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/affil
url2 <- "https://raw.githubusercontent.com/JeffreyAlanSmith/Integrated_Network_Science/master/data/affil

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

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

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.

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.