**DATA 620 Week 4 Homework**

*Riguel Fabre & Lara Burford*

**Dataset**

We found a good network dataset on the UCI Network Data Repository website at the link: <https://networkdata.ics.uci.edu/netdata/html/divorce.html>. The data set is called the “Legal basis for divorce in the USA”, and it links each of the 50 states with each of 9 different legal bases for divorce including incompatibility, cruelty, desertion, nonsupport, alcohol, felony, impotence, insanity, and separation. The state is already categorical, but we thought we may add a region as well just to make the graph a little more interesting.

**Data Analysis**

The data format is in R, so we’ll load the data into R and then convert it to a csv file so we can read it into python for analysis. We’ll use GraphLab to produce the graph. Currently, the data is set up to be a bipartite graph that just links each state to every legal bases for divorce in that state. If we add the region, then we’ll group the states by region within the graph.

Hypothetically, we would be able to predict which state and/or region has the highest divorce rate based on the degree centrality of each state and/or region. That information is available on the CDC website at <http://www.cdc.gov/nchs/nvss/marriage-divorce.htm> so we can compare that data to our results.