Data Science Seminar - MSAI 339

Checkpoint 4: Graph Analytics

December 2nd, 2021

Professor

Jennie Rogers

Students

Aleksandr Simonyan

Dimitrios Mavrofridis

Donald Baracskay

Introduction

Based on our initial project proposal for the graph analytics part of this project, we investigated the connections of the officers involved in the Tactical Response Report incidents and complaints. For TRRs, we decided to filter on high action response categories, more than five. For complaints, we filtered to only use force complaints. We chose these filters because our larger topic is about use/abuse of force among the police. In order to achieve this, our team decided to operate on the Google Colab platform by implementing pySpark for processing, visualizing, generating, and analyzing the appropriate graphs.

Findings & Results:

For the completion of checkpoint 4, our team decided to identify and analyze the connections between Chicago PD police officers involved in the same tactical responses and complaints by generating and easy to understand graphs. We developed additional queries which allowed us to gather the necessary data and information while also filtering on the action response category for TRRs and use of force for complaints.

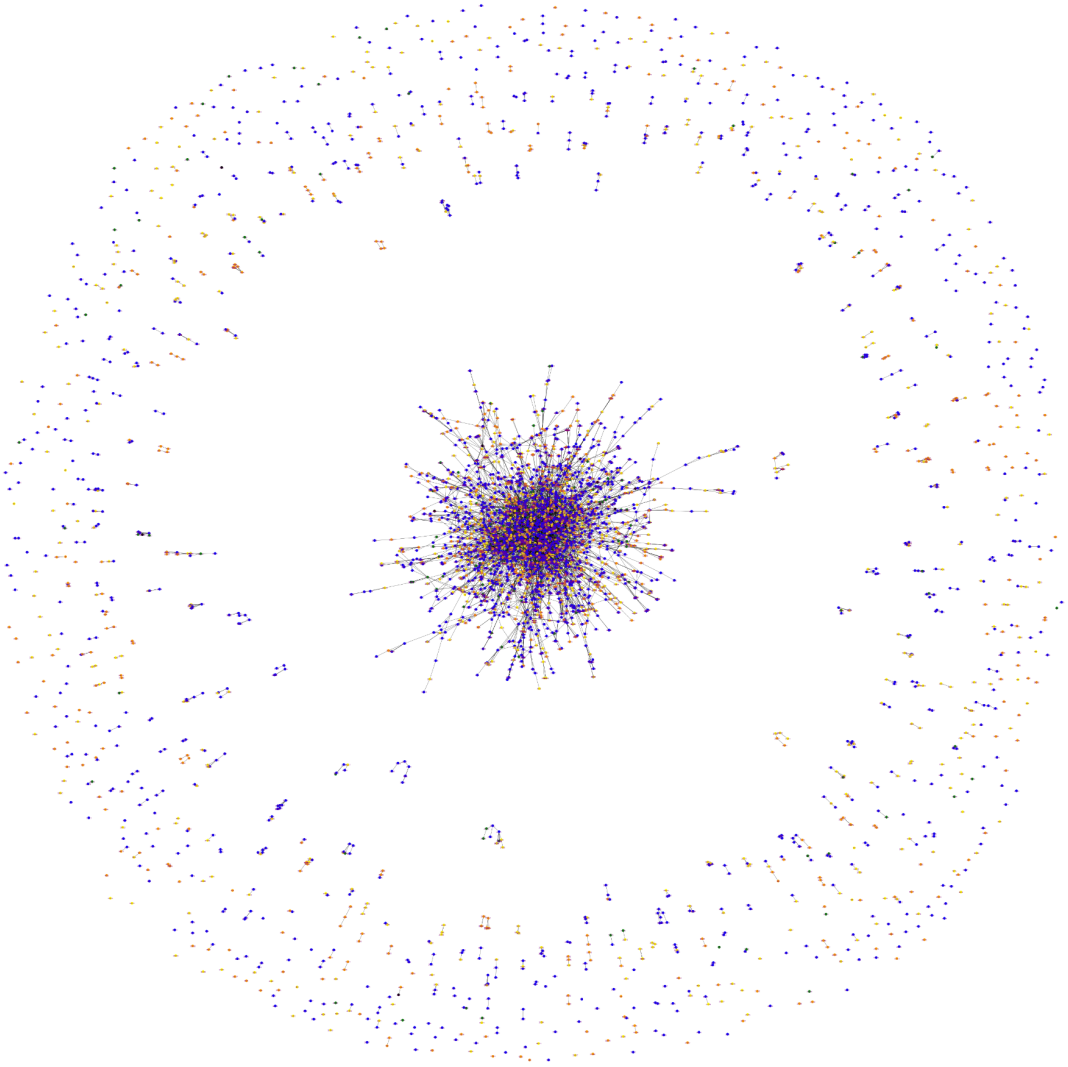

Figure 1: Officers connected by violent TRRs

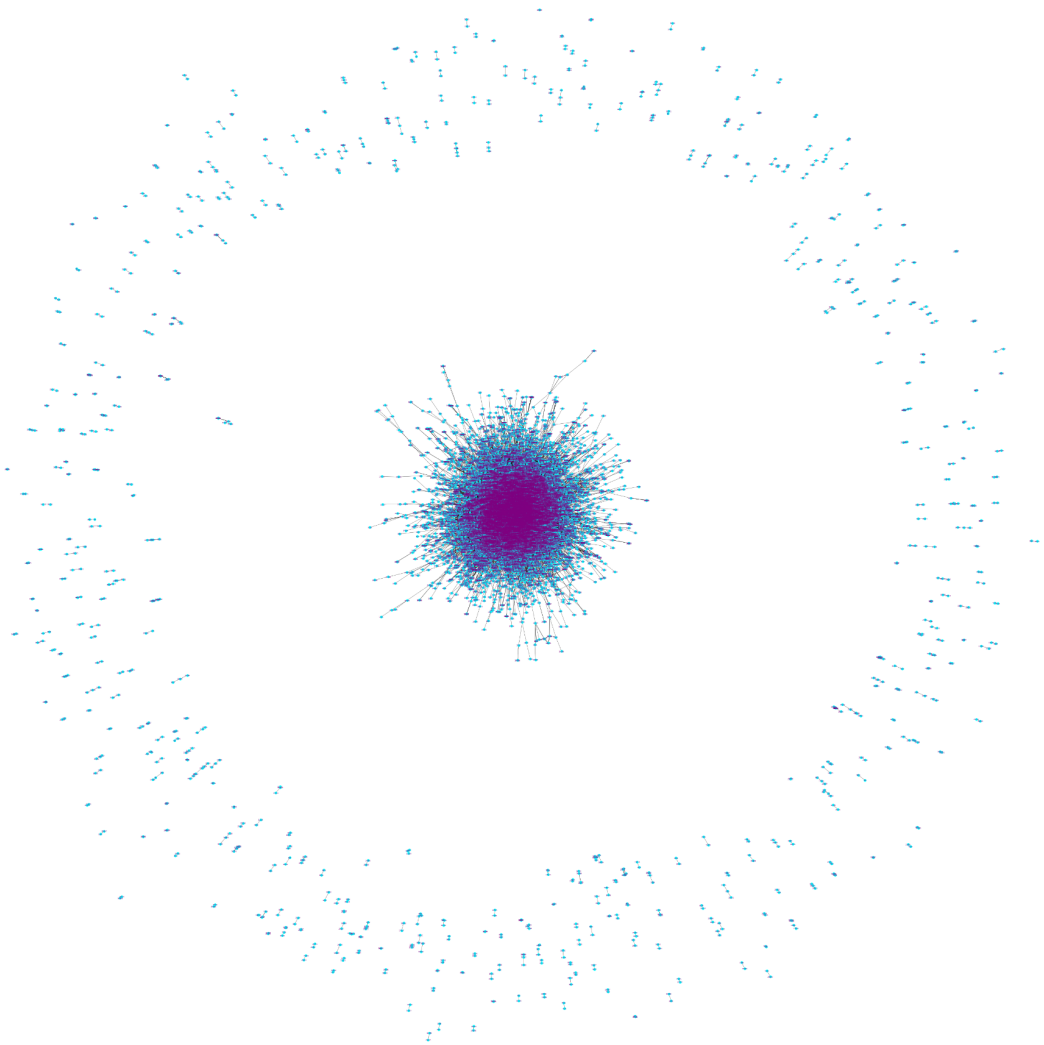

Figure 2: Officers Connected by Use of Force Complaints

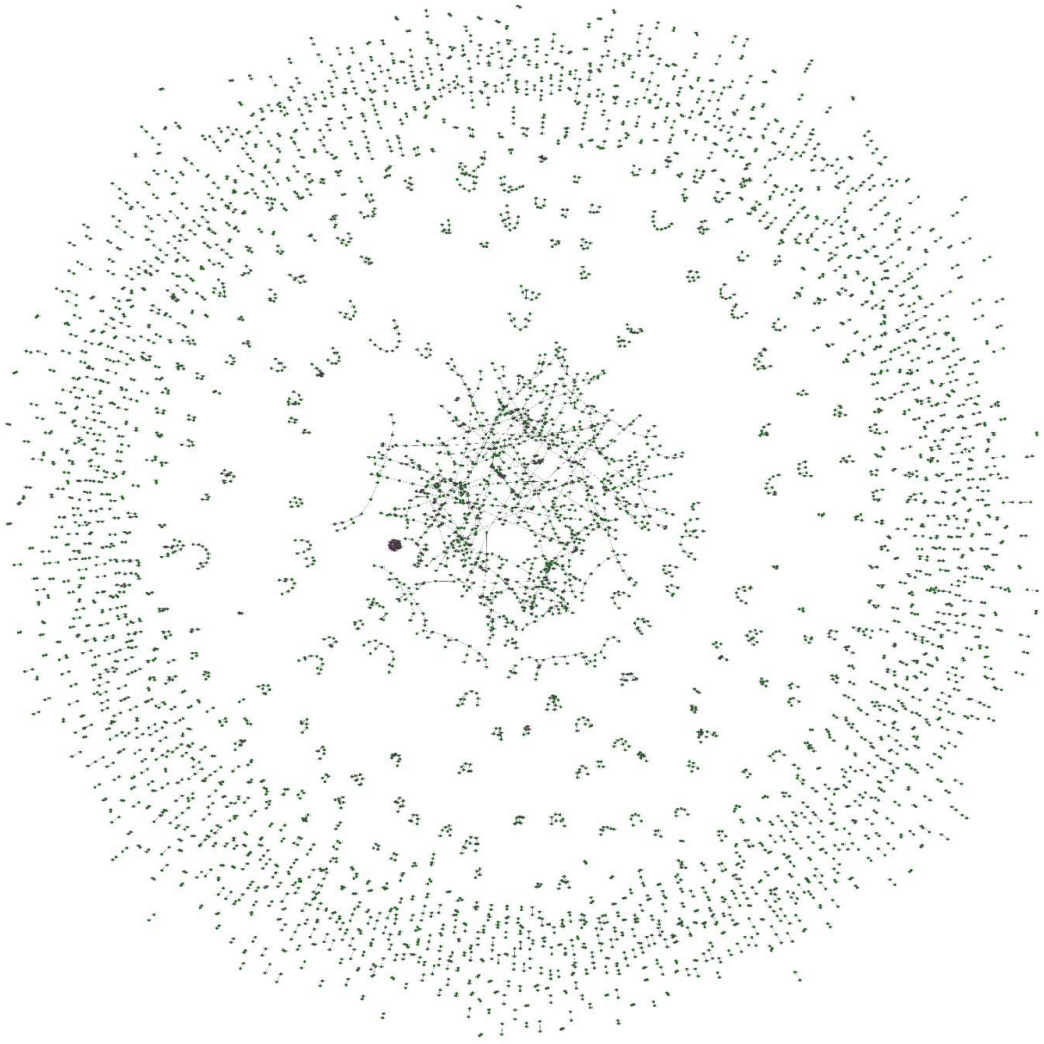

Figure 3: Officer Partner Relationships

TRRs

For TRRs, we first counted the number of officers involved in a TRR, dropping any that had a count of one or had an action response category at or above 5 (5 and up are incidents where the police used violence, eg. fists, guns). We then generated a table of officers involved in these incidents (vertices) and a table of officer to officer connections (edges, basically if two officers are in the same TRR, we create an edge between them) as seen in Figure 1. We also created a partners table, which contained partner relations between officers. We determined partner relations by finding officers who signed in at the same time and beat over 100 times. We figured that officers that performed these actions were partners. We use these partner relationships as a baseline to compare against for both TRRs and use of force allegations, as they are relatively banal meetings that do not involve violence or civilian complaints. We also removed these connections from TRR edges as we wanted to find officers who appear together at the scene of crimes that weren't partners. From all of this, we generated a graph of officers who were at the scene of a violent tactical event with officers who weren't their partner. We then ran page rank and triangle count on this graph.

From page rank, we wanted to find officers who were highly connected to many violent TRRs. For the officers with the top ten TRR page rank scores, we compared their page rank in violent TRRs with their respective page rank in terms of partners. However, we found early on that officers with a high page rank also had very high TRR percentile, and they also had either no partner (based on our partner criteria) or a very low partner page rank score. This likely means that these officers are appearing together at the scene of especially violent police responses, but do not regularly log in with other officers.

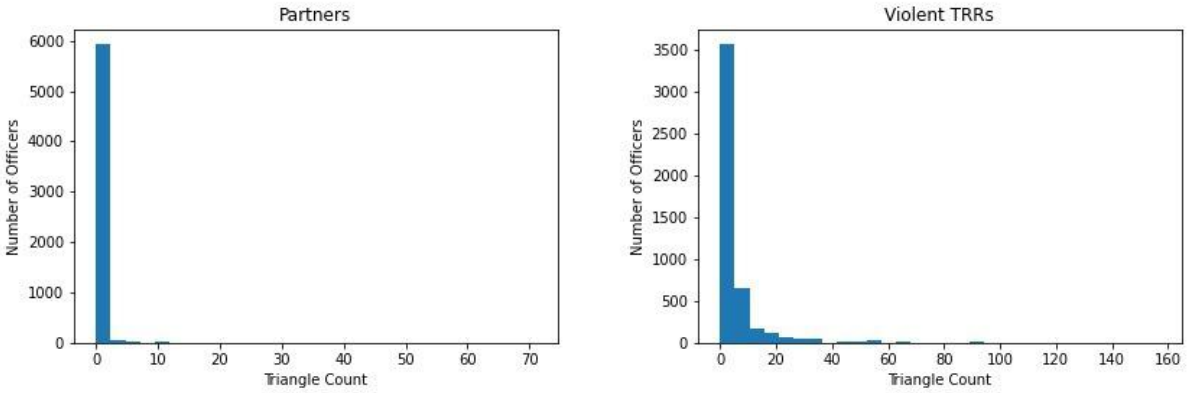

Figure 4: Partner (a) and Violent TRR (b) Triangle Count Histogram

Using triangle count, we attempted to measure the level of clustering in the TRR events and compare to the baseline partner data. We could therefore determine if there were clusters of officers who appear at crimes; this could serve as a more systemic measure than the previous page rank approach. We generated a histogram from both partners and TRR event triangle counts and compared the two as our measure. The partner histogram (seen in Figure 4a) has relatively low deviation with most officers only being linked to a single triangle, if that. A small number of officers have a few more respective triangles and an even smaller number have more than 10 associated triangles. This establishes a trend where most officers only associate with their partners with only a small number clustering beyond this. While the TRR histogram (Figure 4b) is similar in shape, there are a number of deviations from the trend. Notably, the deviation in terms of number of triangles is higher (there is a greater range of triangles). This would seem to indicate that there is a greater level of clustering among the TRR data. Also, there are more outliers, which would point to a subset of officers who are bizarrely more clustered in TRR events than in partners.

Use of Force Complaints

For Use of Force complaints, we obtained the allegations with multiple officers listed as well as the tag "Use of Force." We then performed the same page rank and triangle count as with the previous section.

From page rank, we found officers who were highly connected to others with respect to use of force allegations. These should be officers who exist in a community of officers with many allegations and are therefore repeat offenders and co-accused. High page rank here should indicate not only that they co-offend, but also that those that they offend with also co-offend. The fact that they are so highly connected in this regard may just indicate a high level of connectedness with the rest of their department, so we compared use of force page rank with their partner page rank. When we did this, out of the top 20, very few (~2) actually had what we defined as a partner. This would seem to indicate that they are not well connected to their department when it comes to logging in for work, but are when it comes to allegations.

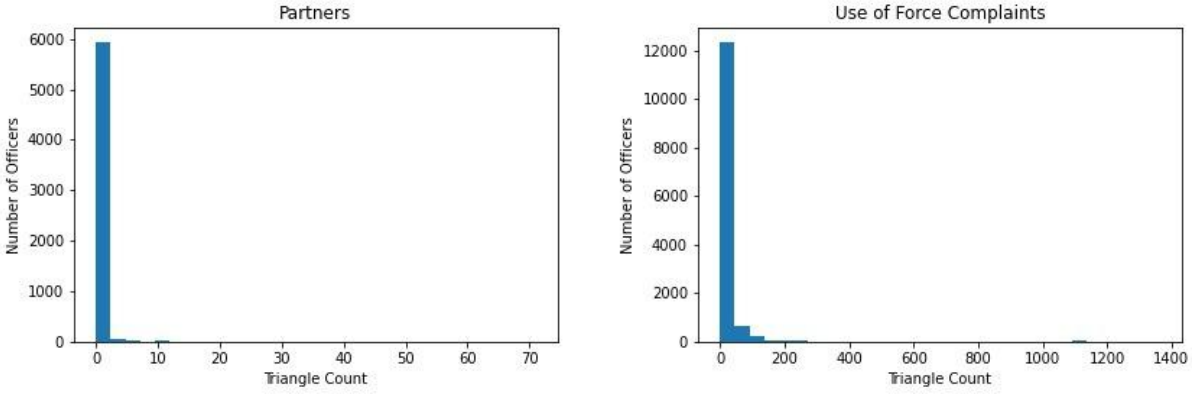

Figure 5: Partner (a) and Use of Force (b) Triangle Count Histogram Comparison

From triangle count, we once again pursued a general metric for how much clustering there is amongst the officers who received use of force complaints (again comparing against the baseline partners triangle count). This is shown in Figure 5. Similar to the TRR triangle count, this one also has a distribution like the partner histogram, but has a much higher deviation and

many outliers. This once again seems to indicate a subset of officers who co-offend as a sort of group.

Conclusion:

From our analysis of TRRs we conclude that there is a subset of officers who do not have partners, but simultaneously become embroiled in violent TRRs regularly. One possibility is that these officers are part of another officer's call for backup. Perhaps they are called upon for their experience or because they are located in areas with a great deal of violent crime. However, another possibility is that these officers either seek out violent situations to become involved with or are actually what insights the violence in these situations.

Our analysis supports that of Papachristos (Charette & Papachristos, 2017), in that there is definitely a community of officers who offend in terms of use of force and they have co-offend together. They have much higher connectedness to other officers implicated in complaints and also rarely have long term partners.

With these facts combined, it would seem there is a community of police officers who are much more involved with violent crimes than the baseline and a community of co-offenders involved in use of force complaints. It seems likely that these two groups would have a fair amount of intersection. This paints a picture of a group of police officers who potentially either seek out or are called to violent situations and then respond in such a way that the subject is compelled to file a use of force report. The fact that they do so is perhaps not notable, but what is more important from our analysis is that they do so as a group, showing perhaps a more systemic issue with the police force.

In addition, our findings allow us to possibly identify police engaging in abuse of force and thus is well within the scope of our project.

Experience with Google Collab & Pyspark:

Our experience working with Pyspark has not been positive. To begin with, we attempted to install Pyspark as a local copy on our portable computers in order to integrate it into Pycharm. However, we faced a number of difficulties especially when it comes to installing it on M1 Macs. Our reasoning behind using Pycharm for the completion of our projects is its ability to support active multi-user code editing. However, after encountering a number of issues and spending a significant amount of time in making Pyspark work locally, we decided to migrate to Google Colab. Although Google Colab offers a familiar interface (Jupiter), it does not allow for active multi-user editing. Thus, our team was unable to work on the same code at the same time. This significantly impacted our efficiency as a group. Nevertheless, after getting familiar with the interface, we were able to successfully write the appropriate code in order to generate and analyze the generated graphs.

Work Cited:

Charette, Y., & Papachristos, A. V. (2017). The network dynamics of co-offending careers. *Social Networks*, *51*, 3–13. https://doi.org/10.1016/j.socnet.2016.12.005