

# Project 5

Logan Bolton

2025-02-23

*Acknowledgement:* This code was created through the repurposing of code found in the lecture notes and through collaboration with Claude 3.5 Sonnet and o3-mini. These AI tools were very helpful for me while fixing errors and determining the correct syntax to plot graphs.

```
url1 <- "https://raw.githubusercontent.com/JeffreyAlanSmith/Integrated_Network_Science/master/data/affi
url2 <- "https://raw.githubusercontent.com/JeffreyAlanSmith/Integrated_Network_Science/master/data/affi

affiliations96 <- read.delim(file = url1, check.names = FALSE)
affiliations97 <- read.delim(file = url2, check.names = FALSE)
dim(affiliations96)

## [1] 1295  91
```

## 1 - 1996 Dataset

a - Which student clubs serve to integrate the school and which are more peripheral?

b - Which student clubs tend to share members at high rates?

c - What is the shared feature, or theme, that brings these clubs together in a cluster?

## 2 - 1997 Dataset

```
G <- graph_from_incidence_matrix(as.matrix(affiliations97))

## Warning: `graph_from_incidence_matrix()` was deprecated in igraph 1.6.0.
## i Please use `graph_from_biadjacency_matrix()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

order_G <- gorder(G)
size_G <- gsize(G)
density_G <- edge_density(G)

cat("Order (number of vertices):", order_G, "\n")

## Order (number of vertices): 1386

cat("Size (number of edges):", size_G, "\n")

## Size (number of edges): 2641
```

```
cat("Density:", density_G, "\n")
```

```
## Density: 0.002751601
```

a - What is the order, size, and density of G?

```
is_connected(G)
```

```
## [1] FALSE
```

```
cat("The graph is NOT connected\n\n")
```

```
## The graph is NOT connected
```

```
comp <- components(G)
```

```
# Size of largest component
```

```
largest_comp_size <- max(comp$size)
```

```
fraction_largest <- largest_comp_size / gorder(G)
```

```
# Create subgraph H
```

```
H <- induced_subgraph(G, which(comp$membership == which.max(comp$size)))
```

```
cat("Percent of vertices belonging to largest component: ", fraction_largest, "\n")
```

```
## Percent of vertices belonging to largest component: 0.6789322
```

b - Is the network G connected? If not, what fraction of vertices belong to the largest connected component? If the network is not connected, consider only the largest component H for the remaining questions.

```
average_path_length_H <- average.path.length(H, directed = FALSE)
```

```
## Warning: `average.path.length()` was deprecated in igraph 2.0.0.
```

```
## i Please use `mean_distance()` instead.
```

```
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was  
## generated.
```

```
average_path_length_H
```

```
## [1] 3.738472
```

c - What is the average path length of H?

d - Is H scale-free? Provide statistical evidence (e.g., by examining the degree distribution and fitting a power-law distribution)

e - What is the fraction of edges that are attached to the top 10% of high-degree vertices?

f - What distributions do the following centrality measures follow:

Eigenvector centrality

Betweenness centrality

Closeness centrality

g - How does the clustering coefficient of vertices change with vertex degrees?

h - Does H exhibit assortative mixing in terms of vertex degrees? Provide the assortativity coefficient and interpret its value.