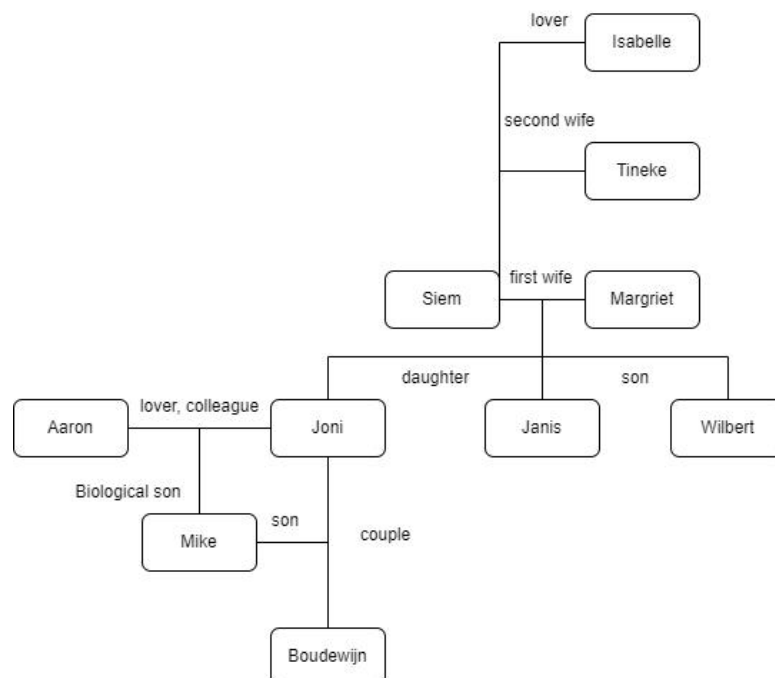
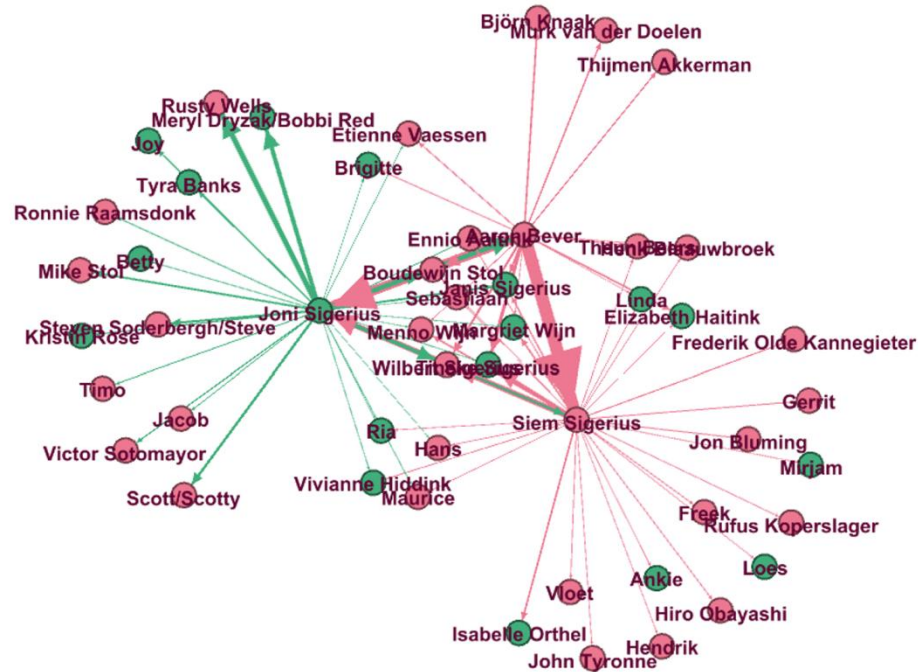


Part A Collaborated with Haozhe Bai



After Gephi's analysis, the relationship of the characters can be presented very clearly.

The book uses Joni, Aaron, and Siem as the main nodes from which the different characters are drawn. There are 29 characters related to Joni, 17 characters related to Aaron, and 29 characters related to Siem. The number of characters associated with multiple protagonists is 8. The connections between these people show the framework of the story very clearly. In the picture, the thicker the line, the closer the relationship between the two characters. The two thickest lines connect Joni, Aaron, and Siem, Aaron, which are the most important relationships in the text. Secondly, there are also some relationships that have significant meaning in the text, such as Siem, Joni and Joni, Rusty and Joni, Mike, etc. These relationships are well represented in the text. These relationships are well reflected in the storyline, where Siem becomes the Minister of Education, his marriage to two women, and the unpleasant events between him and his son Wilbert. the experience of running a pornographic website between Joni and Aaron and the pain of separation can be seen in these relationships. However, there are still some elements that have been overlooked, such as the reference to Thom de Graaf, who introduced Siem as Minister of Education. But this relationship is not reflected in the picture.

In Moretti's article, he uses a thick line to show that there is a direct connection between the two characters and a dashed line to represent the indirect connection between the two in the narrative of the story. I think this is a better way to show the content of the storyline in more aspects and levels, and we can even analyze the tendency of the writing method from it. However, Gephi's analysis is unable to do so, and it can be said that some of the information in the article is somehow left out.

Moretti mentions symmetry in his article and analyzes in detail the similarities and differences of symmetry between Chinese and English works. I think the result in the analysis about Bonita Avenue can be seen as a centrosymmetric figure rather than an axisymmetric figure. We get the information that the plot is quantified, that not only explicit but also implicit connections are used in the character network, and that the thickness of arrows and lines can be used to visualize weight and direction.

The information we lose is that we usually consider the dualism of roles: protagonists vs. secondary roles, or "round" vs. "flat" (Moretti) only see the relationships of characters around the three main roles, regarding the relationships of secondary roles. In Gephi's distribution we do not see the hierarchy of their roles, nor can we determine the different genres, whether tragedy or comedy.

Reference

Moretti, F. (2011). *Network theory, plot analysis*. Universitätsbibliothek Johann Christian Senckenberg.

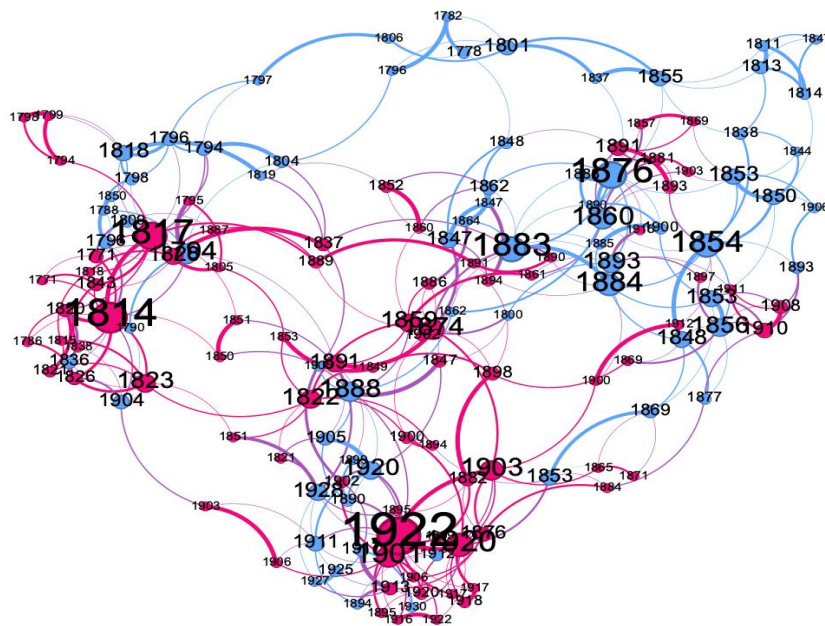
Part B

Analyzing stylistic relationships as networks in textual analysis can help visualize the similarities and connections between different texts or authors. By viewing texts or authors as nodes in a network and stylistic similarities as edges connecting them, graph theory and network analysis techniques can be used to uncover patterns and insights that are difficult to discern using traditional methods. It can also detect communities in the network that correspond to different writing genres or periods.

Stylometry is a method that utilizes quantitative analysis of writing styles to uncover patterns and trends in literary works, which may not be discernible through conventional qualitative means. In this report, we will use a network analysis approach to investigate the relationships between 150 English novels published between 1771 and 1930, using the Burrows Delta measure to determine their stylistic similarities. To visualize and interpret the network of similarities between these novels, we will use Gephi, a popular open-source network analysis tool.

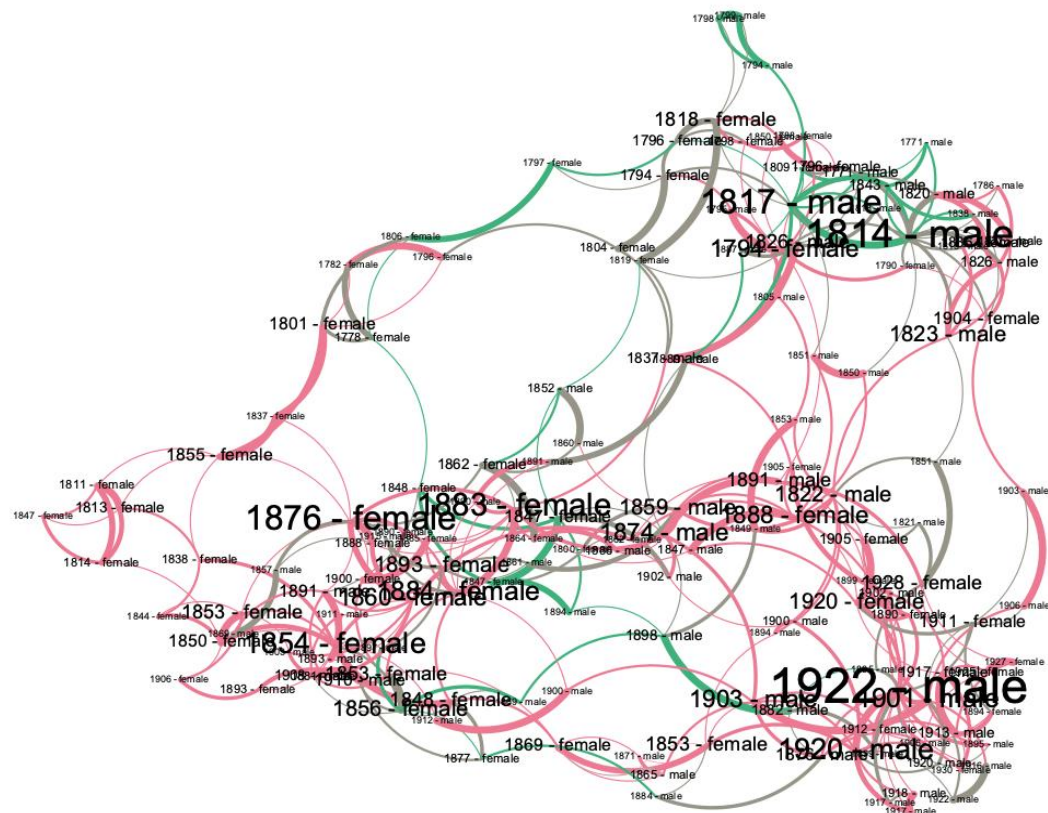
The dataset analyzed in this study is visualized through three graphs which provide insights into the year of publication, author gender and narrative perspective. Each node in these visualizations represents a novel with its size indicating the degree of centrality; denoting how influential it is within the community or a cluster based on the number of connections it has with other nodes. The positioning of nodes also highlights their interconnectivity, while edge thickness signifies a strength between two novels regarding similarities such as genre or language choice among others. Hence, thicker edges imply stronger connection implying greater similarity between them.

Provided by the dataset, we visualized the following three graphs according to author gender, publication year, and narrative perspective. Each node in these three visualizations symbolizes a novel, and its size indicates how central it is to the community or cluster, or how many connections it has with other nodes. The larger the node, the more influential the novel is in the community or cluster. The node's position also suggests that the novel is more interconnected than usual. The strength of the connection between the nodes is represented by a node's edge and weight. The relationship between two novels is stronger, the higher coarser the weight of the edge between them.



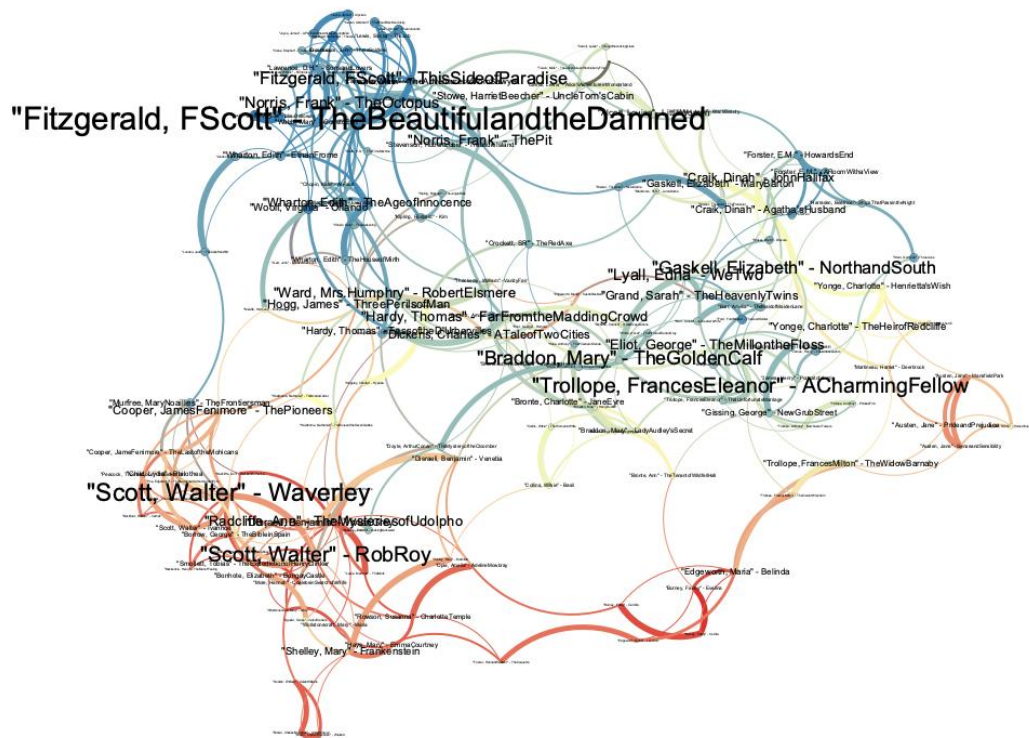
Graph 1. Network Analysis of relationships among British novels between 1771 and 1930 categorized by gender

The distribution of male and female authors is shown in Graph 1 together with labels for the publication dates of their respective novels. Male authors are represented by the blue nodes, while female authors are represented by the pink nodes. The results show that male authors are slightly more than female authors, according to the color distribution. It might also be contended that male writers generate more influential novels than female writers because they also tend to publish more books that are central to their individual societies. It also implies that female writers are becoming more influential over time when combined with the labels for the year of publication. The results also demonstrated that there is a connection between the authors of the same gender. We can also observe that the blue and pink nodes are highly concentrated throughout the network, constituting their own communities. Books written by male authors are similar to each other in that of writing style and use of words, and the books written by female authors also have similarities with each other.



Graph 2. Network Analysis of relationships among British novels between 1771 and 1930 categorized by narrative perspective

Graph 2 displays the point of view of each novel's narrative, with labels specifying the year of publication and the writer's gender. First-person narrative perspectives are shown in cyan on the graph, third-person are shown on pink nodes. The graph demonstrates how novels written in the first person had total centrality towards the end of the 18th century, whereas novels written in the third person had taken control by the start of the 20th century. It also demonstrates that female authors were comparatively more likely to use the third-person point of view in the narrative process of producing their works when combined with the labels of the writer's gender.



Graph 3. Network Analysis of relationships among British novels between 1771 and 1930 categorized by year

Based on the year of release, Graph 2 depicts a network analysis of these 150 British books. Red denotes a publishing year that is closer to 1771, while blue denotes a publication year that is closer to 1930. We discovered that several novels had a dominant position in the network after analyzing the relationships between the novels in the network. They were often the books in the network that were the most connected to one another. For instance, we found that F. Scott Fitzgerald's novel *The Beautiful and Damned* had the highest degree of connection to other novels in the network, followed by Walter Scott's *Waverley* and *Rob Roy*. In addition, no nodes are as noticeable in the center of Figure 2 as they are in the other two communities. We can speculate that the British writers of the late 19th and early 20th centuries may have been developed into different genres, while the clusters are not necessarily mutually exclusive, and various writers of this era were influenced by and associated with multiple groups.