Data Science Seminar - MSAI 339

Checkpoint 4: Graph Analytics

November 18th, 2021

Professor

Jennie Rogers

Students

Aleksandr Simonyan

Dimitrios Mavrofridis

Donald Baracskay

Background and Theme:

By looking at the Tactical Response Reports, we intend to examine the areas and officers involved in violent incidents. We would then hypothesize about the crime in said areas and the officers tangled in such altercations. We will achieve this by filtering the incidents that involve certain actions associated with violence while grouping by location. For example when the police officer's taser or firearm is used against the subject. Additionally, the subjects and officer's physical state will also be taken into consideration along with the number of shots fired during the incident. Moreover, we would like to know whether officers listed in TRRs receive more awards than those who deal with less violent incidents or vice versa. One important aspect that we are going to be covering is understanding the relationship between the number of trr-s and the number of officers in a certain location. We are planning on adding additional dimensions to the data such as race and gender. This will allow us to understand whether there is over policing in certain areas of Chicago and whether the TRRs connected in said areas are likely to be violent in nature.

Introduction

Based on our initial project proposal, for the graph analytics part of this project, we aspired to show the connections of the officers involved in the Tactical Response Report incidents and complaints. For TRRs, we decided to filter on high action action response categories, more than five. For complaints, we filtered to only use force complaints. Finally, we attempted to add additional dimensions to our data, such as race (which we visualized in our graphs). 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.

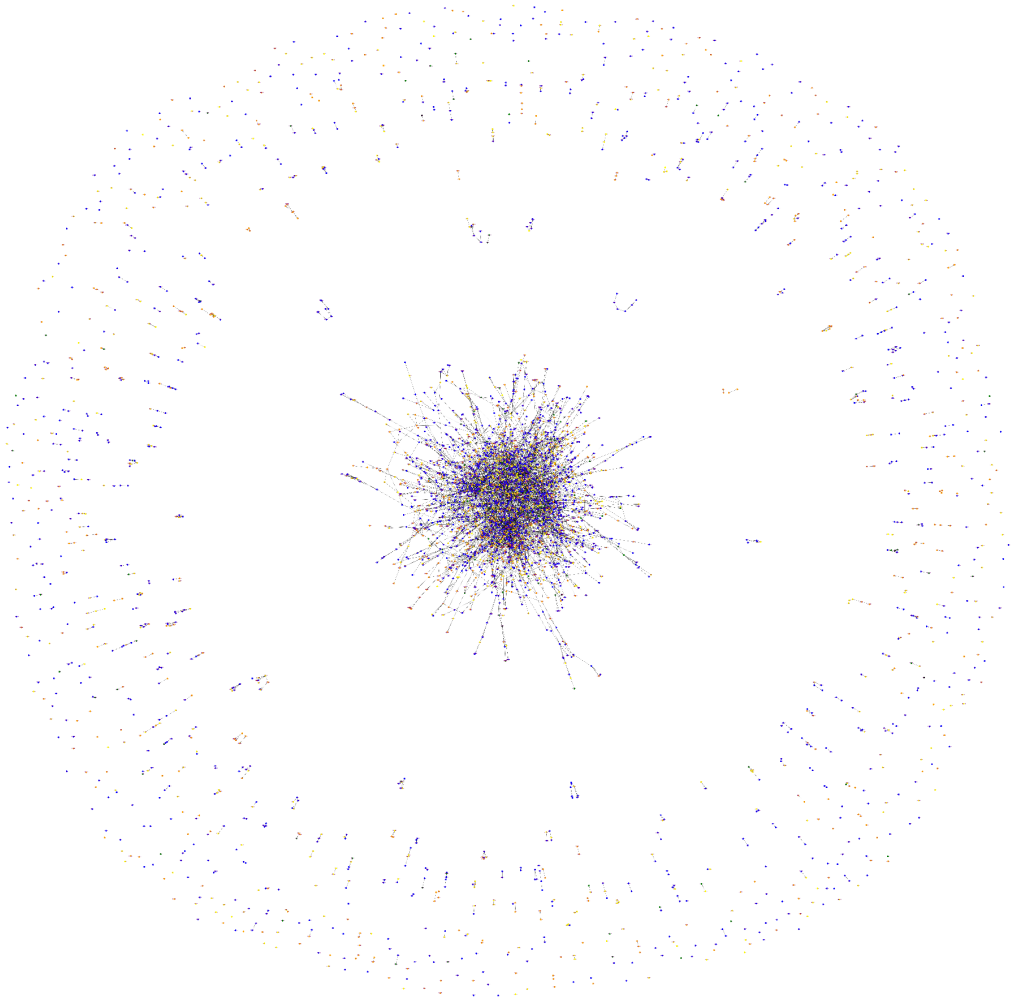

Figure 1: Officers connected by violent TRRs

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 below 5. 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). We also created a partners table. This contained partner relations between officers. We determined this 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 effectively chose to use this as a baseline of relationships to compare against for both TRRs and use of force allegations. 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 to try to determine two things.

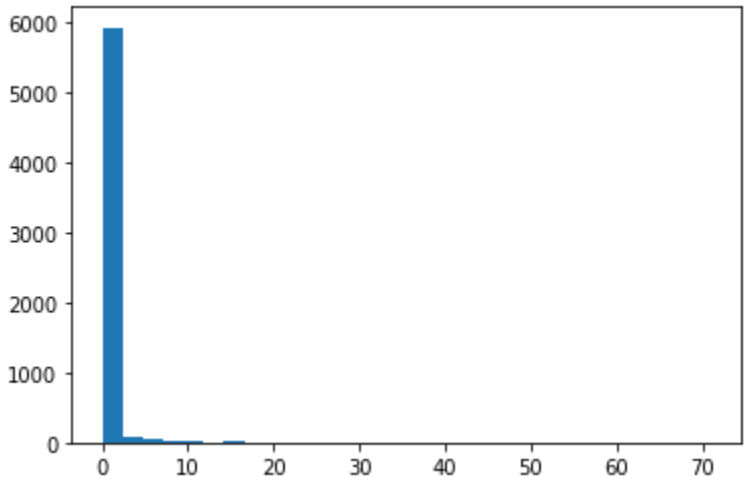

Figure 2: Partner Triangle Count Histogram

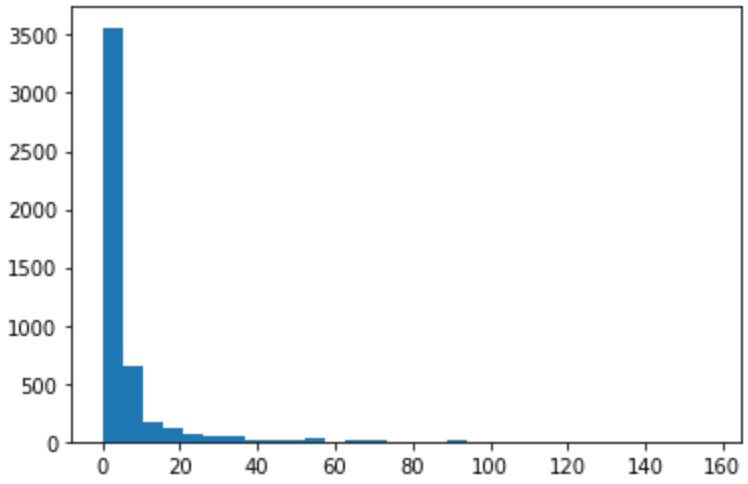

Figure 3: Violent TRR Triangle Count Histogram

From page rank, we wanted to find officers who were highly connected to many events. We compared against the partners page rank as a baseline level of connectedness. 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 crimes, but do not regularly log in with other officers.

From 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, sort of 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 has relatively low deviation with most officers only being linked to a single triangle, if that. A small section of officers have a bit more and an even smaller number have more than 10. While the TRR histogram 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 simply grabbed 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.

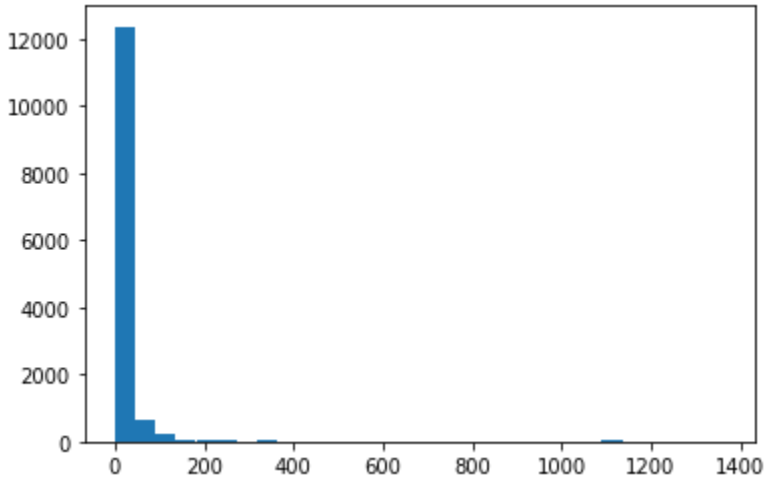

Figure 4: Use of Force Triangle Count Histogram

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 sort of community of officers with many allegations and are therefore repeat offenders and co accused. High page rank here should indicate that 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, however, 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.

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). 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 would have to conclude that there is a subset of officers who do not have partners, but simultaneously become embroiled in violent events 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 actually what insights the violence in these situations.

From our analysis of use of force complaints, we would likely have to agree with Papachristos, 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.

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.