In many ways, this checkpoint is the most important to our project. Since we defined our goal as being to figure out **first, whether there is "use of force contagion" between officers that share activities with each other, and second, whether certain groups of officers are more likely to incur civilian allegations than others, regardless of allegation count**, understanding the graph dynamics of CPD is understandably a key resource for us to use. More formally, we concluded that in this checkpoint, our goal was to To construct our graph, we decided to use the data_allegations table to populate our edges and officer names, ids, allegation counts, and civilian allegation percentiles to populate our nodes. Initially, our group had decided to use the activity card table to populate our edges, however we found that it produced a graph much too sparse. Thus, since we were using allegations to construct a network of relationships, we had to modify the ways in which we analyzed our data. As a result, we conducted a few methods of analysis upon our graph and used them all to construct a comprehensive idea of use-of-force contagion within CPD.

First, in pursuit of our question we first decided to perform a triangleCount on the graph. Social Network Theory has suggested that more triangles suggests not only higher connectivity (maybe obviously), but also stronger connections within the network. Thus we then found the correlation between triangleCount and civilian_allegation_percentile. Interestingly, there was nearly no correlation between these two factors. This suggests that both very well connected and very poorly connected officers are likely to receive civilian allegations, and that perhaps contagion (at least across officers that share allegations) is unlikely to occur.

Second, we chose to run a PageRank algorithm on our graph. The PageRank algorithm ranks nodes by popularity according to the graph we have constructed. This output proved very different from our previous one, and quite striking at that. While the mean civilian allegation percentile is around 45%, every single one of the top 20 PageRank outputs was significantly above that margin, with a majority above 90%. However, the correlation between the PageRank and the civilian allegation percentile was much weaker than expected given these other results, with a coefficient of around 0.21. This suggests that only a small amount of the most connected officers have this high civilian allegation percentile, supporting the "few bad apples" theory. This helps to answer both of our questions, but in particular provides very strong evidence that certain, very well connected individuals are much more likely to receive a civilian allegation.

Third, we chose to run a label propagation algorithm on our graph. After this, we displayed the nodes with the highest civilian allegation percentiles and their respective labels. What we found was that most of the top 20 nodes could be attributed to similar nodes, despite thousands of labels being present. This suggests that there are a few subgraphs with very high civilian allegation percentiles wherein contagion occurs. This, combined with the previous analysis, suggests that our second hypothesis is true.

Finally, we chose to run a custom query that found the mean of neighboring nodes' civilian allegation percentiles and sorted them based off of this metric. While this seems like selectively picking which nodes to look at, if the nodes with the highest means are extremely high, that might suggest that there exist many nodes wherein the network that they interact within consists largely of individuals with high civilian allegation percentiles. True to this, all of the twenty nodes with the highest civilian allegation percentiles consisted of mean neighbor percentiles greater than 95%.

Ultimately, all of these methods of analysis paint a pretty specific picture of contagion within CPD: contagion is not wide but rather deep. Instead of affecting many of the nodes evenly, we can trace civilian allegations to small, sequestered communities of individuals that negatively influence each other.