Title: A Comparison of Text Analysis and Social Network Analysis using Twitter Data
Authors: John Myles White (Princeton), Drew Conway (NYU)

Abstract

Within the social sciences, interest in statistical methods for the automatic analysis of text has grown considerably in recent years. Specifically, within the field of political science there has been continued interest in using text analysis to extend well-known spatial models of political ideology. Simultaneously, interest in how the structure of social networks affects political attitudes and outcomes has been both growing and controversial. In an attempt to bridge the gap between these two research areas we have collected a sample of Twitter data related to the U.S. Congress. Twitter provides a rich data set containing both text and social network information about its members. Here we compare the usefulness of text analysis and social network analysis for predicting the political ideology of Twitter users — two methods that are, in principle, applicable to both members of Congress (for whom roll call data and precise spatial estimates of political ideology already exist) and to the surrounding network of Twitter users (for whom precise estimates of political ideology do not exist). To compare text analysis methods with tools from social network analysis, we fit a variety of L1- and L2-regularized regression models that use word count data from individual tweets to predict the ideal points of Members of Congress. We then compare the performance of the resulting text models with the performance of social network models that employ techniques developed for predicting the spread of transmissible diseases to predict the ideal points for the same Members of Congress.

Contact:
John Myles White
Graduate Student
Department of Psychology and Princeton Neuroscience Institute
Princeton University
jmw1729@princeton.edu