Project 5

This project focuses on two-mode network analysis and large scale structure of networks.

You will work with the affiliation data collected by Daniel McFarland on student extracurricular affiliations. It is a longitudinal data set, with 3 waves - 1996, 1997, 1998, although you will only use the second wave, 1997.

1) Perform a network analysis similar to the example shown in Chapter 7 using the 1996 dataset, and answer each of the following questions.

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?

```
library('igraph')
```

```
##
## Attaching package: 'igraph'
```

```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

```
## The following object is masked from 'package:base':
##
##     union
```

```
url1 <- "https://github.com/JeffreyAlanSmith/Integrated_Network_Science/raw/master/data/affiliations_19

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

dim(affiliations97)
```

```
## [1] 1295   91
```

2) Consider the student-to-student network G projected from the 2-mode network generated from the 1997 dataset. Answer each of the following questions with statistical verification:

a) Basic Network Properties:

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

b) Connectivity:

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) Average Path Length:

What is the average path length of H?

d) Scale-Free Property:

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

e) Edge Concentration:

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

f) Centrality Distributions:

What distributions do the following centrality measures follow:

Eigenvector centrality

Betweenness centrality

Closeness centrality

g) Clustering Coefficient vs. Degree:

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

h) Assortative Mixing:

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