Digital Humanities: Tools and Methods Assignment 4

Part A
Collaborated with Spring Wan

Moretti (2011) examines the use of network analysis to comprehend the links and interconnections between characters in novels in his article "*Network Theory, Plot Analysis*". For *Hamlet* and *The Story of the Stone*, he provides a visual summary of the network of characters, which depicts the relationships between the characters as lines on a network.

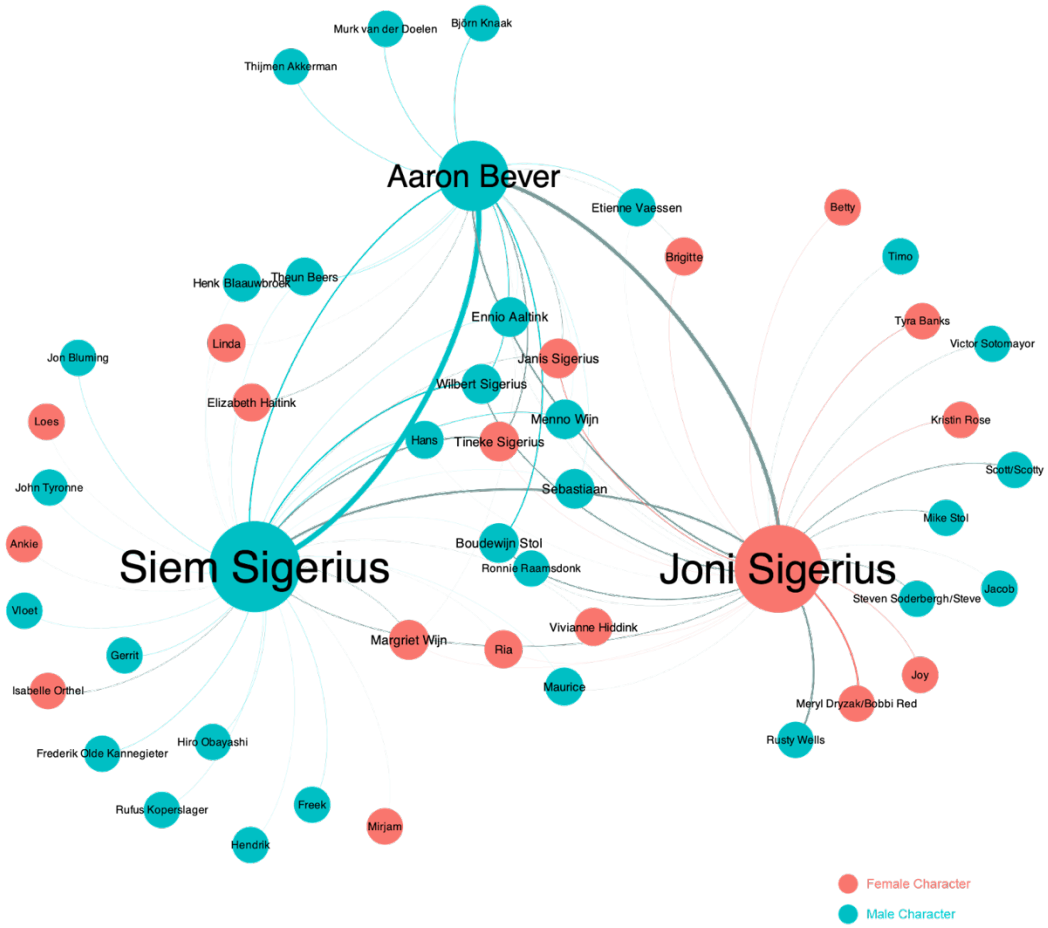
The fact that this visual summary (Graph 1) gives a clear and succinct understanding of the relationships between the characters in *Bonita Avenue* is one of its advantages. The visual summary makes it simple to identify which characters are connected to one another and to comprehend the patterns and trends in their interactions by representing these relationships as lines on a grid.

The great advantage of Graph 1 is when we talk about a character in *Bonita Avenue*, it is no longer an abstract literary theory, but automatically turns to 'consciousness' and 'interiority' (Moretti, 2011). Specifically, firstly, centrality. We can judge the centrality of Aaron Bever, Siem Sigerius and Joni Sigerius by the position of the characters. In other words, Aaron Bever, Siem Sigerius and Joni Sigerius are the typical characters in the novel. Secondly, clustering. The triangle formed by Aaron Bever, Siem Sigerius and Joni Sigerius which is the densest part of the whole network. Almost everyone in the network is associated with them and the clustering is almost 100%. Lastly, plot. According to the sparseness of the lines between the different characters, that is, the complexity of their interactions, we can infer the direction of the novel's plot. Moreover，we can conclude that the number of male characters is roughly the same as the number of female characters based on the legends, with slightly more males.

This graphic explanation does have certain drawbacks, though. For instance, it offers no details regarding the nature of the relationships between individuals. Therefore, beyond the relative strength of the associations, this visual network analysis offers no information about the nature of the links between the characters. For instance, it is clear that there is a strong connection between Aaron Bever and Siem Sigerius, but it is impossible to determine if this bond is friendly or hostile.

Additionally, as was already indicated, a network analysis of this kind cannot adequately capture the complexities of the character associations, and as a result, crucial details may be lost when they are summarized. Examples include the particular events and interactions that shaped these connections. Such a visual network analysis might exclude certain potentially intriguing but subtle connections by concentrating exclusively on the most crucial or central character interactions.

Graph 1: Network Analysis of relationships among characters in Bonita Avenue



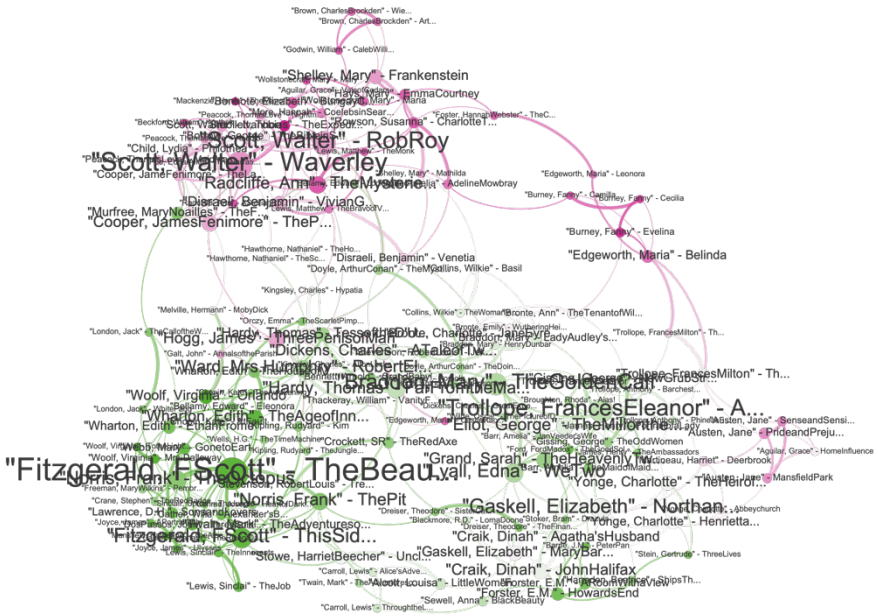
Part B

In textual analysis, analysing stylistic relationships as networks can visualise the similarities and connections between different texts or authors. By considering texts or authors as nodes in a network and their stylistic similarities as edges that connect them, it is possible to use graph theory and network analysis techniques to discover patterns and insights that are difficult to discern using traditional methods. It can also be used to detect communities in the network that may correspond to different genres or periods of writing.

Literature from different eras, genders and countries can exhibit different textual styles, so for this study we used 150 British novels published between 1771 and 1930, downloaded from the NovelTM corpus, as source data, and the output of Stylo from this dataset as input data for Gephi, to analyse the stylistic relationships between these 150 novels. The output of Stylo from this dataset was used as the input data for Gephi to carry out a network analysis of the genre relations between these 150 novels and to output the visualised results.

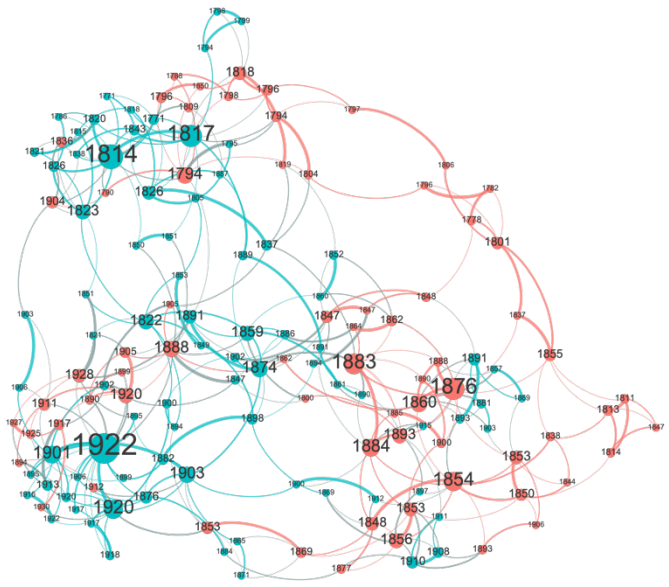
The following three graphs present visualisations of the dataset analysed by year of publication, author gender and narrative perspective respectively. In these three visualisations, each node represents a novel, and the size of the node represents the degree centrality of the node, i.e. the number of connections it has with other nodes, the more connections it has, the more influential the novel is in the community or cluster. The position of the nodes also indicates that the novels are more connected to each other. The edge of the nodes and its weight represent the strength of the connections between the nodes. The higher the weight of the edge between two novels (the thicker it is), the stronger the connection between the two novels, i.e. the more similar they are in terms of genre, language choice or other aspects.

Graph 2 shows a network analysis of these 150 British novels based on their year of publication. The colour pink represents the year of publication closer to 1771, while the colour green represents closer to 1930. This means that British novels published in the same decade have greater textual similarity. Label size in Graph 2 is related to the level of influence of the node in the community, so we can see that, for example, in the late 18th century's community, Scott Walter's work has a greater influence. Also in the lower right-hand part of Graph 2, there are no nodes that appear as prominent as in the other two communities, which may also mean that in the 19th century, the English novel showed a relatively more blossoming development.



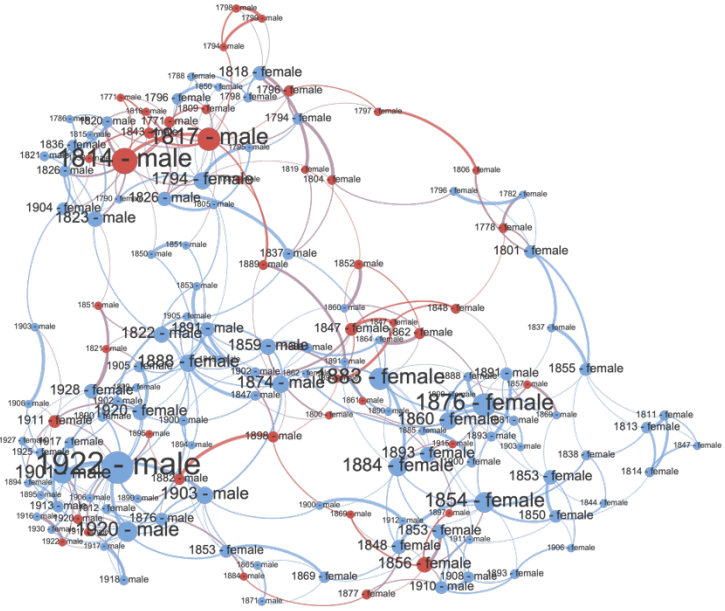
Graph 2: Network Analysis of relationships among British novels between 1771 and 1930 categorised by year

Graph 3 then represents the position of male and female writers in the overall network, with the label representing when the novel was published. The blue nodes represent male writers and the coral nodes represent female writers' works. Firstly, the colour distribution shows that there are more male writers than female writers in this dataset, which spans 160 years. Moreover, more of the novels that are at the centre of their respective community are also from male writers, whereby it could also be argued that male writers produce novels that are more influential than female writers. But with the combination of the labels, which represent the year of publication, also suggests that female writers emerged gradually over time (there were also some influential female writers at the end of the 19th century). We can also see that the blue and coral nodes are relatively concentrated throughout the network, forming their own community, thus suggesting a degree of exclusivity in the influence of male and female writers within their respective gender groups of writers.



Graph 3: Network Analysis of relationships among British novels between 1771 and 1930 categorised by gender

Graph 4 shows the narrative perspective of each novel, with red being first person, the cyan nodes representing third person narratives, and the labels being the year of publication as well as the gender of the writer. The graph shows that at the end of the 18th century, novels written in the first person perspective had absolute centrality, while by the beginning of the 20th century novels written in the third person perspective had taken over. Combined with the labels of the writer's gender, it can also be seen that female writers, relatively speaking, tend to use the third-person point of view in the narrative process of writing their novels.



Graph 4: Network Analysis of relationships among British novels between 1771 and 1930 categorised by narrative perspective

The overall structure and relationships of a text corpus can be learned via analysing textual relationships as network graphs. It's important to keep in mind that this method might not fully capture the texts in the dataset. For instance, it might not include differences in genre, content, or theme, which are also crucial for comprehending the relationships between novels. In addition, this type of stylometric approach means that results may be influenced by the particular vocabulary and features used in the analysis. Overall, network analysis can be a useful tool, but for a more complete picture, it should be used in addition to other methods.

Reference

Franco Moretti, ‘Network Theory, Plot Analysis’. New Left Review 68 (2011): 80-102.