A5. Network Data Analysis (Word Adjacency Data Set)

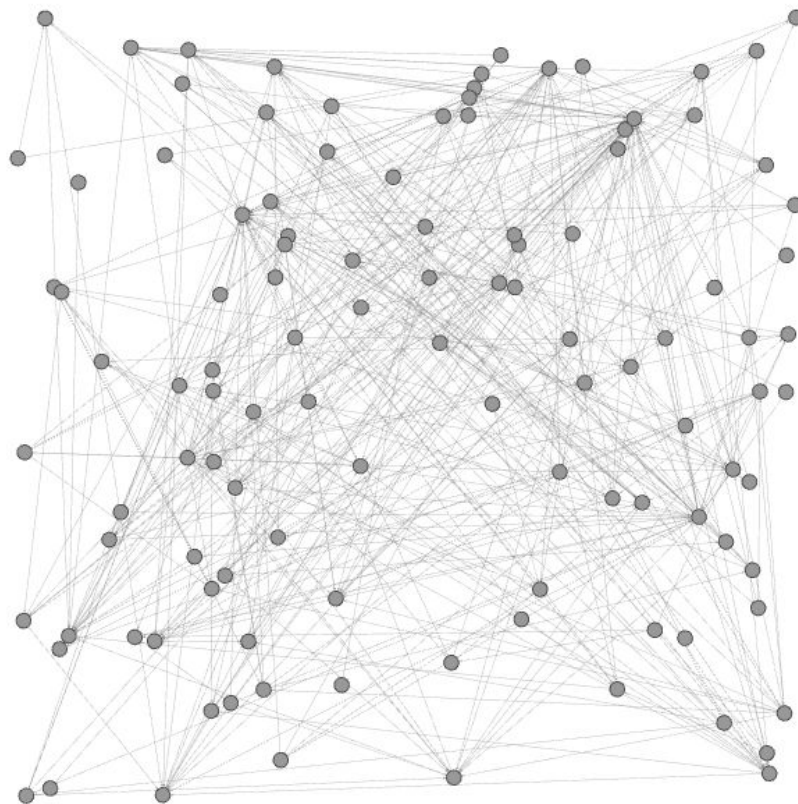## Greg Ceccarelli | March 31st, 2016

W209 - 4

For this assignment I chose to explore the adjacency network of common adjectives and nouns in the novel David Copperfield by Charles Dickens.
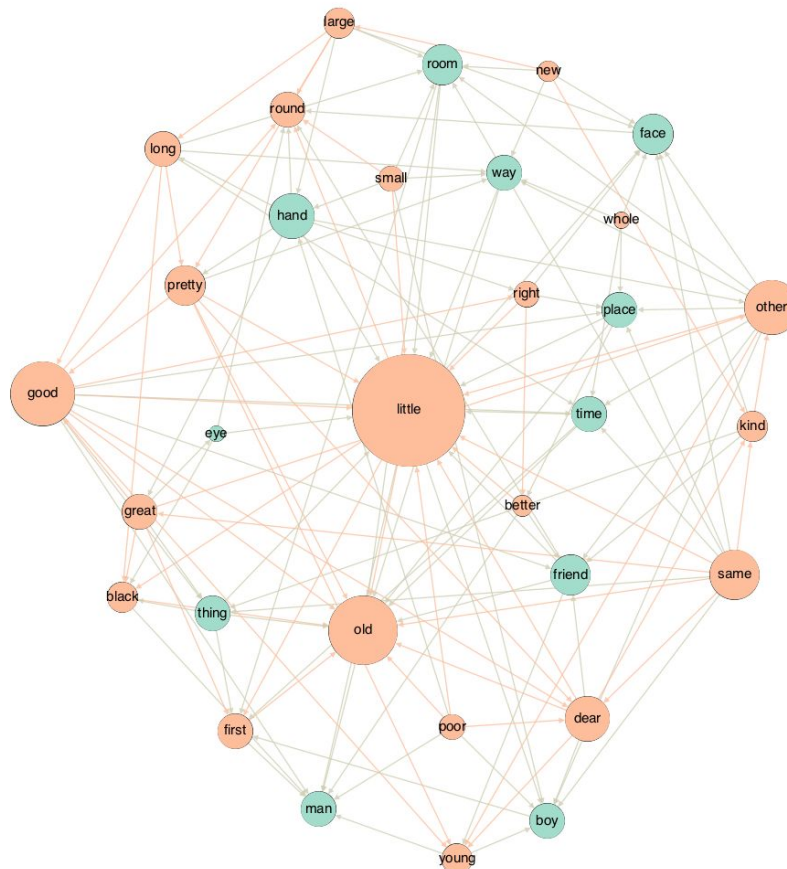
Upon import into gephi, the graph consisted of **111 nodes** and **424 edges.**

Figure 1: Unadulterated Graph
Clearly some work is required to make this interpretable!

Figure 2: Modified Graph

The modified graph above is the result of a number of revisions and iterations made in an attempt to determine the importance of nodes in the dataset.

Context
- Adjectives are colored melon whereas nouns are colored sea green
  - I attempted to color based on modularity class but this was more compelling
- Node sizing is based on relative total degree (inbound and outbound edges)
- Degree range filtering was applied to only show those nodes with > 10 connections
- Straight edges were preferred over curved in order to show edge arrows which are important for understanding in vs. out degrees and interpreting the adjective phrases

Key Insights:
- Little and old are both key adjectives comprising a great deal of node degree strength
  - Given their centrality, it's reasonable to infer, even without knowing the plot of David Copper, that the it is either a) about or b) contains a <u>dear old man</u> or a <u>small poor young little boy</u>
- Another interesting insight is that network reveals multiple inbound and outbound connections to the nouns place & time. Based on adjacency pairs, we can infer that the story is likely a retelling that covers a great bit of time, examples:
  - <u>Right place, right time;</u> <u>little time;</u> <u>whole place;</u> <u>whole time;</u> <u>other time</u> etc