



Data Science Seminar - MSAI 339

Checkpoint 4

November 18th, 2021

Professor
Jennie Rogers

Students
Anery Patel
Chukwueloka Obi
Ana Cheyre

Graph Analytics

- 1. Network analysis: Is there a connection between investigators and certain officers or districts who are involved in incidents that have unsustained complaints using connected components to test if there is a correlation between the two.**

Using graph analytics, we discovered how relationships between the investigators and officers involved in the same allegation report could have influenced disciplinary actions for the accused officer. The investigator mentioned in an allegation report is responsible for investigating the misconduct by a police officer accused in that allegation. With regard to our theme, we assumed that investigators could be giving a leeway to the culprit officers thus leading to an unsustained complaint as the outcome. Our analysis enforces this assumption in multiple ways.

Representing tabular data in a Graph:

Vertices:

We have two types of vertices. Investigator vertex denoted by investigator_id suffixed with an 'i' and a Police officer vertex denoted by officer_id suffixed with an 'o'. We added suffices because it was impossible to distinguish between investigators and officers since they could have the same id as well.

Edges:

'src' of an edge corresponds to an investigator and 'dst' of an edge corresponds to a police officer while the 'relationship' denotes the allegation_id that 'src' (investigator) is investigating the misconduct of 'dst' (officer). We are only working with unsustained complaints here because these are the ones when an officer was not disciplined.

Graph Analysis:





It is surprising that even basic graph properties can reveal interesting information about the relationships between the investigators and officers.

1. Count of edges between every pair of investigators and officers

src	dst	count
3886_i	23841_o	24
2538_i	32166_o	24
3886_i	8562_o	23
2538_i	12478_o	21
3886_i	3454_o	20
3278_i	2725_o	20
3886_i	18076_o	17
3836_i	8386_o	16
3278_i	32164_o	16
3886_i	27778_o	15
2815_i	25039_o	15
3836_i	8620_o	15
2512_i	31837_o	14
3278_i	21615_o	14
3836_i	18205_o	13

The third column denotes the number of times the investigator and officer pairs have worked together.

Investigator with id 3886 (James Eldridge) can be seen to have investigated many allegations that eventually lead to unsustained complaints. On further inspection, we find that James Eldridge himself is accused for misconduct allegations 85% more than other officers. Furthermore, **we observe that the officers whose cases James has investigated are also seen to be involved in co-accusations with James.**

COACCUSED 5-9 TIMES			
			
Police Officer Jerome Finnigan	Captain of Police Warren Richards	Police Officer Carl Suchocki	Police Officer as Detective Bret Rice
175 Allegations 6 Sustained More than 99.9% of other officers	19 Allegations 1 Sustained More than 66% of other officers	78 Allegations 1 Sustained More than 99.5% of other officers	88 Allegations 4 Sustained More than 99.5% of other officers
58-year-old white male	62-year-old white male	48-year-old white male	52-year-old white male
Coaccused in 8 cases.	Coaccused in 7 cases.	Coaccused in 6 cases.	Coaccused in 5 cases.

The ids co-occurring frequently with James (3886_i) are 23841_o, 8562_o, 3454_o, 18076_o and 27778_o. When we look up the officer details, we note that it's Bret Rice, Jerome Finnigan, John Burzinski, Timothy Mc Dermott and Carl Suchocki respectively in order of decreasing co-occurrence frequency. From cpdp.co, we observe four of them are also mentioned in co-accusals in misconduct by James. The one not mentioned in the top percentile are still observed as co-accused officers in the below percentile. We observed this pattern not with only with 3886_i but other investigators too.

Investigator 2538_i has allegations against him 91% more than other officers, Investigator 3278_i has allegation against him 94% more than other officers.

It gives a clear indication that these investigators and officers can also have a direct relationship or understanding that is opening a free leeway for the misbehaving officers via the investigators. In brief, a silence of code is adopted not only within officers involved in misconducts but also between investigators and officers.

2. Connected Components Algorithm

The number of edges in the graphs are 41210 and vertices are 17353. We get 51 connected components denoting a high reachability between the vertices. This could be due to multiple reasons.

Firstly, the investigators (3011) are less in number as compared to the officers (14342) so the same investigators are likely to investigate cases by multiple officers. Secondly, the officer accused in one allegation might serve as an investigator in the other. So the interconnectivity can increase that way. Thirdly, the investigators and officers could indeed have a close-knit relationship that leads to frequent co-occurrences.

To find out which hypothesis is true, we dig in further. When we check how many nodes are present within a component, we notice a high level of disparity. Component labelled as 0 has

17240 vertices, which includes almost all of them in the graph. We notice many dual relationships appearing once in the graph. We find this result tough in order to deduct any plausible interpretation of connected components.

3. Outdegrees of vertices

We calculate the outdegrees of all the vertices in the network. Clearly, top nodes with maximum outdegrees are the investigator nodes because they are involved in multiple allegations for investigations and also they are less in number. We deviate a bit to look further into the attributes of the top investigators; those who worked on the unsustained allegations most frequently.

id	outDegree
1438_i	2497
1657_i	2093
2376_i	1888
2472_i	1593
1715_i	1500
2970_i	1485
2994_i	1475
1595_i	1448
2985_i	1440
3183_i	1358
3270_i	1286
3062_i	1200
2724_i	1182
3028_i	1163
1419_i	1018
2375_i	1018
1667_i	955
1688_i	949
3204_i	914
1676_i	877

We inspect the police department where these investigators serve by manually checking in the **data_investigator_allegation** table. It is surprising to see that top 6 investigators belong to the same police department unit having unit_id = 71! This hints us towards the line of thought that bad apple behaviour is not only prominent within officer network of misconducts but could also be evident among the investigators. They are not directly accused for misconduct but are passively involved in supporting misconducts.

2. **Centrality analysis: Where are the most influential (negatively) people in the CPD network? where the most unsustained reports are concentrated? We want to track the cascading effect between different police officers. Does one officer with a high number of complaints rub off on his culture and behavior on another officer who has a relatively low number of complaints against him?**

For this analysis, we used the **data_officer** table to get the officer's names and the count of unsustained allegations. On the other hand, we used the **data_officerallegation** table to get the relationships between officers that have been accused in the same unsustained allegations.

1. Triangle count algorithm

First, we did a graph where the vertices were the officers with unsustained allegations in their records and the edges the relationship between officers who have been accused in the same allegation, this way we can have a network with officers that have misconduct together. Then, we applied the "triangle count" algorithm to measure the level of connection that each officer has with the network.

In Table 1 we can see the top 20 officers with the most number of relationships on the network, which can be considered as the center of the network. We add a column named "percentile" to see their positions against the rest of the officer's network. The higher the number, the more relationships they have. In Table 2, we can see the top 20 officers with the most unsustained allegations. When we see the percentile column from Table 2, we can see that all the top 20 officers have a percentile around 90 or higher in the level of connection with the network.

If we take a more generic look, seeing a summary of the number of unsustained allegations, the percentile of the level of connection with the network, and the average number of triangle count, as we can see in Table 3, we can notice a direct correlation between the unsustained allegations count and the triangle count. It is very noticeable that in the higher two ranges of unsustained counts, where officers have over 40 allegations, corresponding just to 6.6% of the officers, they have even 8 times the number of triangle counts than ones from the [0-10] range, which are the better behaved. **We can conclude that the more unsustained allegations an officer has, the more centralized and connected he/she is in the network.**

count	id	officer_name	unsustained_count	index	percentile
32118	6315	Terence Davis	19.0	0	1.0
32117	3033	Raimondo Brown	8.0	1	0.999971866647911
32073	3744	Derek Campbell	5.0	2	0.9999437332958222
27855	18042	Donald Mc Coy	9.0	3	0.9999155999437332
27823	441	Fernando Alonzo	12.0	4	0.9998874665916444
23900	21530	Michael Overstreet	36.0	5	0.9998593332395554
23518	27349	Charles Stanton	2.0	6	0.9998311998874666
23499	5180	Stephen Conner	4.0	7	0.9998030665353776
23487	5667	Jerry Crawley	14.0	8	0.9997749331832888
23477	16747	Evetta Lundin	4.0	9	0.9997467998311999
23475	8844	Thomas Flynn	10.0	10	0.999718666479111
23472	23654	Lloyd Reid	0.0	11	0.999690533127022
23472	14750	William Kissane	13.0	12	0.9996623997749332
20185	19856	Ronald Muhammad	5.0	13	0.9996342664228443
19322	8138	Glenn Evans	74.0	14	0.9996061330707554
18773	29882	Fred Waller	41.0	15	0.9995779997186665
18648	28273	James Taylor	15.0	16	0.9995498663665776
18602	28459	Curtis Thomas	20.0	17	0.9995217330144887
18539	5577	Michael Cox	14.0	18	0.9994935996623998
18502	30841	Teresa Williams	22.0	19	0.9994654663103109

Table 1: Top 20 police officers with higher triangle count

count	id	officer_name	unsustained_count	index	percentile
1251	8562	Jerome Finnigan	112.0	588	0.983457588971726
244	28805	Charles Toussas	87.0	2106	0.9407511605007737
1402	17816	Edward May	87.0	567	0.9840483893655929
244	10890	James Grubbs	79.0	2103	0.9408355605570403
94	21837	Joe Parker	78.0	4689	0.8680827120551414
329	29033	Jerome Turbyville	76.0	1529	0.9569841046560698
129	4807	Maurice Clayton	75.0	3703	0.8958221972147982
19322	8138	Glenn Evans	74.0	14	0.9996061330707554
118	13788	Broderick Jones	73.0	4001	0.8874384582923055
1369	2356	Harold Bone	67.0	572	0.9839077226051484
125	13391	Tyrone Jenkins	67.0	3786	0.8934871289914194
266	13095	Gregory Jackson	66.0	1935	0.9455619637079758
270	31631	Adam Zelitzky	65.0	1904	0.9464340976227318
2242	8658	Corey Flagg	63.0	492	0.9861583907722605
215	17285	Edward Maras	62.0	2398	0.9325362216908145
453	25306	James Sanchez	61.0	1099	0.9690814460542974
76	28740	Ignacio Torres	59.0	5455	0.8465325643550429
774	32166	Emmett Mc Clendon	58.0	710	0.98002532001688
9933	12478	Ronald Holt	56.0	365	0.989731326487551
153	1901	Joe Berry	56.0	3233	0.9090448726965819

Table 2: Top 20 police officers with higher unsustained count

Unsustained Count	Percentile of network connection	Triangle count average
50 >	0.938	1,956.5
50 - 40	0.934	1,660.3
40 - 30	0.904	992.1
30 - 20	0.877	759.6
20 - 10	0.825	629.9
0 - 10	0.468	205.7

Table 3: Summary of unsustained allegations, percentile and triangle count average according to connection to network level

2. Motif finding

To analyze the propagation of misconduct between officers we used motifs in PySpark. This indicates a path of officers who have been involved in allegations in a cascade: “Officer A” → “Officer B” → “Officer C” → “Officer D”, we used a chain of 4 officers because of the limitation of the processor of the system. From the 35,545 officers we have as vertices; 17,215 of them have been the origin of a cascade of a length of 4 or more. With this information we can prove that the misconduct of officers is something that spread along with them. We would like to do further analysis with a more powerful machine to find the longest path of each officer and compare with the number of unsustained allegations they have.

a	e	b	e2	c	e3	d
{113, Donna Adams...	{113, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 28412, 5}	{28412, Timothy T...
{113, Donna Adams...	{113, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 18069, 4}	{18069, James Mc ...}
{113, Donna Adams...	{113, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 5782, 3}	{5782, Willie Cro...
{113, Donna Adams...	{113, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 26437, 2}	{26437, Theresa S...
{113, Donna Adams...	{113, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 1412, 1}	{1412, Martin Bar...
{113, Donna Adams...	{113, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 1509, 1}	{1509, Tyrone Bat...
{113, Donna Adams...	{113, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 6928, 1}	{6928, Dominic Di...
{113, Donna Adams...	{113, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 11485, 1}	{11485, Tommie Ha...
{113, Donna Adams...	{113, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 18183, 1}	{18183, Paul Mc G...
{113, Donna Adams...	{113, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 19905, 1}	{19905, Michael M...
{113, Donna Adams...	{113, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 23933, 1}	{23933, Calvin Ri...
{113, Donna Adams...	{113, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 27871, 1}	{27871, Gerome Su...
{79, Erwin Ac Cox, ...}	{79, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 28412, 5}	{28412, Timothy T...
{79, Erwin Ac Cox, ...}	{79, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 18069, 4}	{18069, James Mc ...}
{79, Erwin Ac Cox, ...}	{79, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 5782, 3}	{5782, Willie Cro...
{79, Erwin Ac Cox, ...}	{79, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 26437, 2}	{26437, Theresa S...
{79, Erwin Ac Cox, ...}	{79, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 1412, 1}	{1412, Martin Bar...
{79, Erwin Ac Cox, ...}	{79, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 1509, 1}	{1509, Tyrone Bat...
{79, Erwin Ac Cox, ...}	{79, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 6928, 1}	{6928, Dominic Di...
{79, Erwin Ac Cox, ...}	{79, 200, 1}	{200, Richard Agu...	{200, 241, 1}	{241, Willie Aker...	{241, 11485, 1}	{11485, Tommie Ha...

Table 4: Sample view of cascading effect of officers involved on unsustained allegations

3. Outdegrees:

To analyze the propagation of misconduct from another point of view, we used the algorithm of outdegrees. This indicates how many connections each officer has with other officers in the

network. First, we obtained a sample of the complete table of outdegrees count that can be seen in table 5, where it is possible to identify that officers with more unsustained allegations also have a higher level of contacts in the network. Then we did a summary table to be able to analyze the general trends, where it can be seen that the greater the number of unsustained allegations, the more connections they have with other officers, making it more possible to spread the misconduct.

	id	outDegree	officer_name	unsustained_count
0	3033	332	Raimondo Brown	8.0
1	441	330	Fernando Alonzo	12.0
2	3744	313	Derek Campbell	5.0
3	6315	300	Terence Davis	19.0
4	8138	290	Glenn Evans	74.0
...
18838	2800	1	John Bribiesca	1.0
18839	13812	1	Curtis Jones	0.0
18840	3376	1	Marcellus Burke	0.0
18841	17961	1	John Mc Carthy	0.0
18842	19260	1	Kenneth Mitchell	1.0

Table 5: Sample of Outdegree count from police officers

Unsustained Count	Outdegree average
50 >	55.5
50 - 40	43.7
40 - 30	32.8
30 - 20	27.2
20 - 10	20.9
0 - 10	9.5

Table 6: Summary of unsustained allegations and outdegree average per range

Conclusion

After this network-focused analysis, we were able to identify some variables and relationships that can help us explain that some allegations were falsely investigated leading to an unsustained outcome. The main findings are:

1. Investigators frequently investigating the unsustained allegations also fall in the category having a high percentile of allegations against them. This implies that if they commit misconduct, they allow misconduct.
2. Investigators frequently involved with specific officers also share co-accusals with them on other cases. This implies that a close-knit relationship with the co-accused is satisfied by giving the officer partner a leeway for misconduct. It is akin to a passive involvement in misconduct.
3. Investigators belonging to the same police department unit display similar behavior. They are observed to top the list of investigators handling the highest number of unsustained allegations.
4. Officers with the highest number of unsustained allegations are the ones that are more connected and centralized in the network, having contact with many officers around.
5. This great connectivity also explains the influence they have on other officers, which then triggers the cascade effect of misconduct. We could see that a large percentage of them have reached at least 4 levels of misconduct spread.