

Checkpoint 4: Graph Analytics

In alignment with our group's theme to find out what drives an allegation against a police officer to be sustained, our group analyzed the relationships between the police officers in the most predominant categories of allegations. We analyzed the rank of Police officers for our analysis as they are the ones with maximum number of positions in the police department and have the greatest number of allegations and in turn sustained allegations against them. We used the different algorithms provided by Graph frames in Pyspark and found some interesting findings.

- 1) Amongst officers who have a rank of 'Police Officer' in the police department, what is the frequency of two officers being co-accused in the same allegation category 'Operation and Personnel Violations' and having their allegations sustained?
 - We filtered the officer data based on their ranks and whether their allegations had been sustained or not. We chose the rank of police officers for analysis as they have the highest representation in the police department. Then we filtered on the allegation category of 'Operation and Personnel Violations' which was the category with one of the highest number of sustained allegations resulting in severe punishments. The allegation category was determined by our analysis in the previous checkpoints. The officers who belonged to these ranks are taken as nodes, and the edges represents the relationship between two officers who were a part of a crime together and hence had a common allegation id.
- 2) Amongst officers who have a rank of 'Police Officer' in the police department, what is the frequency of two officers being co-accused in

the same allegation category 'Conduct Unbecoming (Off-Duty)' and having their allegations sustained?

- We filtered the officer data based on their ranks and whether their allegations had been sustained or not. We chose the rank of police officers for analysis as they have the highest representation in the police department. Then we filtered on the allegation category of 'Conduct Unbecoming (Off-Duty)' which was the category with one of the highest number of sustained allegations resulting in severe punishments. The allegation category was determined by our analysis in the previous checkpoints. The officers who belonged to these ranks are taken as nodes, and the edges represents the relationship between two officers who were a part of a crime together and hence had a common allegation id.

Results :

- 1) Amongst officers who have a rank of 'Police Officer' in the police department, what is the frequency of two officers being co-accused in the same allegation category 'Operation and Personnel Violations' and having their allegations sustained?

Vertices and Edges for the graph:

id	officer_name	sustained_count
6	Anthony Abbate	2
17	Moulay Abdullah	1
18	Jason Abejero	1
21	Deborah Abelson	3
29	Henry Abrams	1
47	Rafael Acevedo	1
50	Nadine Acevedo	2
54	Jeff Acevedo	1
58	Manuel Acevedo	5
59	Edward Acevedo	2
75	Mario Acosta	1
79	Erwin Acox	1
80	Zaireh Acsvecs	1
89	Randy Adamovitz	1
90	Craig Adams	3
97	Wilbur Adams	1
100	Marvin Adams	2
105	Christoph Adams	2
108	Vicky Adams	2
110	Lloyd Adams	3

Figure 1(a): Vertices of the graph consisting of officer_id, officer name, count of sustained allegations

src	dst	relationship
27740	30737	3
309	6075	2
9491	31328	2
28278	28999	2
19501	31818	2
2400	15354	2
8156	18223	2
22906	30108	2
13594	13788	2
6084	19477	2
7829	28560	2
15195	24973	2
9517	27740	2
30737	32294	2
9517	30737	2
11731	18042	2
9517	32294	2
12160	15760	2
30849	31248	2
5829	20792	2

Figure 1(b): Edges of the graph consisting of src(officer_id 1), dst(officer_id 2), relationship (count of allegation_category_id)

Implemented Graph Algorithms:

1. Triangle Count: The triangle count defined as the 'count' column in Figure 2 shows that many police officers are involved in a single allegation. Sometimes people with less allegation count also have higher triangle count which shows that these officers have been involved in less crimes overall but in crimes with very high number of police officers involved in each crime. Below image shows the list of Police officers who are involved with many other police officers in the allegation category of 'Operation and Personnel Violations' who have sustained allegations against them.

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

id	officer_name	count
12788	William Humphrey	3723
6387	Percy Davis	3636
5667	Jerry Crawley	3573
6260	James David	3570
10986	Vincent Guiffra	3570
258	Robert Alanis	3570
13401	Johnny Jenkins	3570
4699	Joseph Cistaro	3570
13794	Benjamin Jones	3570
8134	Roy Evans	3570
11537	Andre Harris	3570
10227	James Golich	3570
9555	Wilman Garcia	3570
9713	Gary Gaski	3570
944	Raymond Augle	3570
5756	Lenwood Cross	3570
923	Jesse Atilano	3570
911	Tyrone Askins	3570
13732	Lorenzo Jones	3570
10556	Jack Granadon	3570

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

2. Page Rank: Page rank like the triangle count gives an idea of how much involved a police officer is in the crimes. Police officers with higher page rank values are involved in either more sustained allegations against

them or are involved in sustained allegations where there are a greater number of people involved per allegation. Every person involved in a sustained allegation would link to every other person involved in it. Below image shows the list of Police officers who are involved with many other police officers in the allegation category of 'Operation and Personnel Violations' who have sustained allegations against them.

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

id	officer_name	sustained_count	pagerank
32192	Dereck Miller	2	40.97297768180862
31756	Jose Zuniga	15	23.550374153558035
30342	Christoph Wells	1	14.710267738714268
30663	Ronald Williams	2	11.814155914780049
29807	Richard Walker	1	11.462546289907223
30294	Daniel Weems	1	10.586822550956482
31648	Andres Zepeda	1	10.057465668696398
29046	Isaiah Turner	2	9.457956097235726
28556	Robert Thorpe	3	8.084936168448614
27708	Antionett Strong	1	7.081695913969587
30223	Gerald Waxmonsky	3	7.032093966103569
26851	Ronald Smith	3	6.932595904355005
27349	Charles Stanton	2	6.318782167932732
30653	Alan Wilks	1	6.021661743717718
29063	Andrew Turner	2	5.45884531205926
26774	Curtis Smith	1	5.10349246278389
24653	Edward Rosales	3	4.82875030915085
32384	Edwin Utreras	3	4.823832321495567
26674	Stephen Smith	2	4.7040342337817265
32291	Victor Rivera	1	4.6921344697321405

Figure 3: List of police officers in descending order of PageRank score

3. Degrees: The degree of a node shows the number of edges it has incoming on it and outgoing from it. In our analysis, the degree of a node shows the number of police officers that one police officer is connected to in the category of 'Operations and Personnel Violation' which resulted in a sustained allegation. Below is a list of police officers with the highest degrees.

id	degree
12788	103
6387	97
9555	89
5667	88
31756	88
18042	87
23534	87
26851	87
10389	86
2604	86
13108	86
6534	86
24185	86
21442	86
13732	86
10556	86
13366	85
13794	85
13401	85
8057	85

Figure 4: List of police officers in descending order of degree

4. Label Propagation: It is the only algorithm in graph frames which gives us communities in the graph. In our analysis, we use label propagation to find the number of communities that are formed of police officers who have sustained allegations in the department of 'Operation and Personnel Violation'. Figure 4(a) tells us that there are 4041 communities in our sample graph and gives us the labels of the communities that an officer belongs to. Figure 4(b) gives us a list of all the communities in our sample graph with the highest number of members in it.

```

communities = cpdb.labelPropagation(maxIter=40)
communities.persist().show(10)
print (f"There are {communities.select('label').distinct().count()} communities in this sample graph.")

```

```

+-----+-----+-----+-----+
|  id|      officer_name|sustained_count|label|
+-----+-----+-----+-----+
|  29|      Henry Abrams|          1|29807|
| 2214|      Calvin Blunt|          3|23072|
| 5556|    Ramon Covington|          1| 5556|
|11190|    Mahmoud Haleem|          1|11190|
|13401|    Johnny Jenkins|          8|29807|
|19979|    Dennis Murdock|          1|22809|
|28242|      Andre Taylor|          1|28242|
|29824|      Myrta Walker|          1|29824|
|31156|      Emmitt Womack|          1|31156|
|31703|Bernice Ziolkowski|          2|31703|
+-----+-----+-----+-----+
only showing top 10 rows

```

There are 4041 communities in this sample graph.

Figure 4(a): List of police officers with their corresponding labels

```

communities.groupby('label').count().sort('count', ascending=False).show()

```

```

+-----+-----+
|label|count|
+-----+-----+
|29807|  113|
|22809|   25|
| 5439|   19|
|30294|   17|
|11573|   11|
|29403|    9|
|21986|    9|
| 5520|    8|
|20820|    8|
|24247|    8|
| 8481|    8|
|12745|    7|
|30518|    6|
|15936|    6|
|21436|    6|
| 9001|    6|
|23631|    6|
|28153|    6|
|21777|    6|
|31619|    5|
+-----+-----+

```

Figure 4(b): List of communities in descending order of their count of members

- 2) Amongst officers who have a rank of 'Police Officer' in the police department, what is the frequency of two officers being co-accused in the same allegation category 'Conduct Unbecoming (Off-Duty)' and having their allegations sustained?

Vertices and Edges for the graph:

id	officer_name	sustained_count
21	Deborah Abelson	1
29	Henry Abrams	1
45	Pedro Acevedo	1
58	Manuel Acevedo	2
89	Randy Adamovitz	1
117	James Adams	1
160	Mahir Affaneh	1
191	Edward Aguilera	2
199	Kathleen Aguinaga	2
220	Terry Ahlgrim	1
225	Maurice Ahmed	1
241	Willie Akerson	1
245	Jimmie Akins	1
258	Robert Alanis	1
344	Joel Algarin	1
376	Gloria Allen	1
387	Raymond Allen	1
396	Michael Allen	1
420	Xavier Almaguer	1
422	Joseph Almanza	1

Figure 5(a): Vertices of the graph consisting of officer_id, officer name, count of sustained allegations

src	dst	relationship
985	17287	2
25262	28807	2
191	22319	1
191	29285	1
860	1938	1
923	15982	1
947	15308	1
974	985	1
974	2085	1
974	2323	1
974	2978	1
974	3027	1
974	4249	1
974	4408	1
974	6155	1
974	6250	1
974	6857	1
974	8602	1
974	8658	1
974	8992	1

Figure 5(b): Edges of the graph consisting of src(officer_id 1), dst(officer_id 2), relationship (count of allegation_category_id)

Implemented Graph Algorithms:

1. Triangle Count: The triangle count defined as the 'count' column in Figure 2 shows that many police officers are involved in a single allegation. Sometimes people with less allegation count also have higher triangle count which shows that these officers have been involved in less crimes overall but in crimes with very high number of police officers involved in each crime. Below image shows the list of Police officers who are involved with many other police officers in the allegation category of 'Conduct Unbecoming (Off-Duty)' who have sustained allegations against them.

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

```

+-----+-----+-----+
|  id|      officer_name|count|
+-----+-----+-----+
|  985|      Joseph Avila| 1080|
| 17287|    Charles Maratre| 1080|
|   974|    Bernard Avant| 1038|
| 10902|    David Grubisic| 1035|
| 31906|      James Davis| 1035|
|  2323|      Craig Bolin| 1035|
| 10352| Enriquez Gonzalez| 1035|
|  3027|      David Brown| 1035|
|  8602|      Lolita Fisher| 1035|
| 13207|      Rufus James| 1035|
|  4408| Nicolas Chapello| 1035|
| 13401|    Johnny Jenkins| 1035|
| 19052|        Che Miles| 1035|
| 15668|    Darryl Langston| 1035|
|  8992|      Henry Foster| 1035|
| 11413|    Wadell Hardy| 1035|
|  8658|      Corey Flagg| 1035|
|  2085|      Eural Black| 1035|
| 22172|      Aaron Pena| 1035|
|  6155|      Fred Dammons| 1035|
+-----+-----+-----+

```

Figure 6: Officers placed in the descending order of triangle count

2. Page Rank: Page rank like the triangle count gives an idea of how much involved a police officer is in the crimes. Police officers with higher page rank values are involved in either more sustained allegations against them or are involved in sustained allegations where there are a greater number of people involved per allegation. Every person involved in a sustained allegation would link to every other person involved in it. Below image shows the list of Police officers who are involved with many other police officers in the allegation category 'Conduct Unbecoming (Off-Duty)' who have sustained allegations against them.

```

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

```

id	officer_name	sustained_count	pagerank
32323	Jerry Scaife	1	24.030912366209257
31906	James Davis	1	13.009611474979001
30921	Leigha Wilson	1	9.129551912265967
29817	Herbert Walker	2	7.122709294673093
26497	Lonnie Singleton	1	5.876637866365711
31275	Ernest Wright	1	5.7050918902945815
26167	Teresa Shegog	3	5.022767407150179
23936	Steven Ridgner	2	4.405013497829839
23859	Janice Richard Ka...	2	3.9307799866696236
23265	Robert Quintero	2	3.55450967199834
22172	Aaron Pena	3	3.248224962353675
27911	Rennard Sutton	1	3.0996310600865105
21890	Sammie Parr	4	2.993755725671591
29285	June Valle	1	2.9742098451045087
21860	Robbin Parker	1	2.782190925539482
20286	Andrew Neal	1	2.6002319039392057
17287	Charles Maratre	2	2.511675621652257
20078	Herbert Murphy	1	2.44197254064013
26951	Stephen Sodergren	1	2.3457785183116733
24473	Lloyd Roe	1	2.3175661130684486

Figure 7: List of police officers in descending order of PageRank score

3. Degrees: The degree of a node shows the number of edges it has incoming on it and outgoing from it. In our analysis, the degree of a node shows the number of police officers that one police officer is connected to in the category of 'Conduct Unbecoming (Off-Duty)' which resulted in a sustained allegation. Below is a list of police officers with the highest degrees.

id	degree
17287	55
985	55
974	49
26167	46
9617	46
15668	46
22172	46
19052	46
19856	46
11184	46
31906	46
29817	46
6250	46
16167	46
20078	46
19684	46
18059	46
16875	46
14313	46
8992	46

Figure 8: List of police officers in descending order of degree

4. Label Propagation: It is the only algorithm in graph frames which gives us communities in the graph. In our analysis, we use label propagation to find the number of communities that are formed of police officers who have sustained allegations in the department of 'Conduct Unbecoming (Off-Duty)'. Figure 9(a) tells us that there are 4041 communities in our sample graph and gives us the labels of the communities that an officer belongs to. Figure 9(b) gives us a list of all the communities in our sample graph with the highest number of members in it.

```

communities = cpdb.labelPropagation(maxIter=40)
communities.persist().show(10)
print (f"There are {communities.select('label').distinct().count()} communities in this sample graph.")

```

```

+-----+-----+-----+-----+
| id| officer_name|sustained_count|label|
+-----+-----+-----+-----+
| 29| Henry Abrams| 1| 29|
| 2214| Calvin Blunt| 1| 2214|
| 2453| Joseph Boston| 2| 2453|
| 3764| Sean Campbell| 1| 3764|
| 13401| Johnny Jenkins| 4| 22172|
| 191| Edward Aguilera| 2| 191|
| 1360| Donna Barnes Simmons| 1| 1360|
| 3280| Andre Bullock| 2| 3280|
| 12967| Sonia Irwin| 1| 12967|
| 22102| John Pechulis| 1| 22102|
+-----+-----+-----+-----+
only showing top 10 rows

```

There are 1246 communities in this sample graph.

Figure 9(a): List of police officers with their corresponding labels

```

#only shows the top 5 rows
communities.groupby('label').count().sort('count', ascending=False).show()

```

```

+-----+-----+
|label|count|
+-----+-----+
| 22172| 47|
| 21197| 9|
| 191| 4|
| 2400| 3|
| 24473| 3|
| 26533| 3|
| 6393| 3|
| 11229| 3|
| 9688| 3|
| 2978| 3|
| 22365| 3|
| 29| 1|
| 1360| 1|
| 2214| 1|
| 3280| 1|
| 25946| 1|
| 2453| 1|
| 12967| 1|
| 8330| 1|
| 3764| 1|
+-----+-----+

```

Figure 9(b): List of communities in descending order of their count of members

Summary: After using the various algorithms provided by graph frames, we came to many interesting observations.

1. Higher triangle count for a police officer does not ensure that he/she has a higher allegation count. It suggests that the police officer was involved in those sustained allegations, where there were many other police officers involved in each allegation.
2. Higher PageRank for a police officer suggests that a police officer has had influence on the other police officers. Higher PageRank suggests that the police officer would influence or be involved in a certain way in an allegation that was sustained for the two allegation categories we analyzed.
3. There are 4041 communities formed of police officers where they are part of an allegation which was sustained, and the allegation category is 'Operation and Personnel Violations'.
4. There are 1246 communities formed of police officers where they are part of an allegation which was sustained, and the allegation category is 'Conduct Unbecoming (Off-Duty)'.