   a. In this network, there are 29894 nodes and 8841868 edges. And its average degree is 591.5480.
   b. We draw a spring layout of this network

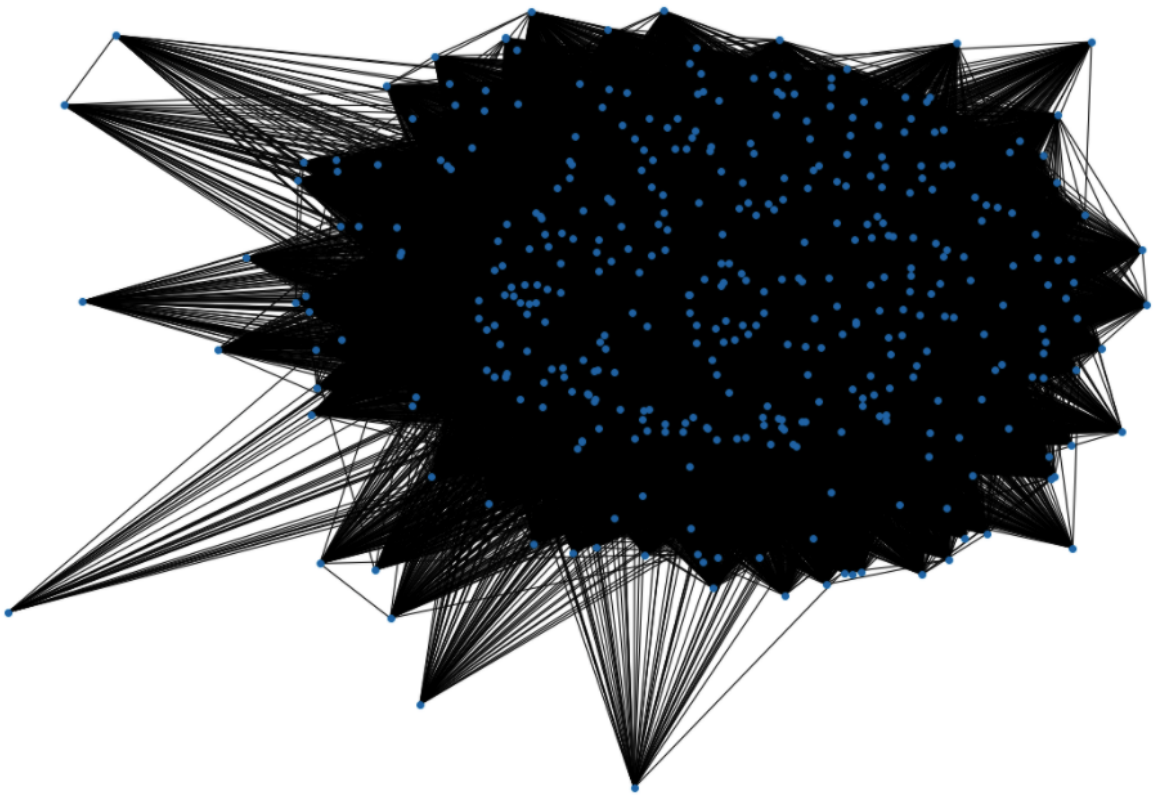


c. We also check the centrality of betweenness of each node, degree of each node, the number of triangles in the network and the clustering of each node. These details can be found in the Jupyter notebook.
d. We draw histogram of degrees:



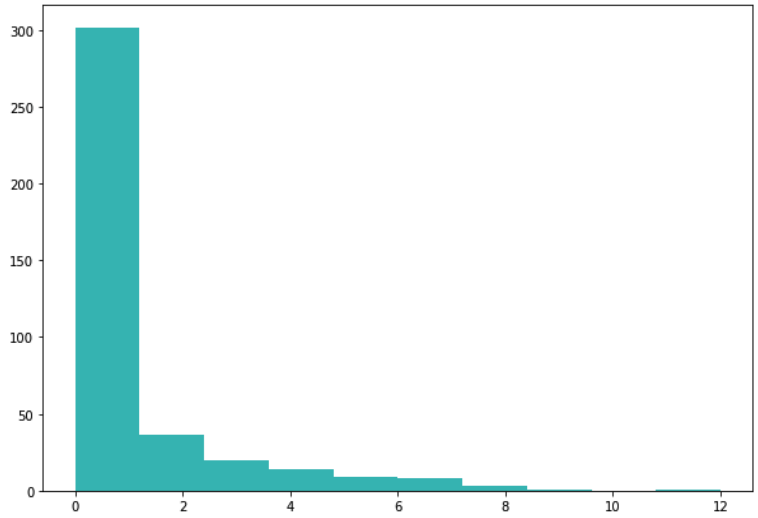
(2) From the clusion above, we collect salary data of officers and the network of salary data. Here we also regard officers who work in the same unit have a connection.

a. In this network, there are 393 nodes and 62343 edges. Its average degree is 317.2672.
b. The spring layout of this network:



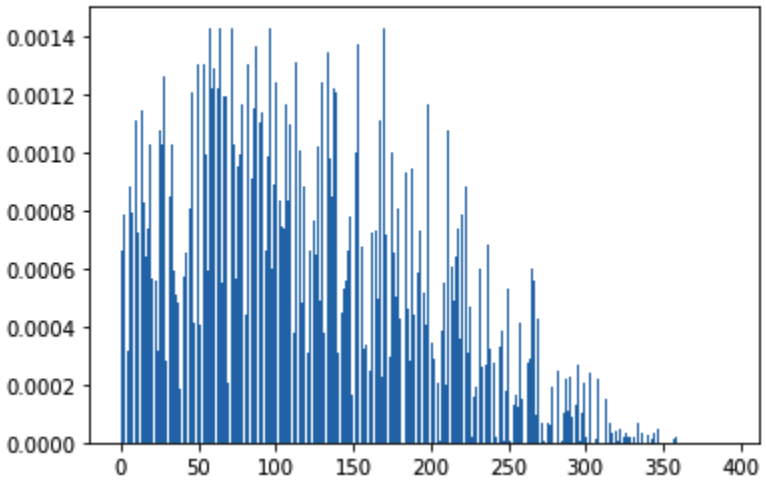
c. We also check the centrality of betweenness of each node, degree of each node, the number of triangles in the network and the clustering of each node. These details can be found in the Jupyter notebook.
d. We draw a histogram of degrees:

```
(array([301.,  36.,  20.,  14.,   9.,   8.,   3.,   1.,   0.,   1.]),
 array([ 0. ,  1.2,  2.4,  3.6,  4.8,  6. ,  7.2,  8.4,  9.6, 10.8, 12. ]),
 <BarContainer object of 10 artists>)
```



e. We draw a bar chart of the betweenness centrality

```
<BarContainer object of 393 artists>
```

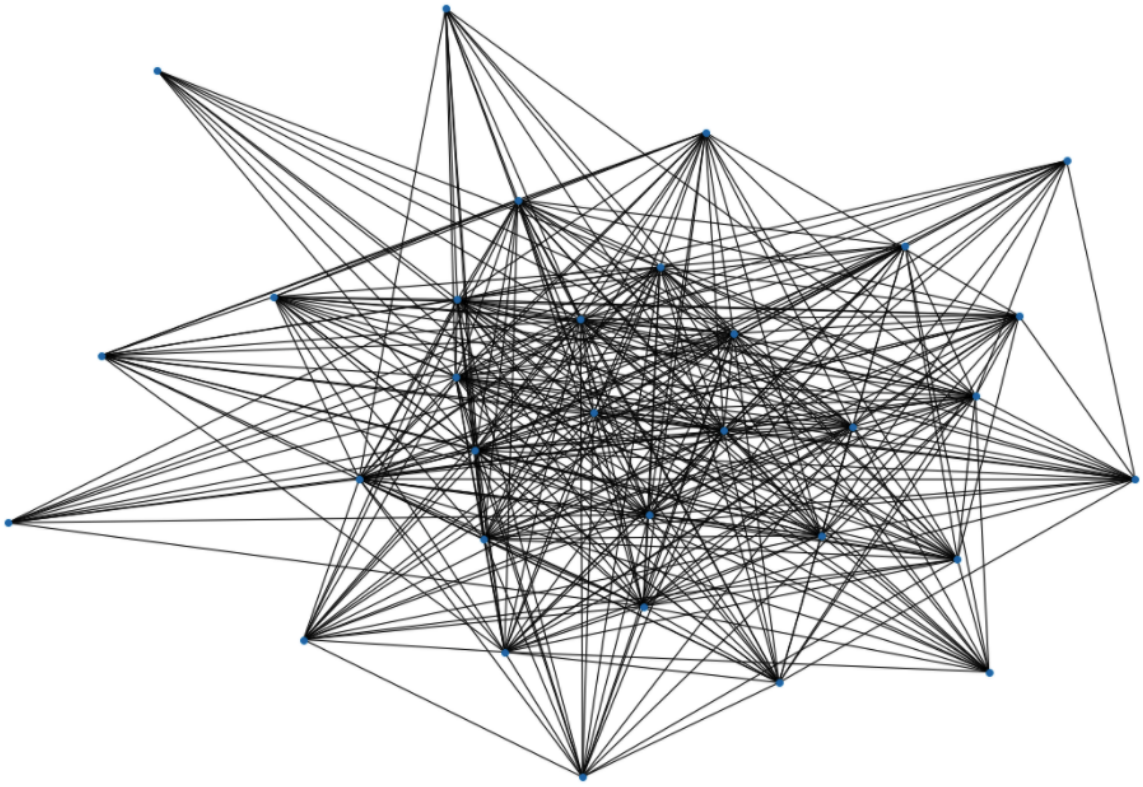


f. Conclusion: there exists cluster of officers regarding their salary

(3) About the rank and award, we also analyze its network in the same method.

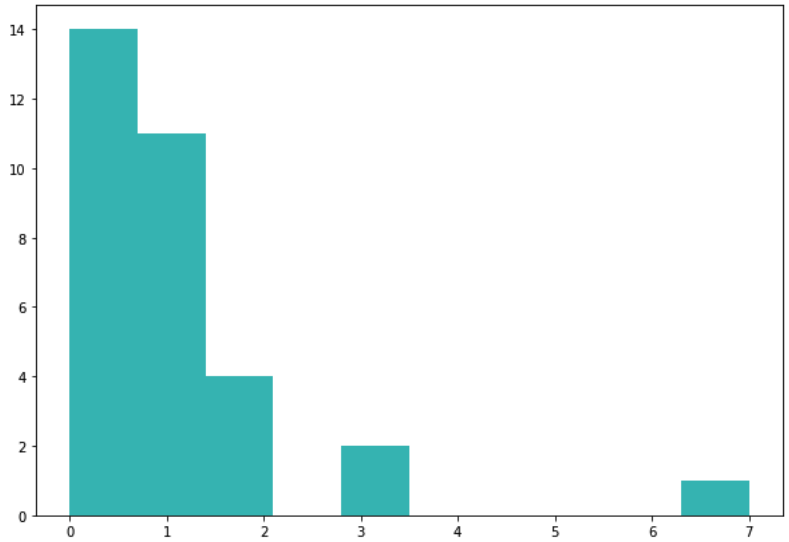
a. In this network, there are 32 nodes, 375 edges and its average degree is 23.4375.

b. We draw a spring layout of this network:

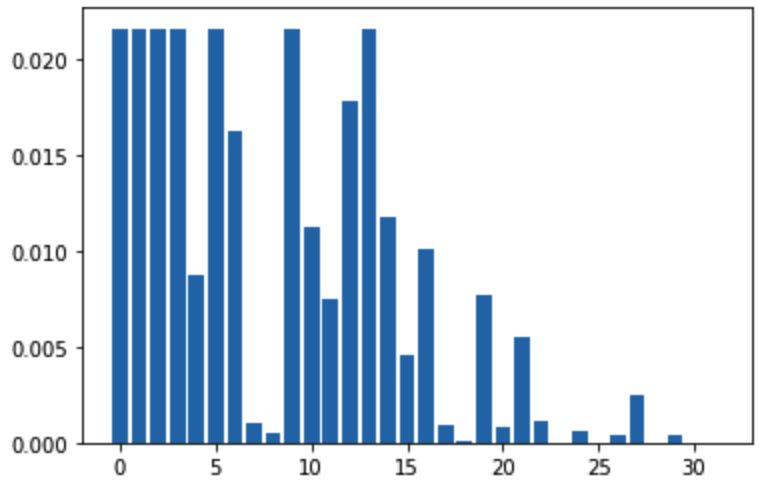


c. We also check the centrality of betweenness of each node, degree of each node, the number of triangles in the network and the clustering of each node. These details can be found in the Jupyter notebook.

d. We draw histogram of degrees:



e. We draw a bar chart of the betweenness centrality:



f. Conclusion: the rank of officers shows high clustering.