Module 2 - Assignment

Start Assignment

- Due Sunday by 11:59pm
- Points 60
- Submitting a text entry box or a file upload
- Attempts 0
- Allowed Attempts 1

This assignment will provide practical experience conducting social network analysis (SNA) on a toy data set from the popular television series *Grey's Anatomy*. The data set was pulled together by Garry Weissman in 2011. All the nodes in the data set represent characters in the show. The links between nodes represent sexual contact between the characters. There is also an attributes file with gender and race included. This is a common application of SNA in understanding sexually transmitted disease and associated epidemics.

Download R and RStudio following the instructions, **https://www.rstudio.com/** ⤷ **(https://www.rstudio.com/)**. Note you should be downloading the free version for your computer. There is a tutorial video on RStudio installation and basics if you need it, **RStudio Tutorial For Beginners.** ⤷ **(https://www.youtube.com/watch?v=mcYcjH-1giM)**

RStudio Tutorial For Beginners | RStudio In...

Minimize Video

Part A (10%)

Use the provided R script for generating a toy 3-node network and basic ergm (lines 12-18).

Sample R Script = **Grays Anatomy.R (https://jhu.instructure.com/courses/83281/files/11341299?wrap=1)** ↓ **(https://jhu.instructure.com/courses/83281/files/11341299/download?download_frd=1)**

Part B (50%)

Use the provided R script for ingesting the provided Gray's Anatomy data. Note there are two files, one for nodes & links and the other for attributes. (lines 20-46).

Nodes & Links = **gray-adj.csv (https://jhu.instructure.com/courses/83281/files/11341300?wrap=1)** ↓ **(https://jhu.instructure.com/courses/83281/files/11341300/download?download_frd=1)**

Attributes = **gray-attr.csv (https://jhu.instructure.com/courses/83281/files/11341291?wrap=1)** ↓ **(https://jhu.instructure.com/courses/83281/files/11341291/download?download_frd=1)**

Use the tutorial file ONA-in-R from your readings to calculate the following:

1. Which nodes have the highest five betweenness centrality scores and report the values.
2. Which nodes have the highest five degree centrality scores and report the values.
3. Plot the two centrality scores against each other (e.g. betweenness on y-axis, degree on x-axis).
4. Generate a plot of the social network where nodes are sized based on a centrality measure of your choice, colored based on an attribute of your choice, and only the labels for the top 5 node (based on some criteria) are displayed. Include a legend.

Part C (40%)

Use the provided R script and Gray's Anatomy data to conduct an exponential random graph (ERGM) analysis of the data to answer the following questions:

1. Does racial homophily or heterophily exist in the data?
2. Does gender homophily exist in the data?
3. What does gender homophily mean in layman's terms (not a coding/analytic question).

Note that for running ERGMs on this data set, the triadic term is very likely to cause convergence issues. In this case, you may choose to omit the term 'triangle'. If you think about what a triad means in a sexual contact network, it might be reasonable to question its appropriateness.

SUBMIT your revised R Script with comments **and** a pdf showing your outputs/plots/response to the above questions. Make sure you provide written explanation that interprets your findings from Part B and Part C for a layman with no knowledge of social network analysis.