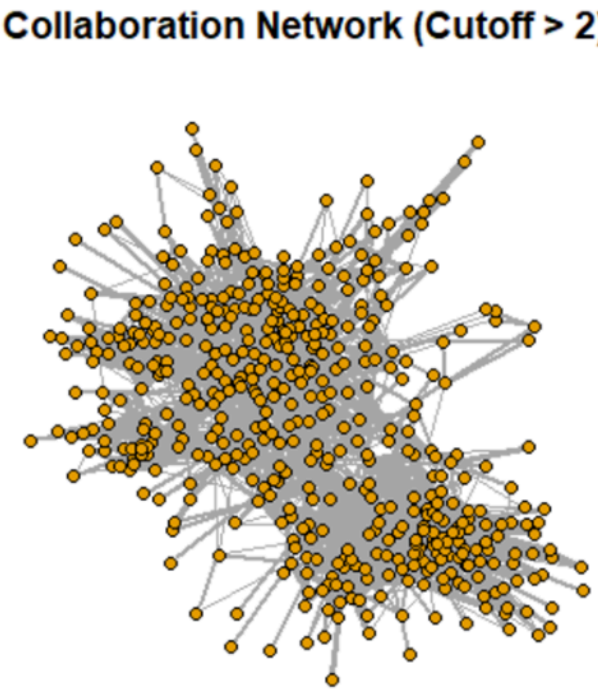
Advanced social network analysis, assignment 1

Grotto Tommaso

1. Network visualization

The Borgatti_Scientists504 dataset represents a collaboration network among scientists. I first applied the cutoff at >2 on the network and this is a visualization of it.

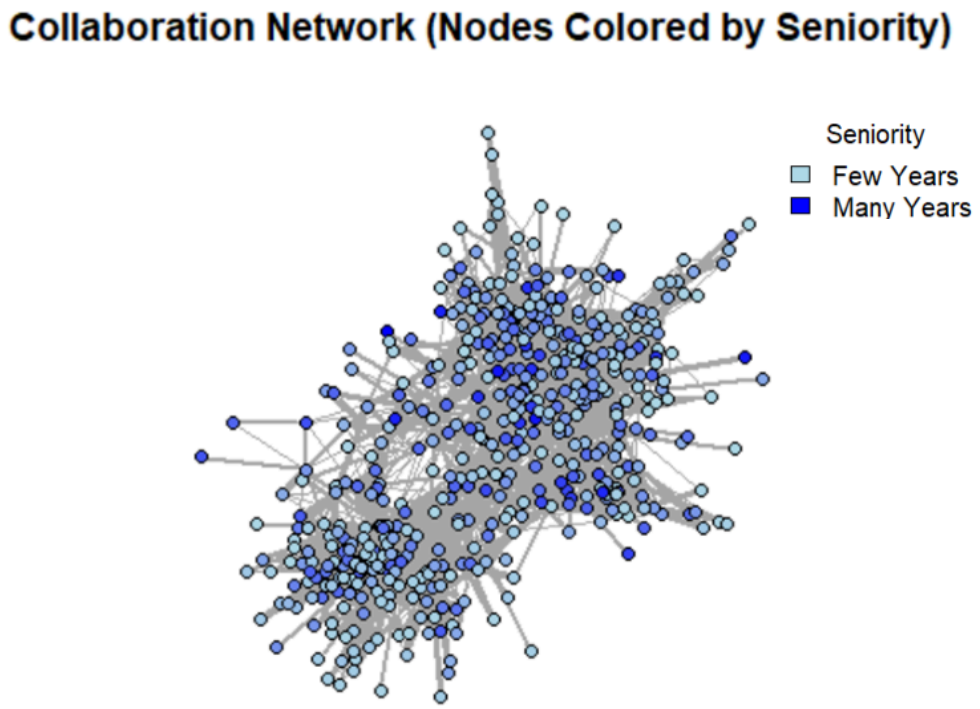


Graph 1: Collaboration network with a cutoff > 2

The network appears sparse, with no immediately apparent patterns.

2. Network visualization with seniority attribute

In this section, the collaboration network filtered by the >2 cutoff was further explored by incorporating the attribute seniority, measured as the number of years each scientist has been employed. To represent seniority visually, node colors were mapped to the number of years worked. A gradient color palette ranging from light blue (indicating fewer years) to dark blue (indicating many years) was used to encode this attribute.



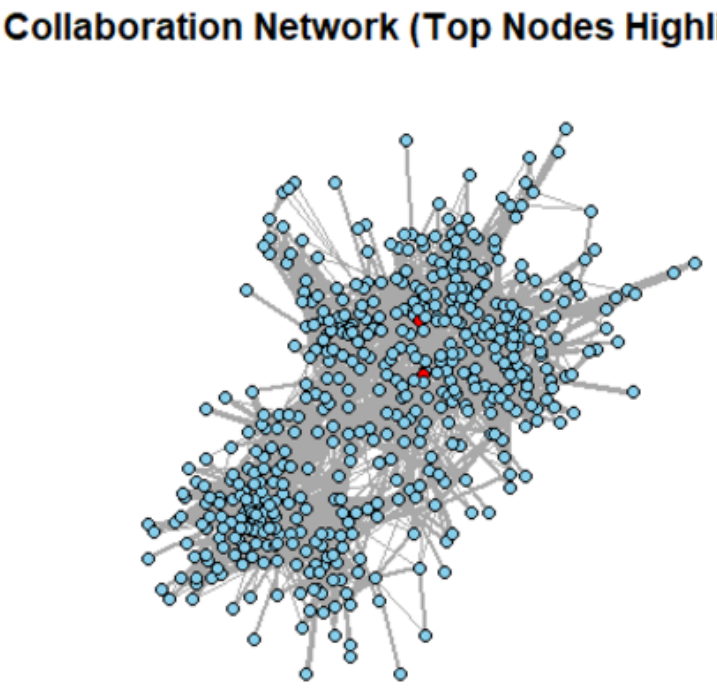
Graph 2: Collaboration network with a cutoff > 2 and nodes colored by seniority

The visualization reveals that peripherical nodes tend to have fewer years, but there seems not to be a very clear distribution of seniority within the network, suggesting that seniority alone does not fully explain a node's position in the network.

3. Centrality measures and key nodes

To better understand the structure of the collaboration network, two measures of position, betweenness centrality and eigenvector centrality, were calculated using the >2 cutoff network. These measures were selected because they capture complementary aspects of a node's role within the network. Betweenness centrality identifies nodes that serve as critical bridges between different groups, emphasizing their importance in maintaining connectivity. In contrast, eigenvector centrality measures the influence of a node based on its connections to other influential nodes, emphasizing integration within highly connected clusters.

The analysis revealed two standout nodes: T0266 emerged as the top node by betweenness centrality with a value of 14815.17. This indicates that T0266 plays a pivotal role in connecting disparate parts of the network. By acting as a bridge, this node facilitates collaboration across otherwise disconnected clusters, making it essential for the network's cohesion. T6111 had the highest eigenvector centrality, with a value of 1, signifying that it is the most influential node in the network. This suggests that T6111 is embedded in a dense and highly interconnected cluster of scientists, amplifying its importance through associations with other prominent nodes.

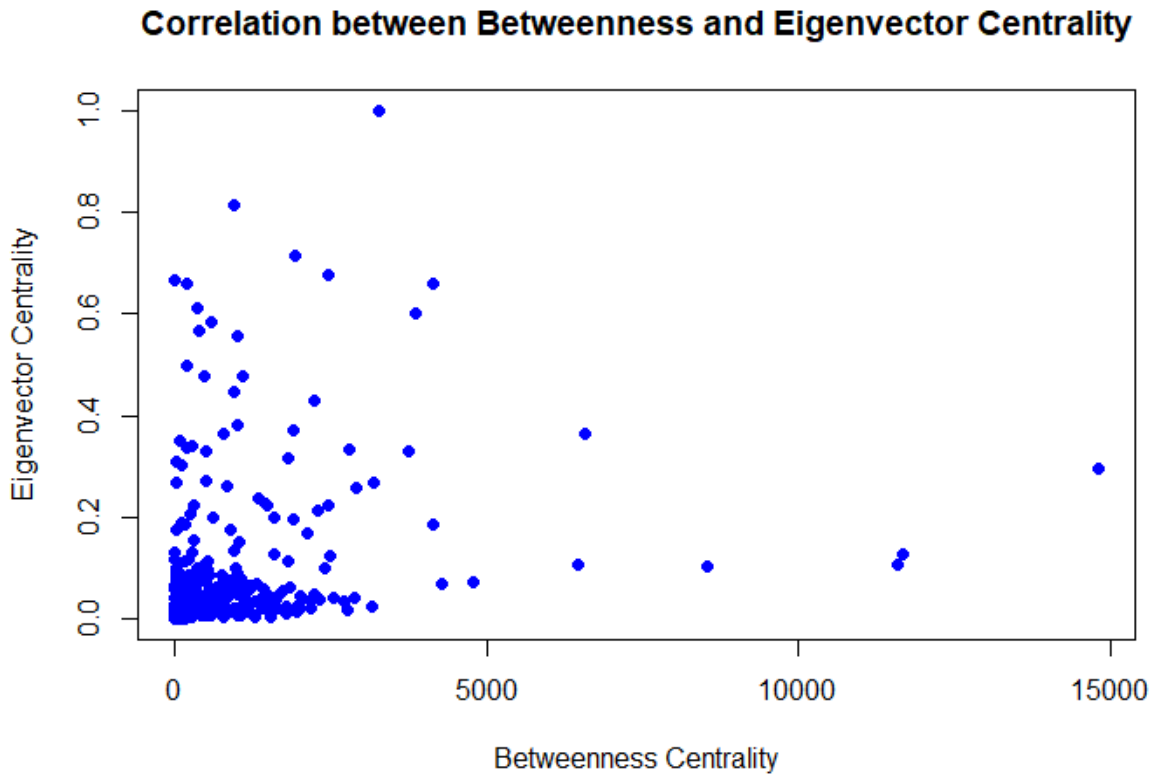


Graph 3: Collaboration network with a cutoff > 2 where the top nodes are highlighted

The visualization highlights the top nodes inside the collaboration network, showing their centrality.

4. Correlation between centrality measures

To explore the relationship between the two centrality measures, betweenness centrality and eigenvector centrality, a Pearson correlation analysis was conducted. The analysis resulted in a correlation coefficient of 0.3010, indicating a weak positive linear relationship between the two measures. While there is some alignment, nodes with higher Betweenness Centrality are slightly more likely to also have higher Eigenvector Centrality, the relationship is not strong. This suggests that these two metrics capture distinct aspects of a node's role within the network.



Graph 4: Correlation between betweenness centrality and eigenvector centrality in the collaboration network with a cutoff > 2

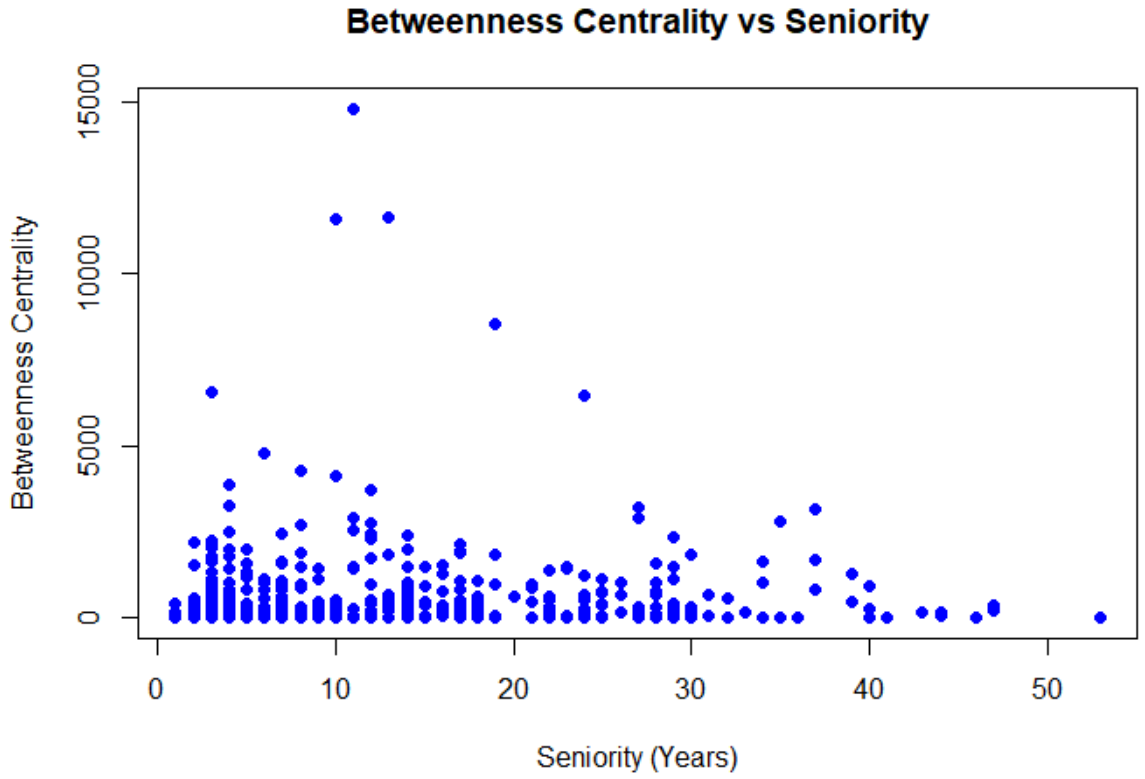
The scatterplot provides a visual representation of this relationship. Most nodes are clustered near the lower ends of both centrality measures, reflecting their peripheral positions in the network. However, a few extreme points with high values in either betweenness or eigenvector centrality stand out.

The weak correlation highlights the distinction between the two centrality metrics: betweenness centrality identifies nodes critical for maintaining the overall connectivity of the network, often acting as bridges between clusters, and eigenvector centrality emphasizes influence within highly connected and cohesive parts of the network.

5. Correlation between centrality measures and seniority

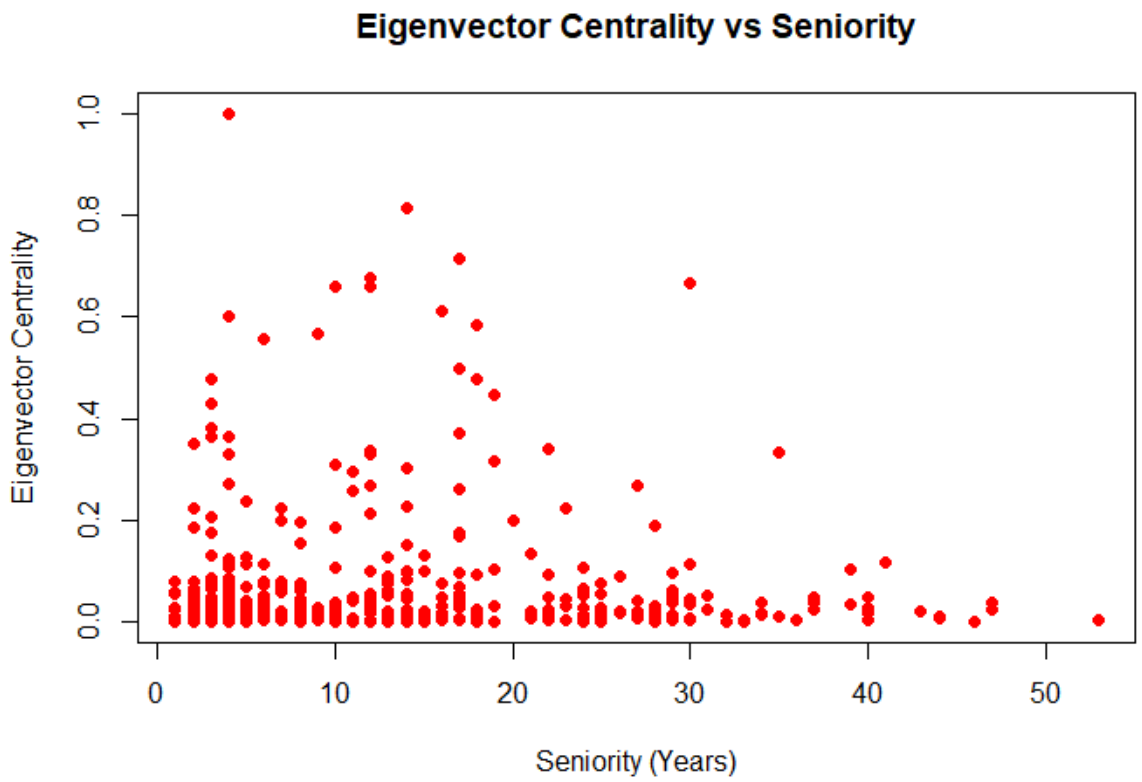
To investigate how network position relates to seniority, correlation coefficients were computed between seniority and the two centrality measures, betweenness centrality and eigenvector centrality. The results provide insights into whether a scientist's experience influences their role within the collaboration network.

The analysis revealed a correlation of 0.0351 between betweenness centrality and seniority, and a correlation of 0.0162 between eigenvector centrality and seniority. Both values indicate an almost negligible relationship between seniority and centrality.



Graph 5: Correlation between betweenness centrality and seniority in the collaboration network with a cutoff > 2

The scatterplot shows no clear trend, with senior scientists and junior scientists appearing at various levels of betweenness. This suggests that the ability to act as a bridge between different parts of the network is not strongly tied to years of experience.



Graph 6: Correlation between eigenvector centrality and seniority in the collaboration network with a cutoff > 2

Similarly, the lack of any discernible trend in this graph reinforces the conclusion that seniority does not significantly affect a node's influence within highly connected clusters. Both senior and junior scientists are found among the nodes with the highest Eigenvector values, suggesting that integration into influential groups is not determined solely by years of employment.

In conclusion, these findings demonstrate that seniority has little to no impact on a scientist's centrality in the collaboration network. Neither the role of a bridge (betweenness centrality) nor the influence within connected clusters (eigenvector centrality) is significantly tied to the number of years worked. This indicates that factors such as interdisciplinary collaboration, departmental roles, or proactive networking likely play a more critical role in determining a scientist's position in the network.