Structural Analysis of Interdepartmental Collaborations: Measure of Positions and Departmental Groupings

Assignment 1

Agnese Cervino (agnese.cervino@studenti.unitn.it)
Damiano Orlandi (damiano.orlandi@studenti.unitn.it)
Maria Amalia Pelle (mariaamalia.pelle@studenti.unitn.it)
Maria Starodubtseva (mariia.starodubtseva@studenti.unitn.it)

Introduction

In this assignment, we focus on analyzing the `Borgatti_Scientists504` dataset, which represents a collaboration network among 504 scientists. This dataset captures the duration of collaborative efforts between pairs of scientists, offering a unique opportunity to explore how network position is associated with individual and organizational attributes.

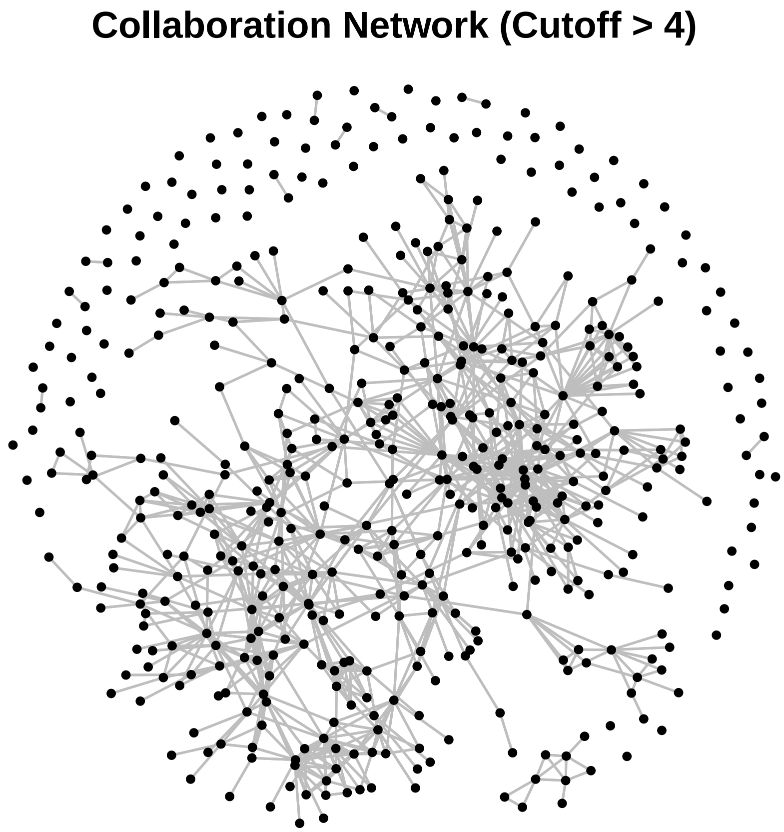
In the sections below, we present the results of our analysis, which include five main parts:

1. We apply a cutoff threshold to filter the collaboration network, focusing on substantial collaborations lasting more than 4 days.
2. We filter the network to analyze collaborations involving scientists from specific departments. Specifically, we group scientists from the Economics and Behavioral Science departments into one category, while scientists from all other departments are grouped into a separate category.
3. We calculate two measures of position, degree and betweenness centrality, to assess the structural positions of scientists within the network. We identify the two scientists showing the highest values for these measures and comment potential factors contributing to their prominence.
4. We perform a correlation analysis to explore the relationship between the measures of position.
5. Finally, we also investigate the correlation between measures of positions and the departmental classification of scientists into the two categories mentioned above.

Part 1

Given the large number of nodes in the network, we followed the professor's guidelines and applied a cutoff **threshold of** $>$ 4 days of collaboration. This threshold ensures a more focused analysis by emphasizing substantial collaborations.

The plot below visualizes the collaboration network after applying the cutoff threshold. The nodes represent scientists, while the edges denote collaborations. It is important to consider that nodes that are not linked in the plot are sensitive to the cutoff, meaning that those scientists might have worked together but less than 5 days.

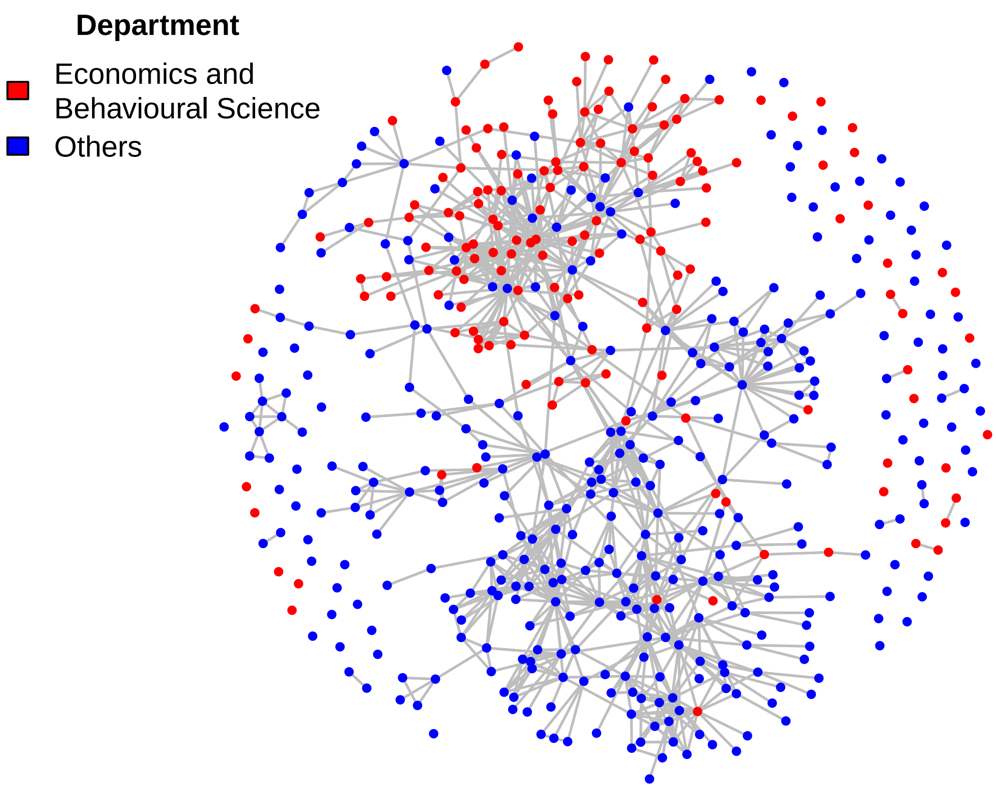


We can notice two separate clusters of nodes that are interconnected. These clusters likely represent groups of individuals who collaborate frequently within their groups but have less collaborations with members outside their clusters.

Part 2

Additionally, for the attribute analysis, we code the department variable such that scientists in "Economics" and "Behavioural Science" departments are grouped as 1, while all other departments are coded as 0. This binary attribute enables us to investigate whether structural network positions correlate with this departmental classification.

The plot below visualizes the collaboration network after applying the cutoff threshold and filtering for scientists in the "Economics" and "Behavioural Scientists" departments. The red nodes represent scientists in these departments, while the blue nodes represent scientists in other departments.



The visualization shows the presence of two clear communities, one representing scientists from the "Economics" and "Behavioural Science" departments and another - scientists from other departments. The red nodes form a relatively dense cluster, showing that scientists in these departments collaborate extensively with each other. Meanwhile, the blue nodes form a more sparse group, indicating a different collaborative approach.

There is only a small amount of blue nodes in the red cluster, potentially suggesting little interdisciplinary collaboration between "Economics" or "Behavioral scientists" with other scientists.

Part 3

In this section we calculate two measures of position in the network: **degree** and **betweenness centrality**.

We choose these two measures because:

- **Degree** captures the number of direct connections a node has, providing insights into the node's visibility and influence within the network. This measure is particularly relevant for our analysis as it helps identify scientists who are well-connected and potentially influential in the collaboration network.

$$\text{Degree} = \text{Number of direct ties}$$

- **Betweenness centrality** measures the extent to which a node lies on the shortest paths between other nodes, indicating the node's potential to control information flow. This measure is important for identifying scientists who act as bridges between different groups or clusters within the network. In the previous section, we saw that the network has two main clusters, making it interesting to explore which scientists act as bridges between these clusters.

$$\text{Betweenness centrality} = \frac{\text{Number of shortest paths passing through the node}}{\text{Total number of shortest paths}}$$

Degree

After computing the degree for each node, we calculate the average degree for the "Economics and Behavioral Science" department (1) and for the others (0).

```
##        0        1
## 5.814286 5.616883
```

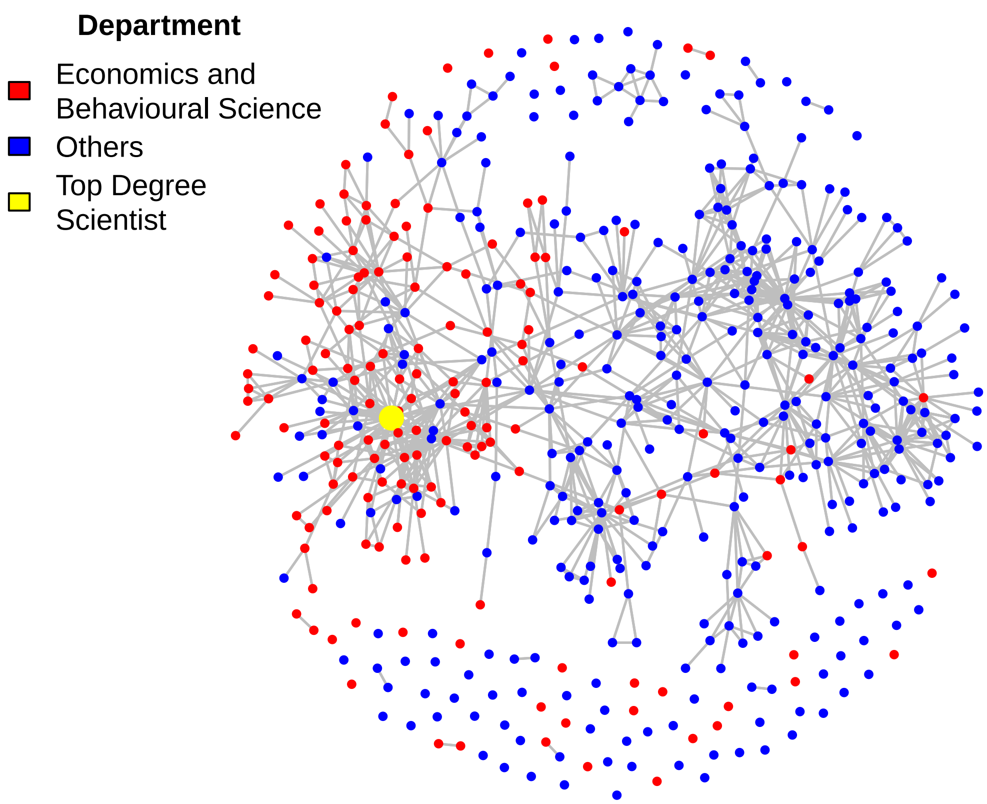
The results indicate that the average degree of the group 0 is higher, even though the graph suggests that blue nodes appear more sparsely connected. This discrepancy can be attributed to the fact that average degree does not account for the total number of potential connections, which is greater in the blue group due to its larger size. By calculating the density, it becomes evident that red nodes form a more tightly connected network compared to the blue nodes.

```
## Density for Class 1 (Econ. and Behav.): 0.02945421
## Density for Class 0 (Others): 0.01525993
```

In order to check what is the node with the highest degree in the whole network, we sort them in descending order and print just the first one with its attributes.

```
##     NodeName Years Sex DeptID IsIn Degree
## 264    N1514    24   1      1    1     35
```

The scientist with the highest degree value in the network is **N1514**, who has a degree of **35**, indicating collaboration with 35 other scientists for periods of 4 days or longer. This person is female, has been working for 24 years and is from Behavioral Science department. Her high degree value is probablu due to her extensive experience, placing her in the top quartile (17.46%) for years of experience. Her position is enlighten in the plot below.

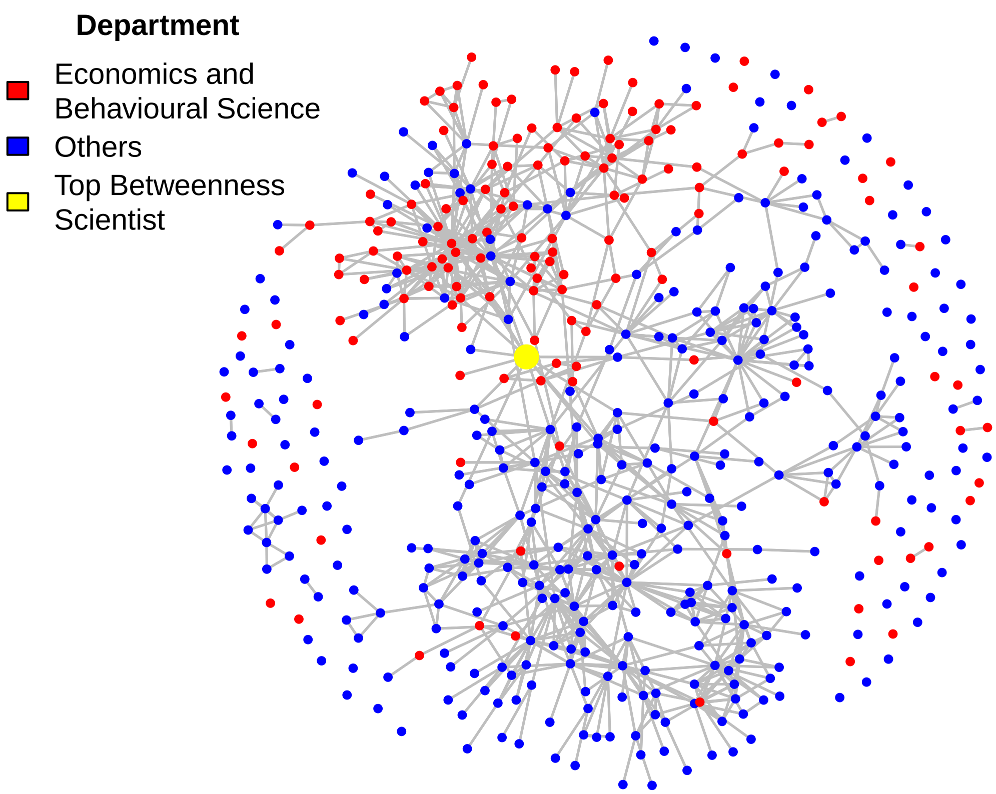


Betweenness centrality

```
##      NodeName Years Sex DeptID IsIn Betweenness
## 259    N0111    19   2     20    0   0.1192227
```

The scientist with the highest betweenness centrality in the network is **N0111**, with a centrality value of **0.1192227**. That means that 11.92% of all the shortest paths between pairs of scientist in the network pass through this individual. Betweenness centrality is a measure of a node's importance in facilitating communication within a network. A high value indicates that the node is a crucial bridge for information flow.

In this case, **N0111's** high betweenness centrality is particularly relevant because it suggests that he serves as a key connector between "Economics and Behavioral Science" and "Others". His demographic and professional attributes — being male, with 19 years of experience, and from a different department — may further highlight his role as the main link between diverse areas of expertise or organizational divisions. This bridging position can influence collaboration and knowledge exchange. His role is illustrated in the plot below, where his position visually reinforces his strategic importance within the network.



Part 4

Proceeding with the analysis, we correlate the degree and betweenness centrality measures to explore how these two measures relate to each other.

The correlation between degree and betweenness centrality is 0.67, indicating an expected strong positive relationship between the two measures. Nodes with a high degree have more connections, which means they are involved in more paths between other nodes. As a result, they are more likely to act as intermediaries or bridges in the network, increasing their betweenness centrality.

Part 5

Finally, we investigate the relationship between the departmental classification and the measures of positions.

By calculating and correlating the measures with the attribute, we could discover patterns or dependencies between the network structure (how nodes are positioned within the network) and node-specific characteristics. This could provide insights into whether certain departments are more central or influential in the network.

The correlation between departmental classification and degree is -0.02, while the correlation between departmental classification and betweenness centrality is -0.06. Both values are so low that they indicate no meaningful correlation between the measures and the attribute under consideration.

Conclusions

This study analyzed the collaboration network of 504 scientists, revealing key patterns in interdepartmental interactions through measures of positions and departmental classifications. It identified two main collaboration clusters, with stronger internal ties in the "Economics and Behavioral Science" departments and more dispersed connections in other fields. High-degree nodes represented influential individuals within groups, while high-betweenness nodes served as bridges between clusters. The analysis found a positive correlation between degree and betweenness centrality, but no significant association with departmental affiliation. In conclusion, the "Economics and Behavioral Science" group showed stronger internal network density.

Appendix

```r
# Preparation
knitr::opts_chunk$set(warning = FALSE,
                      message = FALSE,
                      tidy.opts = list(width.cutoff = 80),
                      tidy = TRUE)
                      #fig.align = "center")

library(igraph)
library(sna)
library(network)
setwd("C:/Users/djagn/Desktop/Università/Advanced Social Network Analysis")
load("Borgatti_Scientists504.RDA")

# Part 1
filtered_list <- ifelse(Borgatti_Scientists504$Collaboration > 4, 1, 0)
net <- as.network(filtered_list, directed = F)

par(mar=c(0,0,1,0))
gplot(net,
      gmode = "graph",
      vertex.cex = .8,
      displaylabels = FALSE,
      edge.col = "grey",
```

```r
      vertex.col = "black",
      main = "Collaboration Network (Cutoff > 4)")

# Part 2
Borgatti_Scientists504$Attributes$IsIn <- ifelse(
  Borgatti_Scientists504$Attributes$DeptID == 1 |
  Borgatti_Scientists504$Attributes$DeptID == 2,
  1,
  0
)

par(mar = c(0, 0, 1, 0))
node_colors <- ifelse(Borgatti_Scientists504$Attributes$IsIn == 1,
                      "red", "blue")
gplot(net,
      gmode = "graph",
      vertex.cex = .8,
      displaylabels = FALSE,
      edge.col = "grey",
      vertex.col = node_colors,
      vertex.border = node_colors)
legend("topleft",
       legend = c("Economics and \nBehavioural Science", "Others"),
       fill = c("red", "blue"),
       title = "Department",
       title.font = 2,
       bg = "transparent", cex = 0.9, box.col = "transparent")

#Part 3
# Degree and avg degree
net.graph <- igraph::graph_from_adjacency_matrix(filtered_list,
                                                  mode = "undirected")
degree_values <- igraph::degree(net.graph)
node_classes <- Borgatti_Scientists504$Attributes$IsIn
average_degree <- tapply(degree_values, node_classes, mean)
print(average_degree)
# Density
V(net.graph)$class <- node_classes
subgraph_class1 <- induced.subgraph(net.graph, V(net.graph)[class == 1])
subgraph_class0 <- induced.subgraph(net.graph, V(net.graph)[class == 0])
density_class1 <- graph.density(subgraph_class1)
density_class0 <- graph.density(subgraph_class0)
cat("Density for Class 1 (Econ. and Behav.):", density_class1,
    "\nDensity for Class 0 (Others):", density_class0)
```

```r
# Sorted degree
sorted_degree <- sort(degree_values, decreasing = TRUE)
top_node_index <- order(degree_values, decreasing = TRUE)[1]
top_node_attributes <- Borgatti_Scientists504$Attributes[top_node_index, ]
top_node_attributes$Degree <- degree_values[top_node_index]
print(top_node_attributes)
# Compute percentage of nodes showing same or highest amount of experience
sort_exp <- sort(Borgatti_Scientists504$Attributes$Years, decreasing = TRUE)
top_perc <- (1-sum(sort_exp<24)/length(sort_exp))*100
top_perc
# Plot node with highest degree
par(mar = c(0, 0, 1, 0))
node_sizes <- rep(0.8, length(node_colors))
node_sizes[top_node_index] <- 2.5
node_colors <- ifelse(Borgatti_Scientists504$Attributes$IsIn == 1,
                      "red", "blue")
node_colors[top_node_index] <- "yellow"
gplot(net,
      gmode = "graph",
      vertex.cex = node_sizes,
      displaylabels = FALSE,
      edge.col = "grey",
      vertex.col = node_colors,
      vertex.border = node_colors)
legend("topleft",
       legend = c("Economics and \nBehavioural Science",
                  "Others",
                  "Top Degree \nScientist"),
       fill = c("red", "blue", "yellow"),
       title = "Department",
       title.font = 2,
       bg = "transparent", cex = 0.9, box.col = "transparent")
# Compute betweenness
betweenness_values <- igraph::betweenness(net.graph,
                                          directed = FALSE,
                                          normalized=TRUE)
# Best betweenness
sorted_betweenness <- sort(betweenness_values, decreasing = TRUE)
top_node_index <- order(betweenness_values, decreasing = TRUE)[1]
top_node_attributes <- Borgatti_Scientists504$Attributes[top_node_index, ]
top_node_attributes$Betweenness <- betweenness_values[top_node_index]
print(top_node_attributes)
# Plot best node
par(mar = c(0, 0, 1, 0))
```

```r
node_sizes <- rep(0.8, length(node_colors))
node_sizes[top_node_index] <- 2.5
node_colors <- ifelse(Borgatti_Scientists504$Attributes$IsIn == 1,
                      "red", "blue")
node_colors[top_node_index] <- "yellow"
gplot(net,
      gmode = "graph",
      vertex.cex = node_sizes,
      displaylabels = FALSE,
      edge.col = "grey",
      vertex.col = node_colors,
      vertex.border = node_colors)
legend("topleft",
       legend = c("Economics and \nBehavioural Science",
                  "Others",
                  "Top Betweenness \nScientist"),
       fill = c("red", "blue", "yellow"),
       title = "Department",
       title.font = 2,
       bg = "transparent", cex = 0.9, box.col = "transparent")

# Part 4
correlation_result <- cor(degree_values, betweenness_values)

#Part 5
cor_degree <- cor(Borgatti_Scientists504$Attributes$IsIn, degree_values)
cor_betweenness <- cor(Borgatti_Scientists504$Attributes$IsIn,
                       betweenness_values)
```