Data Science Seminar: Checkpoint 4 Findings

The Earnest Pirates
Vinit Todai, Shreyas Lele, Tejul Pandit
Checkpoint 4: Graph Analytics

Introduction -

We aim to analyze and evaluate the relationship between the various available features of police officers to assess if they have a significant impact on the trends observed on sustained allegations. The analysis assesses the role played by the rank of a police officer and the unit he/ she belongs to on other officers who are co-accused.

Graph Questions -

1. Amongst officers who belong to a rank that is in the top 10 ranks with the maximum allegation counts, what is the frequency of two officers being co-accused on the same allegation?

    - We filtered the officer data based on the ranks; the officers belong to these ranks as nodes, and the edges represent the relationship between two officers who were co-accused.

2. Amongst officers who belong to a police unit that is in the top 10 ranks with the maximum allegation ratio per officer, what is the frequency of two officers being co-accused on the same allegation?

    - We filtered the officer data based on the unit he belongs to; the officers who belong to these units as nodes, and the edges represent the relationship between two officers who were co-accused.

Results -

1. **Amongst officers who belong to a rank that is in the top 10 ranks with the maximum allegation counts, what is the frequency of two officers being co-accused on the same allegation?**

The subset of ranks selected are as shown in Figure 1. These ranks are obtained by arranging the total number of allegations in descending order and getting top 10 ranks associated.

	rank	total_allegations
1	Police Officer	150977
2	Sergeant of Police	40677
3	Police Officer as Detective	28062
4	Lieutenant of Police	10351
5	Police Officer / Field Training Officer	4466
6	Police Officer Assigned Evidence Technician	2319
7	Captain of Police	2253
8	Commander	1481
9	Police Officer / Mounted Patrol Officer	952
10	Police Officer Assign Youth Off	916

Figure 1(a): Rank of officers used for creating subset

The following figure, Figure 1(b) shows the ranking hierarchies of police officers in Chicago

Title	Insignia
Superintendent of Police	
First Deputy Superintendent of Police	
Chief	
Deputy Chief	
Commander	
Captain	
Lieutenant	
Sergeant	
Field Training Officer	
Police Officer/assigned as: Detective/Youth Officer/Gang Specialist/Police Agent/Major Accident Investigator/ etc	No Insignia
Police Officer	No Insignia

Figure 1(b): Officer Ranking hierarchy

The triangle count defined as the 'count' column in Figure 2(b) shows police officers who are well connected with other police officers. Figure 2(b) shows that officer Raimondo Brown is most connected and co-accused with other officers.

```
tc_cpdb = cpdb2.triangleCount()
tc_cpdb.orderBy('count', ascending=False).show()
tc_cpdb.select("id", "count").show()
```

Figure 2(a): Code for calculating triangle count between nodes

```
+-----+-----+------------------+-----------------+---------------+
|count|   id|              rank|     officer_name|allegation_count|
+-----+-----+------------------+-----------------+---------------+
|29636| 3033| Sergeant of Police|   Raimondo Brown|             17|
|29633| 6315|     Police Officer|    Terence Davis|             38|
|29597| 3744|     Police Officer|   Derek Campbell|              8|
|25678|18042|     Police Officer|     Donald Mc Coy|            22|
|25646|  441|     Police Officer|  Fernando Alonzo|             16|
|22131|21530|     Police Officer|Michael Overstreet|            56|
|21846|27349|     Police Officer|  Charles Stanton|             11|
|21836| 5180| Sergeant of Police|   Stephen Conner|              9|
|21824| 5667|     Police Officer|    Jerry Crawley|             30|
|21814|16747|     Police Officer|    Evetta Lundin|              7|
|21812| 8844|     Police Officer|     Thomas Flynn|             19|
|21809|23654|     Police Officer|       Lloyd Reid|              4|
|21809|14750|     Police Officer|  William Kissane|             23|
|18570|19856|     Police Officer|  Ronald Muhammad|             11|
|17717| 8138|Lieutenant of Police|    Glenn Evans|            132|
|17220|29882|Lieutenant of Police|    Fred Waller|             49|
|17066|28273| Sergeant of Police|     James Taylor|             36|
|17014| 5577| Sergeant of Police|      Michael Cox|             20|
|17012|28459| Sergeant of Police|    Curtis Thomas|             36|
|16982|32074|Police Officer as...|   Ronald Jenkins|             46|
+-----+-----+------------------+-----------------+---------------+
only showing top 20 rows
```

Figure 2(b): Officers placed in the descending order of triangle count

Thus, Figure 2(b) shows that Raimondo Brown, who is Sergeant of Police, is involved in misconduct on a much larger scale with a huge number of other officers, which could potentially imply that he could be misusing his authority.

The page rank gives an idea regarding how likely a police officer is involved in a co-accused allegation. Figure 3(a) highlights that officer Robert Spiegel, who has the highest pagerank, is most influential amongst all the police officers who are co-accused.

Similar to the triangle count, Figure 3(a) shows that it is not necessary that the police officer should be involved in a higher number of allegations for having a higher page rank. Thus, it shows that officer Robert Spiegel is involved in allegations where a higher number of officers are involved.

```
pr_cpdb = cpdb2.pageRank(resetProbability=0.15, tol=0.01)
#look at the pagerank score for every vertex
pr_cpdb.vertices.orderBy('pagerank', ascending=False).show()
pr_cpdb.edges.show()
```

```
+-----+--------------------+------------------+---------------+------------------+
|   id|                rank|      officer_name|allegation_count|          pagerank|
+-----+--------------------+------------------+---------------+------------------+
|32350|      Police Officer|    Robert Spiegel|             20| 59.94255110579391|
|32351|      Police Officer|   Boonserm Srisuth|             25| 58.47588273428055|
|32419|      Police Officer|          Eric Wier|             18| 55.54706144219321|
|32284|      Police Officer|          Mark Reno|             76| 55.05553771973534|
|32384|      Police Officer|      Edwin Utreras|             47| 52.46541790094154|
|32337|Police Officer as...|        Louis Silva|             21| 51.72556067035757|
|32413|      Police Officer|  Carl Weatherspoon|             69|  50.3834356160415|
|32289|      Police Officer|       John Rivera|             44| 49.31297984845296|
|32074|Police Officer as...|     Ronald Jenkins|             46| 47.94275095359477|
|32375|Lieutenant of Police|James Triantafillo|             31| 45.85995467646986|
|32324|      Police Officer|   Kathleen Schmidt|             15| 41.26315879539915|
|32406|      Police Officer|     Reginald Ward|             17|  40.7491314934534|
|32383|      Police Officer|     Armando Ugarte|             47|39.858307108258586|
|31906|      Police Officer|        James Davis|             76| 38.38390375443792|
|32378|      Police Officer|        Mark Trost|             29| 37.91554514385917|
|32355|      Police Officer|    Laurence Stiles|             19|37.352703853575974|
|32341|Police Officer as...|    Robert Slechter|             12| 36.72613447343105|
|32280|      Police Officer|      Eddie Randle|             32| 36.34386614017926|
|32339|      Police Officer|       Keith Sinks|             20| 34.26975822289742|
|32237|      Police Officer|     Louis Ortoneda|             52|33.881065019348554|
+-----+--------------------+------------------+---------------+------------------+
only showing top 20 rows
```

Figure 3(a): Code for calculating pagerank with corresponding tabular representation ordered in descending order of pagerank score

The weight or importance of each edge connecting the source node with the destination node is shown in Figure 3(b).

```
+---+-----+------------+--------------------+
|src|  dst|relationship|              weight|
+---+-----+------------+--------------------+
|  1|  800|           1| 0.14285714285714285|
|  1|10302|           1| 0.14285714285714285|
|  1|16856|           1| 0.14285714285714285|
|  1|24596|           1| 0.14285714285714285|
|  2| 2575|           1|0.045454545454545456|
|  2| 5714|           1|0.045454545454545456|
|  2| 9271|           1|0.045454545454545456|
|  2|11686|           1|0.045454545454545456|
|  2|13744|           1|0.045454545454545456|
|  2|29491|           1|0.045454545454545456|
|  2|32243|           1|0.045454545454545456|
|  2|32402|           1|0.045454545454545456|
|  4| 4380|           1|                0.25|
|  4|13208|           1|                0.25|
|  4|13917|           1|                0.25|
|  4|15095|           2|                0.25|
|  5|  137|           1| 0.02857142857142857|
|  5| 1256|           1| 0.02857142857142857|
|  5| 5044|           1| 0.02857142857142857|
|  5| 5731|           1| 0.02857142857142857|
+---+-----+------------+--------------------+
only showing top 20 rows
```

Figure 3(b): Weight/ importance on each source destination edge

This figure depicts how likely the src (the source police officer id) and the dst (the destination police officer id) are probably going to be involved in an allegation, which is depicted by the weight column.

2. Amongst officers who belong to a police unit that is in the top 10 ranks with the maximum allegation ratio per officer, what is the frequency of two officers being co-accused on the same allegation?

The subset of police units selected are as shown in Figure 4. These police units are obtained by arranging the total number of allegations_ratio_per_unit column which is a ratio of count of total allegations to count of total officers in a police unit in descending order and getting top 10 units associated.

	police_unit	count_allegations	total_officers	allegation_ratio_per_unit
1	97	91	3	30.33
2	173	664	27	24.59
3	175	452	19	23.78
4	113	22	1	22
5	104	22	1	22
6	180	393	19	20.68
7	174	536	26	20.61
8	266	525	26	20.19
9	165	120	6	20
10	156	76	4	19

Figure 4: Police unit data used for creating subset

The triangle count defined as the 'count' column in Figure 5(a) shows that the police officers who are co-accused are involved in allegations which are linked with a large number of officers who are from the same units, or nearby units. Figure 5(a) depicts that officer John Mc Gee from police unit 174 is most well connected with other co-accused officers. Thus, even if the police officer has a lower allegation count against him, he is clustered with a large number of alleged police officers from the same unit.

```
tc_cpdb = cpdb2.triangleCount()
tc_cpdb.orderBy('count', ascending=False).show()
tc_cpdb.select("id", "count").show()
```

```
+-----+-----+------------+------------------+----------------+
|count|   id|last_unit_id|     officer_name|allegation_count|
+-----+-----+------------+------------------+----------------+
|    7|18176|         174|       John Mc Gee|              34|
|    7| 3395|         173|        Adam Burns|              16|
|    7| 1172|         174|   Damen Balesteri|              58|
|    6|30537|         174|   Tracy Whitehead|              19|
|    6|10558|         266|    Anthony Granat|              21|
|    6|27385|         174|      Bill Starling|              29|
|    5|11641|         174|       Craig Hatch|               7|
|    5|12220|         173|      Courtney Hill|              72|
|    4|27439|         173| Robert Stegmiller|              62|
|    4| 2671|         173|      Sean Brandon|              40|
|    4|10663|         174|       Andre Green|              36|
|    4|23600|         173|        Brian Reed|              23|
|    3| 2588|         174|    Jason Bradford|              16|
|    3| 6221|         174|     Herbert Darey|              25|
|    3|  867|         174|Reginald Arrington|              23|
|    3|24324|         174|   Brandon Rodekohr|              16|
|    3|25191|         173|      David Salgado|              31|
|    3| 7982|         173|     Kenneth Epich|              45|
|    3|29613|         173|   Marco Villarreal|              10|
|    2|24521|         175|       Carlos Rojas|              71|
+-----+-----+------------+------------------+----------------+
only showing top 20 rows
```

Figure 5(a): Officers placed in the descending order of triangle count

We can see in the first row that officers from unit id 174 are more likely to be involved in allegations in a group on a much larger scale, followed by unit id 173.

Following figure, Figure 6(a) gives an idea about the pagerank for the officers. It is seen that the pagerank is similar for the officers immaterial of the allegation_count.

```
pr_cpdb = cpdb2.pageRank(resetProbability=0.15, tol=0.01)
#look at the pagerank score for every vertex
pr_cpdb.vertices.orderBy('pagerank', ascending=False).show()
pr_cpdb.edges.show()
```

```
+-----+------------+----------------+----------------+------------------+
|   id|last_unit_id|    officer_name|allegation_count|          pagerank|
+-----+------------+----------------+----------------+------------------+
|31996|         175|    German Gomez|              12| 1.305863820393211|
|25191|         173|   David Salgado|              31|1.1865471624739725|
|27935|         165|   Matthew Swain|              61|1.1847111268651542|
|32406|         173|   Reginald Ward|              17|1.1645157563285098|
|24521|         175|    Carlos Rojas|              71|1.1580852338143208|
|27439|         173|Robert Stegmiller|             62|1.1577471014734335|
|30320|         180| Robert Weisskopf|             23|1.1524574235387521|
|29445|         173|        Luis Vega|             53|1.1440847317012035|
|30537|         174|  Tracy Whitehead|             19|1.1299951650932252|
|23600|         173|       Brian Reed|             23| 1.121360147183769|
|30387|         173| Mark Wesselhoff|              12| 1.119362425837314|
|18176|         174|      John Mc Gee|             34|1.1001864159642845|
|32310|         174|    Bradley Ruzak|             26|1.0931619943906823|
|28335|         180|   Jesse Terrazas|             60| 1.069512583185627|
|32235|         174|    Kevin Oneill|              16|1.0671639477854686|
|28734|         175|    Daniel Torres|              9|1.0658958202979916|
|27385|         174|    Bill Starling|             29|1.0620978011392457|
|10558|         266|  Anthony Granat|              21| 1.058339206340742|
|19834|         174|     Todd Mueller|             27| 1.057729659160206|
|32400|         175|    Adam Wallace|              16|1.0402479378562848|
+-----+------------+----------------+----------------+------------------+
only showing top 20 rows
```

Figure 6(a) : Code for calculating pagerank with corresponding tabular representation ordered in descending order of pagerank score

```
+----+-----+------------+-------------------+
| src|  dst|relationship|             weight|
+----+-----+------------+-------------------+
|1825|18588|           1|               0.04|
|3395| 4936|           1|0.021739130434782608|
| 672|17251|           1|0.045454545454545456|
|1932|13598|           1|0.004504504504504...|
|1932|27686|           1|0.004504504504504...|
|2671|12932|           1|0.014705882352941176|
|3395| 5131|           1|0.021739130434782608|
|3395|23913|           1|0.021739130434782608|
|6451|32287|           1|0.012345679012345678|
| 867|22423|           2|0.030303030303030304|
|1932|31176|           1|0.004504504504504...|
|5571|18462|           1|0.022727272727272728|
| 672| 1177|           1|0.045454545454545456|
| 847|18266|           1| 0.07142857142857142|
|1932| 4993|           1|0.004504504504504...|
|1932|12247|           1|0.004504504504504...|
|1932|24716|           1|0.004504504504504...|
|5406|24596|           2|0.016666666666666666|
|5494| 6009|           1| 0.07142857142857142|
|3743|27229|           2| 0.02564102564102564|
+----+-----+------------+-------------------+
only showing top 20 rows
```

Figure 6(b): Weight/ importance on each source destination edge

This figure 6(b) depicts how likely the src (the source police officer id) and the dst (the destination police officer id) are probably going to be involved in an allegation, which is depicted by the weight column.

Conclusion -

Graph analytics helps in identifying the police officers who are most influential and are most well connected with other police officers. This can provide an early insight into the behaviour of an officer and how likely they are to influence other police officers.