# Computational Journalism: Assignment 6

## Pankhuri Kumar (pk2569)

The initial dataset, when loaded into Gephi, does not give out too much information at first glance. It is easy to see which nodes have more heavily weighted edges. It is hard to discern the relative importance of central nodes, however. The graph also does not have any sense of centrality based on connections or weights.
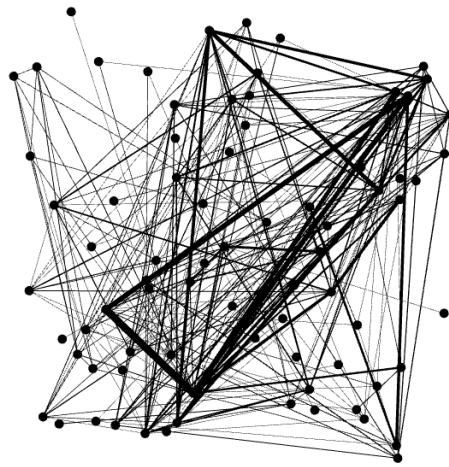


Fig 1. The original Les Miserables graph as loaded by Gephi

However, using the Edit tool, I can still figure out Valjean, Cosette, Javert and other weighted nodes.
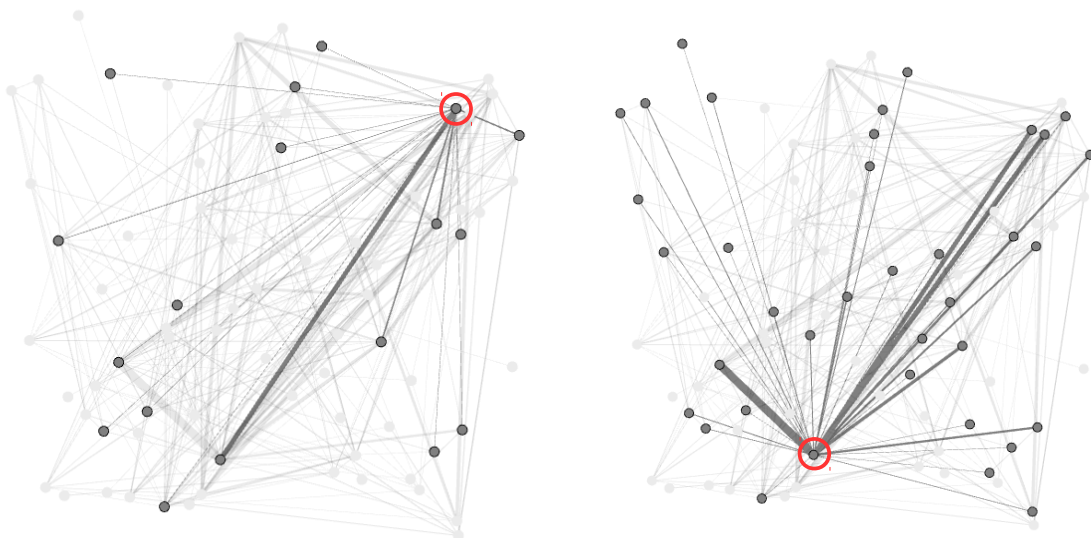


Fig 2. Highlighted nodes for Javert and Valjean, respectively

After running the Force Atlas 2 algorithm, the graph re-arranges to give some notion of centrality. It is easy to notice that most of the important characters align along the central axis of the graph, like Valjean and Cosette. However, there are other characters that appear to be important according to this arrangement, like Gavroche and Woman1/2, who are relatively unimportant in the plot.

It seems like the Force Atlas 2 algorithm gives importance to the number of connections, but also the weights assigned to each connection. However, it is hard to differentitate between the importance of Cosette and Gavroche. Honestly, if I didn't know the plot, I would mark Gavroche more important than Cosette by this graph.
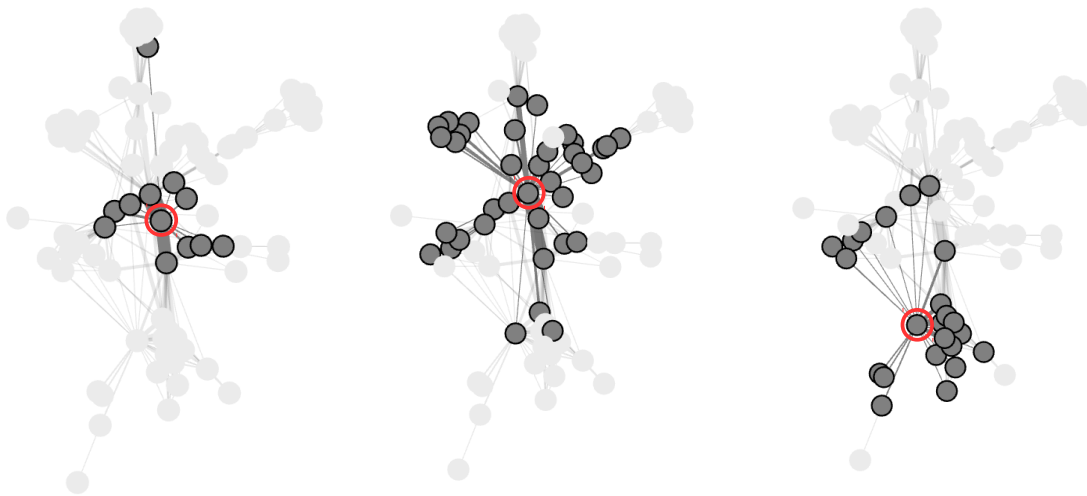


Fig 3. Force Atlas 2 graphs for central characters like Cosette and Valjean along with non-central characters like Gavroche respectively

Coloring the graph according to degree does not agree with the algorithm layout. Like I previously mentioned, the degree for Gavroche is higher than Cosette, and is hence colored darker than Cosette. However, we know from the plot that Cosette is one of the central characters to the story. In fact, even Javert and Marius are colored to be more important than Cosette – though I agree with Javert's color, I do personally believe Marius to be more *central* to the plot than Cosette.
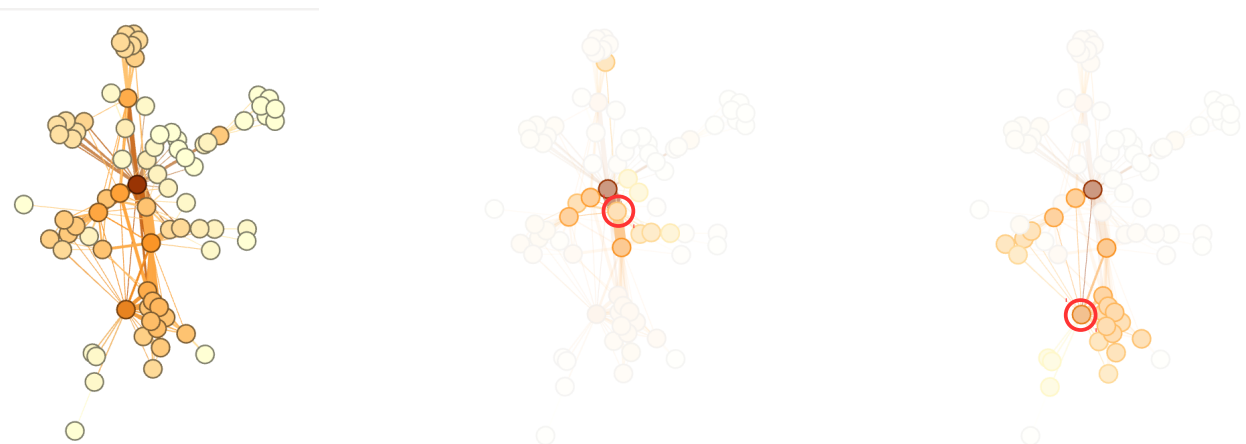


Fig 4. Degree-based colors for the entire graph, Cosette and Gavroche respectively

The Betweenness and Closeness centrality are interesting measures of importance. The Betweenness Centrality measures "how often a node appears on shortest paths between nodes in the network." Thus, it gives more importance to nodes that join relatively disjoint parts of the graph, or nodes that act as bridges between relatively disconnected parts of the graph, like Fantine and Myriel. This is different from degree coloring, where more connections leads to higher importance.

The Closeness Centrality measures "the average distance from a given starting node to all other nodes in the network." I'm not sure if the note on inverse scale for Closeness centrality is valid – my program assigns Valjean the highest centrality score (0.644), and fringe characters (Jondrette – 0.26, MotherPlutarch – 0.28) lower scores.

However, the Closeness Centrality seems to again, value the number of connections and their weights over other factors. More central characters seem to have higher importance while characters on the outer edges of the graph (with higher distances to other nodes) have lighter colors. Here, nodes with heavier weights are considered more important than nodes with lesser weights, if they are equally on the outside of the graph. The colors for the Closeness Centrality seem to become lighter as you move towards the outer edges of the graph.

Again, I disagree with the Closeness Centrality marking Gavroche and Marius as more important than Cosette.
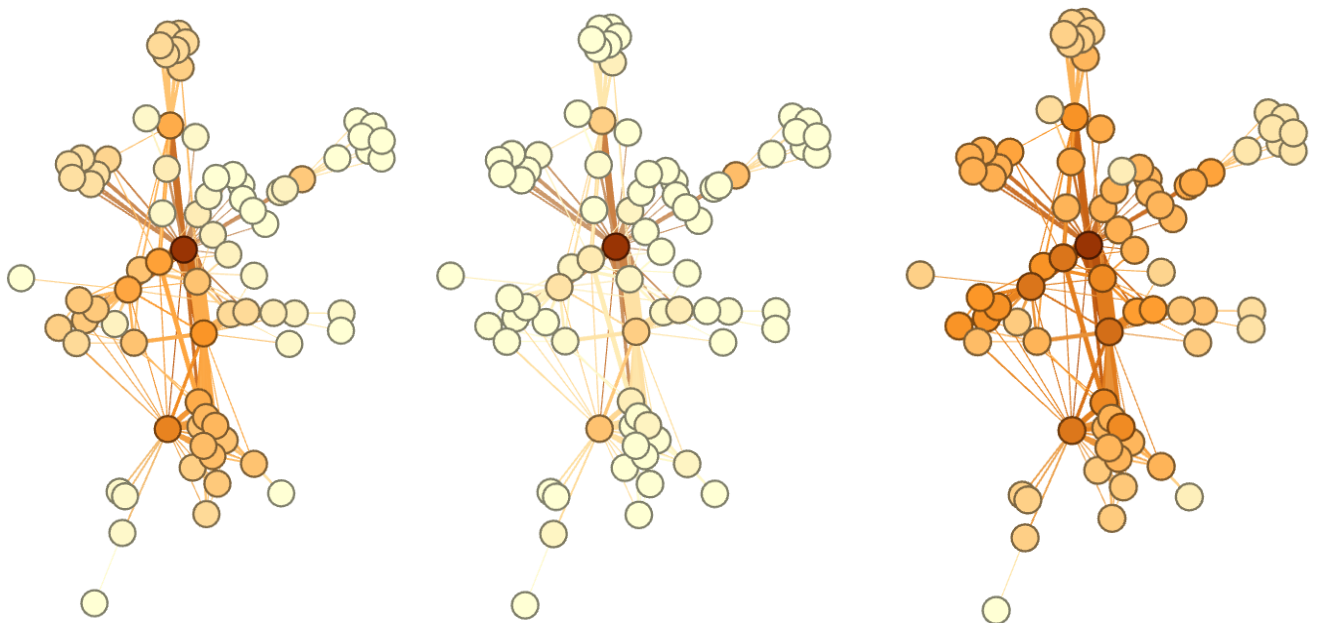


Fig 5. Coloring of nodes based on degree, betweenness centrality and closeness centrality respectively

Amongst the three methods of centrality, I would prefer the Closeness Centrality over other measures. The Closeness centrality seems to take not only the number of connections, but also the weight of these connections into account. This makes it easier to judge even an unknown network, where a poorly connected character (Cosette) is given equal or more importance than other characters because the weight of her connections are large. Characters like Eponine, who talks to

more characters, but has lesser number of interactions with these characters (and is a minor character in the play) is given lesser importance.

Betweenness centrality ends up giving more importance to characters that connect different parts of the play. Thus, characters like Fantine (who interacts with Valjean and Javert, but also with factory workers) and Myriel (who interacts with Valjean, but also the police officers), who appear in only in limited parts of the play get more importance than Cosette, Javert or Marius who are more central characters to the play. The betweenness centrality is important to find characters who play a significant but minor role in the play, or change the direction of the play, but underscores the importance of other characters. It is easy to validate this with a known play like Les Miserables, but if this were an unknown network, the betweenness closeness misleads us about the relative importance of central characters.

The Closeness Centrality still manages to give due importance to these characters, who either have multiple connections, or have higher weights for some portion of the play, and seems like a better measure of centrality or importance. However, it does not diminish the importance of central characters, who are all colored darker than these minor-significant characters.

The degree centrality is an incomplete measure – minor characters may talk to a lot of people, but do little to forward the plot of the play. In a more-real world scenario, these would be people who carry out the grunt work of an (illegal) operation, but are rarely the people who might be making large scale decisions of the direction of said operation. Giving these people more importance only indicates that they might have more information about other characters, but ignores how (in)valuable that information might be.