

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 several articles. The original article was written by Andrew Rafferty and published by NBC News (Rafferty, 2016). The Hill published information regarding the Secretary of the Interior (Henry, 2016), and the New York Times did the same for the Secretary of Labor (Schieber, 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)

4.3 Roles & Responsibilities

We have worked together on prior projects and plan to work collaboratively once again on this project. In general, Daniel will be leading the graph analysis, while Justin will lead the text analysis. Both will contribute at some level to all aspects of the project from this proposal and data set acquisition, to final project write up.

5 References

Henry, D. Trump to pick Rep. McMorris Rodgers for Interior secretary. Dec. 2016. URL: <http://thehill.com/policy/energy-environment/309628-trump-to-pick-rep-mcmorris-rodgers-for-interior-secretary-report>.

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>.

Schieber, N. Trump's Labor Pick, Andrew Puzder, Is Critic of Minimum Wage Increases. Dec. 2016. URL: http://www.nytimes.com/2016/12/08/us/politics/andrew-puzder-labor-secretary-trump.html?_r=0.