Final Project Proposal

DATA620: Web Analytics

Daniel Dittenhafer & Justin Hink

December 11, 2016

1 Problem Statement & Research Questions

For the project we will analyze President-elect Trump's Cabinet nominations from a social and graph analytics perspective. We will look at the network of relationships between the individuals as well as the organizations in which they play a prominent role.

Additionally, we will attempt to unearth common themes in the descriptions of the above described entites from a textual analysis of the currated descriptions of them on Wikipedia.

2 Statement of Objectives

Using graph and text analytics, we will explore the individuals whom have been (or are expected to be) nominated for positions in President-elect Donald Trump's cabinet. Note: Since a number of these positions have yet to be finalized, these data are changing quite frequently. The project will be completed with data current as of 12/11/2016

3 Data Set

The data set is derived from an article written by Andrew Raffery and published by NBC News (Rafferty, 2016). We plan to augment this data set with web-based data including Wikipedia articles about each of the nominees and their organizations.

4 Evaluation

4.1 Graph Analysis:

Basic graph analysis metrics will be extracted along with visualizations of the following categories:

- Prior Organization
- Prior Organization Type
- Prior Position / Title

4.2 Text Analysis:

- A custom library will be written that scrapes raw wikipedia content and converts it into a visual display of the most commonly used words (known olloquially as a word cloud)
- Will leverage python packages: 'BeautifulSoup' and 'wordcloud'
- The accumulated text will be aggregated and analyzed for general language patterns (common metric such as Zipf's law)

5 References

Rafferty, A. Donald Trump's Cabinet Picks So Far. Dec. 2016. URL: http://www.nbcnews.com/politics/politics-news/donald-trump-s-cabinet-picks-so-far-n690296.