Assignment 4:
Social Network Analysis

Acauã Sperl de Faria
asf2169

Shivangi Saxena
ss4733

3.
It is possible to see that some characters are more connected and have stronger connections (appears together in more chapters, represented by thicker edges). This is a strong indication that those characters are important to the plot. However, it is clear that it is not easy to evaluate the graph properly without reorganizing it and/or using a better visualization.

6.
Characters with most central location:
Valjean, Javert, Cosette, Marius and Gavroche
The characters given the most central location do, in fact, point to the most important characters, except for Gavroche, who is not as important. It is also worth noting that the "Friends of the ABC" characters are all clustered together.

7.
The darkest node, Valjean, is the main character and also the most central node. It is also possible to see a correlation between the degree of the node and how close to the center it is A lot of the major characters, like Valjean, Marius, Javert are located at the centre.
However, the central position has a higher correlation with the importance of the character than the degree of its node. It is worth noting that Gavroche's node is one of the darkest, although it is not so important in the plot. This seems to occurs because he is connected to a lot of characters from the "Friends of the ABC".
Conversely, we can see the character of Cosette to be having a lower degree in the graph, despite being one of the key characters in the plot.

8-1.
Characters marked as important based on betweenness:
Valjean (much higher), Gavroche, Myriel, Marius and Fantine

Betweenness seems to exclude characters that are on the edge of the network, even if the character has a high degree. A node that has a high betweenness tends to have a high degree but the opposite may not be true (This is true for Les Miserables network but not in general. There is not much correlation between degree and betweeness. A node can have a low degree but if it is the only connection between large disconnected groups, then its going to have a high betweenes). Gavroche, for example, connects the "Friends of the ABC" to the plot, so he has a high degree and also a high betweenness. Javert, on the other hand, has a high degree but not a high betweenness, since he interacts with a lot of characters but does not connect peripheral characters to the plot.

8-2.
Degree is local (who you are directly connected to) while closeness is global (who in the whole graph you can reach and how easily).

Both closeness and betweenness are global but while closeness is about how many degrees on average separates the node from the others on the graph, betweenness is more about how well a node connects different nodes or clusters of nodes in the graph.

9.

Centrality measures can be used to find the most important character in the plot, on the assumption that the most important character is also the one 'most centrally located' in the network and would have the most number of interactions with other characters.

The Degree centrality tells us how many nodes a character has interacted with in the plot. It gives a sense of how "popular" a character may be, based on its number of connections. We can see from the graph that characters like Valjean, Javert and Marius have high degrees. However, degree centrality is not very accurate. This can be seen in the "Friends of the ABC" cluster, which although has a high degree, isn't actually very central to the plot. If an unimportant character interacts with a lot of other unimportant characters, then he would have a high degree but this does not necessarily make him more important.

Betweenness centrality measures how many shortest paths between nodes pass through a particular node. In general, such nodes would indicate a higher control on the information relay in the network. In the case of this plot, however, Betweenness is not completely accurate either. This is because a not-so-important character that connects a lot of unimportant characters to the plot would have a high betweenness. As seen in the graph, Gavroche and Bishop Muriel have a high betweenness, despite not being very central characters in the story.

Closeness centrality tells us how closely connected a character is to the rest of the nodes in the graph. It can be perceived as how connected a node is to more "popular" nodes, and tells us how easily it reaches other nodes. It is best illustrated by the common "Six Degrees of Kevin Bacon" game.
However, in the case of this story, closeness gives a rather scattered outlook as to the importance of characters. In the graph, we can see several clusters being marked with high closeness, with very low importance to the plot. Examples are the characters of convicts (like Brevet and Chenildieu) and also friends of Fantine - Favourite, Dahlia etc, all of whom have relatively high closeness. So closeness tends to give more noise, at least with respect to this plot.

All the three algorithms clearly mark Valjean as the main character of the plot. For identifying the important characters in this plot, betweenness does the best job. Degree also identifies the other important characters but have some false positives. Closeness does not make much distinction between most characters, except for Valjean and therefore is the worst for this plot.

Without knowing the plot, it would be hard to validate the accuracy of the algorithms and choose the best option. More research data would be necessary. In this case, by reading a summary of the plot. More generally, in journalism, this would require cross referencing with other data.