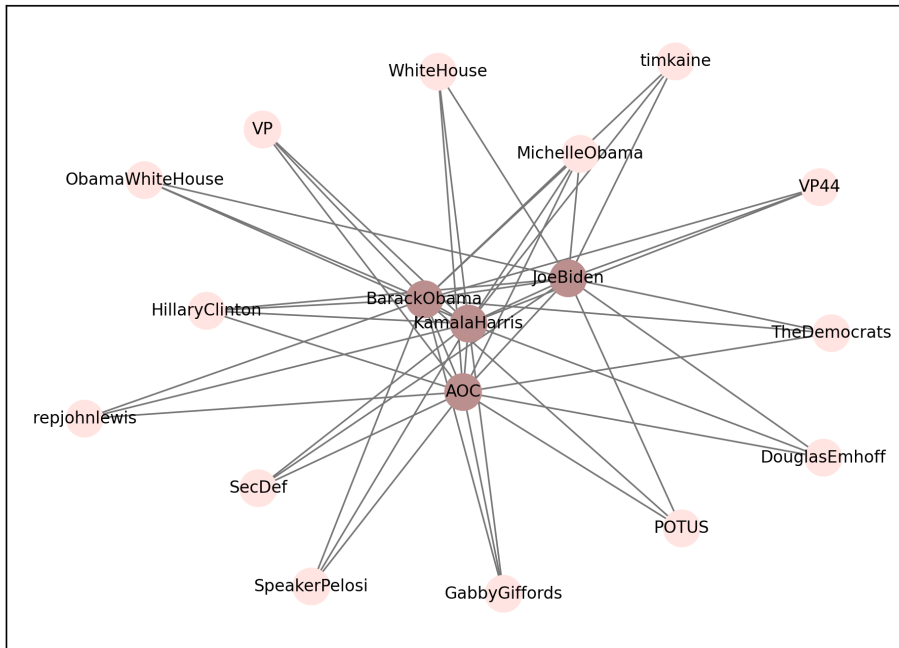


## Part 1 A-B

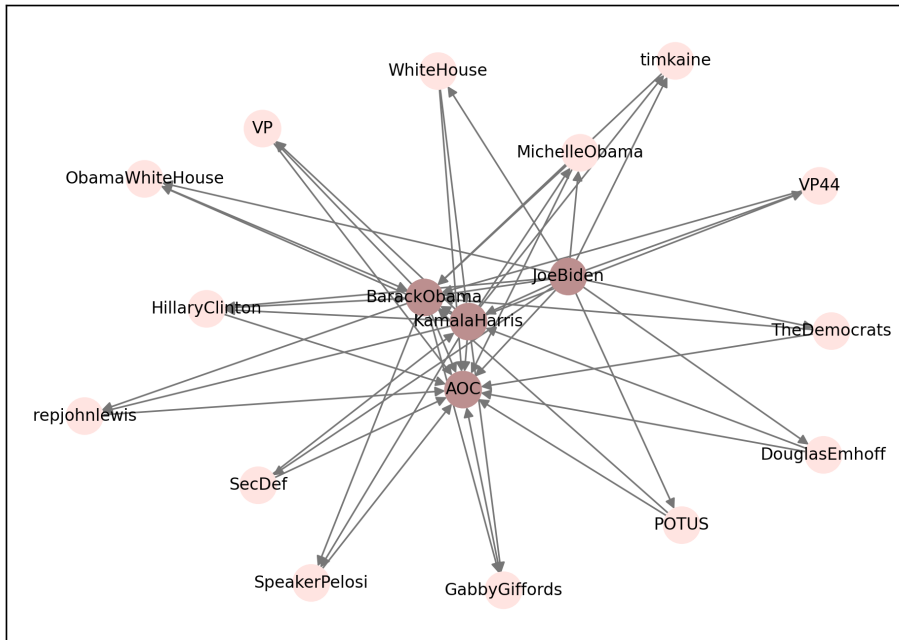
```
Target twitter accounts:
['JoeBiden', 'BarackObama', 'KamalaHarris', 'AOC']
```

## Part 1 C

Undirected graph of Twitter targets and accounts following



Directed graph of Twitter targets and accounts following



## Part 1 D

```
Dijkstra shortest path:
['WhiteHouse', 'JoeBiden', 'SecDef']
```

## Part 1 E

Degree centrality, undirected: JoeBiden: 0.76 POTUS: 0.18 SecDef: 0.18 WhiteHouse: 0.18 DouglasEmhoff: 0.18 KamalaHarris: 0.88 timkaine: 0.18 HillaryClinton: 0.24 ObamaWhiteHouse: 0.18 VP44: 0.18 TheDemocrats: 0.18 MichelleObama: 0.24 BarackObama: 0.82 VP: 0.18 GabbyGiffords: 0.18 repjohnlewis: 0.18 SpeakerPelosi: 0.18 AOC: 0.82	Degree centrality, directed: JoeBiden: 0.88 POTUS: 0.18 SecDef: 0.18 WhiteHouse: 0.18 DouglasEmhoff: 0.18 KamalaHarris: 1.0 timkaine: 0.18 HillaryClinton: 0.24 ObamaWhiteHouse: 0.18 VP44: 0.18 TheDemocrats: 0.18 MichelleObama: 0.24 BarackObama: 0.94 VP: 0.18 GabbyGiffords: 0.18 repjohnlewis: 0.18 SpeakerPelosi: 0.18 AOC: 0.82
Closeness centrality, undirected: JoeBiden: 0.81 POTUS: 0.55 SecDef: 0.55 WhiteHouse: 0.55 DouglasEmhoff: 0.55 KamalaHarris: 0.89 timkaine: 0.55 HillaryClinton: 0.57 ObamaWhiteHouse: 0.55 VP44: 0.55 TheDemocrats: 0.55 MichelleObama: 0.57 BarackObama: 0.85 VP: 0.55 GabbyGiffords: 0.55 repjohnlewis: 0.55 SpeakerPelosi: 0.55 AOC: 0.85	Closeness centrality, directed: JoeBiden: 0.18 POTUS: 0.19 SecDef: 0.19 WhiteHouse: 0.19 DouglasEmhoff: 0.19 KamalaHarris: 0.18 timkaine: 0.19 HillaryClinton: 0.24 ObamaWhiteHouse: 0.19 VP44: 0.19 TheDemocrats: 0.19 MichelleObama: 0.24 BarackObama: 0.18 VP: 0.19 GabbyGiffords: 0.19 repjohnlewis: 0.19 SpeakerPelosi: 0.19 AOC: 0.0
Betweenness centrality, undirected: JoeBiden: 0.16 POTUS: 0.0 SecDef: 0.0 WhiteHouse: 0.0 DouglasEmhoff: 0.0 KamalaHarris: 0.23 timkaine: 0.0 HillaryClinton: 0.0 ObamaWhiteHouse: 0.0 VP44: 0.0 TheDemocrats: 0.0 MichelleObama: 0.0 BarackObama: 0.19 VP: 0.0 GabbyGiffords: 0.0 repjohnlewis: 0.0 SpeakerPelosi: 0.0 AOC: 0.19	Betweenness centrality, directed: JoeBiden: 0.01 POTUS: 0.0 SecDef: 0.0 WhiteHouse: 0.0 DouglasEmhoff: 0.0 KamalaHarris: 0.02 timkaine: 0.0 HillaryClinton: 0.0 ObamaWhiteHouse: 0.0 VP44: 0.0 TheDemocrats: 0.0 MichelleObama: 0.0 BarackObama: 0.01 VP: 0.0 GabbyGiffords: 0.0 repjohnlewis: 0.0 SpeakerPelosi: 0.0 AOC: 0.0
Eigenvector centrality, undirected: JoeBiden: 0.38 POTUS: 0.15 SecDef: 0.16 WhiteHouse: 0.16 DouglasEmhoff: 0.16 KamalaHarris: 0.41 timkaine: 0.16 HillaryClinton: 0.21 ObamaWhiteHouse: 0.16 VP44: 0.16 TheDemocrats: 0.15 MichelleObama: 0.21 BarackObama: 0.39 VP: 0.16 GabbyGiffords: 0.16 repjohnlewis: 0.16 SpeakerPelosi: 0.16 AOC: 0.39	Eigenvector centrality, directed: JoeBiden: 0.21 POTUS: 0.21 SecDef: 0.21 WhiteHouse: 0.21 DouglasEmhoff: 0.21 KamalaHarris: 0.21 timkaine: 0.31 HillaryClinton: 0.31 ObamaWhiteHouse: 0.31 VP44: 0.31 TheDemocrats: 0.21 MichelleObama: 0.31 BarackObama: 0.21 VP: 0.21 GabbyGiffords: 0.21 repjohnlewis: 0.21 SpeakerPelosi: 0.21 AOC: 0.0

**Discussion:**

The degree centrality for the undirected and directed graphs are generally similar, except for three of the target accounts (JoeBiden, KamalaHarris, BarackObama) where the degree centrality is slightly higher in the directed graphs for these targets compared to the undirected graph. Since the directed graph takes into account both incoming and outgoing edges whereas in the undirected graph this measure only examines the total number of edges. This disparity between the centrality means these targets have more connectivity in the directed graph and it would be interesting to dig deeper to determine whether they have more followers or if they are following more people in the rest of the follow list. The AOC target did not have a difference in degree centrality when examining directed and undirected graphs.

The closeness centrality measure examines the shortest geodesic distance from one node to all the other nodes in the network in undirected graphs and in directed graphs it only takes into account the connections from that node to outgoing nodes. For this reason we see higher values in the undirected graph than in the directed graph. Especially drastically lower values for the targets. This is because when examining the visualization of the directed graph these accounts hardly follow other accounts, so hence they have a lower closeness centrality. I particularly want to highlight the AOC account that has a centrality of 0.85 in undirected and 0 in the directed graph, which upon visual examination is because that account hardly follows others.

Betweenness centrality examines the shortest path distance for each pairwise node in the network and nodes that pop up more often in these lists have a higher betweenness centrality score. Since these twitter accounts have low following numbers many of these scores are low or 0. However, the targets are the only accounts in the undirected graph that have non zero values, so they are hubs for shortest path distances between nodes in the network. This mostly continues into the directed graph except the AOC account has a 0 value, which means this account is less of a hub for shortest path connections. This tracks to the closeness centrality measure where we indicated the AOC account has a 0 value as well because it does not follow many other accounts, so it therefore would not be a hub for a directed graph. A further examination of this would be to see how these values change by using other path search algorithms e.g. DFS or BFS.

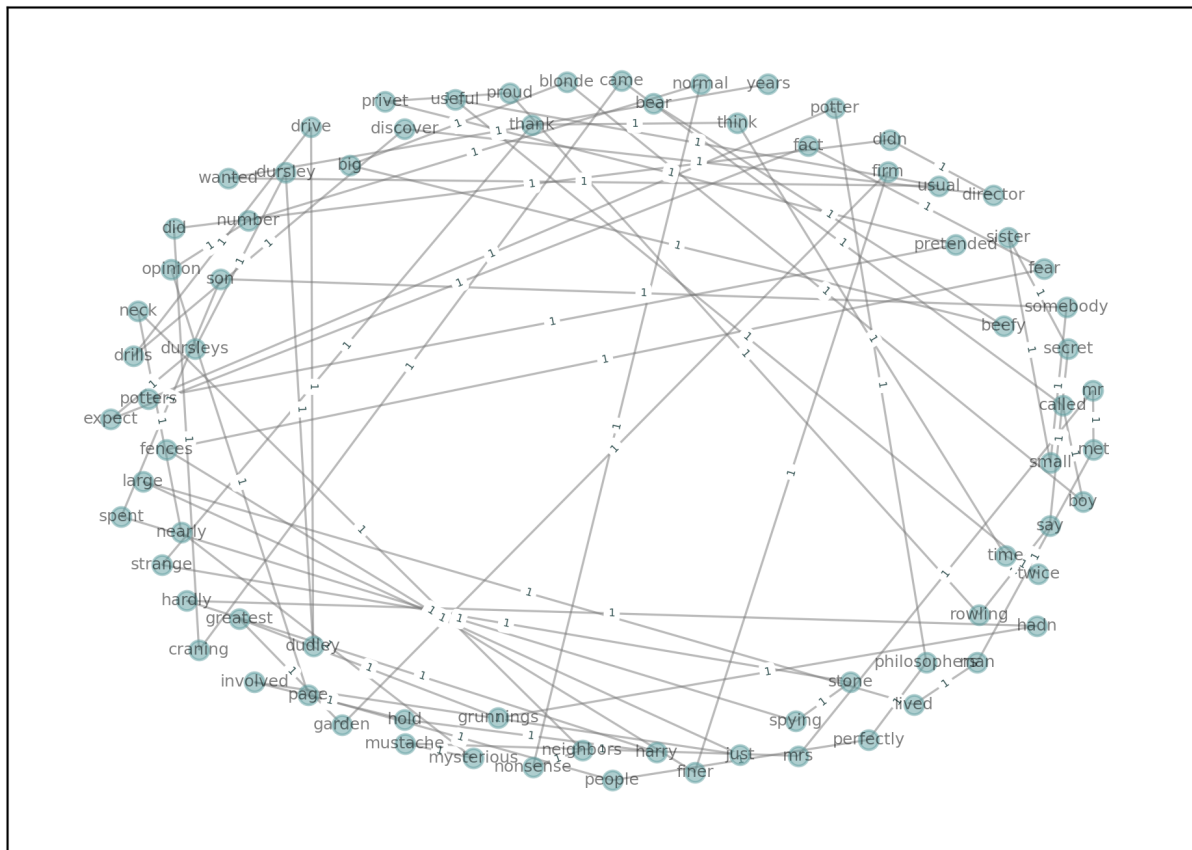
Eigenvector centrality is a measure that determines the importance of a node based on how many other nodes its neighbours are connected to. For e.g. is node A has many direct connections but the neighbours of node A do not have many connects then node A will have a lower Eigen score compared to for e.g. node B that may have less direct neighbours but their neighbours are more highly connected with other nodes in the network. As is typical in the undirected graph the Eigenvalues are higher overall with the targets having higher values; because this measure does not need to account for directedness. The scores flatten in the directed graph i.e. the targets and other nodes have around the same values, which is also expected because this target group does not follow alot of other accounts. A further examination of this would be to include a large sample size and examine how this value varies with accounts that follow alot of other accounts.

## References

- 1) Bhasin, J. (2019). Graph Analytics - Introduction and concepts of centrality. Towards data science: <https://towardsdatascience.com/graph-analytics-introduction-and-concepts-of-centrality-8f5543b55de3>
- 2) Aomar, A. A. (2020). Notes on graph theory - Centrality measures. Towards data science: <https://towardsdatascience.com/notes-on-graph-theory-centrality-measurements-e37d2e49550a>

## Part 2

Undirected weighted graph of bigram count



### Discussion:

The above weighted undirected graph shows the bigram counts obtained from text data of the first few pages of Harry Potter book 1. The text data was preprocessed into tokens and then bigrams were determined. The edges show the connectivity between words, edge weights show the number of times the particular bigram appeared in the dataset and the nodes show the words forming the bigram. Since this is a small subsampled dataset each bigram only had a count of one. Though the weight count is low, interesting connectivity patterns can still be determined e.g. what words have connections to other bigrams (neck), and words that do not have connections (craning and involved).