Reg Gonzalez CS 6350.0U1

Assignment 3, Part 2 - Report

Query Outputs:

(a.) Find the top 5 nodes with the highest outdegree and find the count of the number of outgoing edges in each

name	outDegree
RepJeffDuncan	120
RepFranklin	121
RepJohnRose	108
GOPLeader	127
RepDonBeyer	109

(b.) Find the top 5 nodes with the highest indegree and find the count of the number of incoming edges in each

name	inDegree
RepBobbyRush	111
SpeakerPelosi	210
GOPLeader	157
SteveScalise	89
SenSchumer	97

(c.) Calculate PageRank for each of the nodes and output the top 5 nodes with the highest PageRank values. You are free to define any suitable parameters.

id	name	pagerank
367	SpeakerPelosi	6.6716229665456925
322	GOPLeader	4.9570044878155874
393	RepBobbyRush	3.7250982484163635
436	RepMarkTakano	3.0940686376861075
179	rosadelauro	2.877991799020614

Assignment 3, Part 2 - Report

(d.) Run the connected components algorithm on it and find the top 5 components with the largest number of nodes.

component	count
0	475

(e.) Run the triangle counts algorithm on each of the vertices and output the top 5 vertices with the largest triangle count. In case of ties, you can randomly select the top 5 vertices.

count	id	name
3048	322	GOPLeader
2799	367	SpeakerPelosi
1908	254	LeaderHoyer
1789	208	RepFranklin
1786	393	RepBobbyRush

Summary and Insights:

Before analyzing the results of each query, it's important to remember what this dataset is and what it represents. This set contains data for the Congressional Twitter network and was collected using the Twitter API.

Query A finds the top 5 nodes with the highest outdegrees. Unfortunately, most of the names in the result are unfamiliar to me. The top 4 nodes are all members of the Republican Party, while the last node (RepDonBeyer) is a member of the Democratic Party; that may be of importance. Outdegrees represent connections *going out* to other nodes, and in this instance, this list of congresspeople represents the number of connections they have going out to other congresspeople on Twitter. Without having more data available, perhaps you could say that Republicans—at least in Congress—are likely to be more active on Twitter and tweet out more to other

Assignment 3, Part 2 - Report

congresspeople than Democrats. That seems plausible especially because this is an election year and Republicans currently do not hold the Presidency or the House of Representatives. Maybe they're campaigning more, and so, are using social media to interact with other politicians more often.

Query B finds the top 5 nodes with the highest indegree. Indegrees represent the connections *going into* other nodes, and again, in this instance, it'd be representative of how many other congresspeople are contacting *these* individuals over Twitter. Unlike Query A, there are more names here that I recognize and are probably recognizable to most people who know about politics. Unlike Query A, though, most of the results here are Democrats (Rep. Rush, Speaker Pelosi, and Sen. Schumer). None of these are shocking results other than Bobby Rush. I'm unfamiliar with Bobby Rush, but a quick Google search shows that he has been a representative for Illinois' 1st congressional district for 3 decades. He was a civil rights activist during the 1960s and co-founded the Illinois chapter of the Black Panther Party. Clearly, he's had several accomplishments over his time as a politician, which is maybe why he's been contacted a lot via Twitter. The other 4 nodes (Pelosi, Schumer, GOP leader, and Scalise) are all major players in Congress, so it doesn't surprise me that they have a lot of interaction on Twitter considering their roles in government.

Query C performs the PageRank algorithm and outputs the top 5 nodes with the highest PageRank values. The PageRank algorithm gives an importance rating based on how many pages link to it. In this case, it'd be connections made via Twitter. 3/5 of the nodes are the same as the previous query (Pelosi, GOP Leader, and Bobby Rush), but there are two new ones: Mark Takano and Rosa DeLauro. The 3 previously mentioned nodes aren't surprising since they were already featured in other query results; however, Takano and DeLauro are. I'm unfamiliar with them as congresspeople and based on some—admittedly, surface-level—research I did on them, neither of them seem to have huge presences in their respective districts. However, it is important to remember that this data resolves around Twitter, so perhaps they're more active there than other politicians, which is why their PageRank value is higher.

Query D finds the top 5 components with the largest number of nodes. Connected components of a graph is a subgraph in which each pair of nodes is connected to each other via a path. As you can see from the output, there was only one connected component that was found. When running this through Databricks, I found that many of the Spark Jobs were skipped, so that may have to do with the lack of other found

components. However, there's also the possibility that every node is connected in some way to every other node.

Finally, Query E performs the triangle count algorithm and outputs the top 5 nodes with the largest triangle count. As the name implies, this algorithm finds the number of triangles for each node in a graph. Again, Speaker Pelosi, GOP Leader, and Rep. Bobby Rush are all present; this is unsurprising since they were already in other query outputs. The two new nodes are Leader Hoyer and Rep. Franklin. Leader Hoyer, in reference to Steny Hoyer, is a representative for Maryland's 5th congressional district and served as the House Majority Leader from 2007-2011. The fact that he has a high triangle count should not be shocking since I'm sure his high position in Congress meant that he interacted a lot with other congresspeople on social media. Rep. Franklin, who I assume is about Scott Franklin, is a member of the Republican Party and serves as the representative for Florida's 15th and 18th congressional districts. According to Wikipedia, he is a new addition to Congress, having only served since 2021. Maybe since then, he's become a lot more politically active, and therefore active on Twitter.