

Benjamin Rothschild  
MACS 30200, Dr. Evans  
Problem Set #2  
Critical Review of Computational Methods

Justin Grimmer, “Measuring Representational Style in the House: The Tea Party, Obama and Legislators’ Changing Expressed Priorities,” in *Computational Social Science: Data and Prediction*, ed. R. Michael Alvarez, Cambridge University Press (2016).

**a) State the research question of your assigned paper.**

How do legislators change their representation of constituents in response to shifts in electoral pressure and changes in party control of Congress in the context of the 2008 election?

**b) What data did the paper use?**

The author uses Congressional press releases to measure how legislators present their work to constituents. He uses nearly 170,000 House press releases which is every press release from each House office from 2005 to 2010.

**c) What theory did the paper reference to interpret the data?**

The author cited several theories of topic modeling for unstructured text to justify the methods he uses for his analysis: An explanation of theories he sites are:

- “Topic models are an increasingly popular tool for studying large collections of text (Blei, Ng, and Jordan, 2003; Quinn et al., 2010)
- “[The author] uses a standard set of techniques, though slightly varying the recipe to account for the idiosyncratic features of Congressional press releases (Grimmer and Stewart)”.
- “This [article] uses a model first introduced in Blaydes, Grimmer, and McQueen (2014) that estimates two different sets of topics, similar to the nesting of topics in Panchenko Allocation (Li and McCallum, 2006).
- The author explains how it can be difficult to determine the number of topics in an unsupervised learning model (Grimmer and King, 2011) and that different analysis implies that a different number of topics are ideal (Chang et al., 2009). He addresses this problem by providing two sets of topics (granular and coarse)

He cites these theories to prove the point that topic modeling of unstructured text has been successful in the past. He also explains how the method used in this paper is ideal for his dataset. For example, to apply statistical models to the collection of press releases, he preprocesses the texts – representing its content as number and discards word ordering.

In addition, he cites several theories to aid in the interpretation of his dataset. He says:

- Communication is a central component of representation (Mansbridge, 2003; Disch, 2012). Legislators invest time and resources in crafting speeches in Congress, composing press releases to send to newspapers, and in distributing messages directly to their constituents (Yiannakis, 1982; Quinn et al., 2010; Lipinski, 2004; Grimmer 2013)

He uses these papers to explain that other researchers have proven the connection between the kinds of speeches and press releases a politician gives and how they represent their constituents, a key part of his argument.

**d) Was your assigned paper a descriptive study, an identification exercise, a numerical solution to system of equations study or some combination of the three?**

This paper was primarily an identification paper as the author is trying to find the relationship between two abstract entities: the type of representation politicians provides with the political climate of the time.

Secondarily, this paper has some aspects of a descriptive paper. It is not using a new data origin data source (Congressional press releases) but he is using a new method to analyze the topics and model them. The result of this method provides a new dataset which could be used to analyze other research questions. In the conclusion, the author mention a few research topics that this dataset could be useful for such as analyzing what politicians say about their work versus what they accomplish or how politicians work together on topics of common concern?

**e) What computational methods did this paper use to answer the research question? What was their result or answer to the question?**

The author used an unsupervised learning tool, topic modeling to estimate both coarse and granular topic in congressional press releases. He divided each press release into 44 granular topics and 8 coarse topics. From each topic, the author analyzes how Democrat vs Republican politicians take credit or attack the president in a press release.

He then validates the dataset he produces by comparing it to a human coded dataset of 800 of the press releases and finds a strong correlation between the two datasets. He also validates the dataset by showing that representatives that speak of farming the most represent districts with a higher proportion of farmers.

From here the author uses dataset to answer his research question. He states "In response to the changing electoral and institutional pressures, Republicans altered how they presented their work to constituents ... after Obama's election Republicans allocated a much smaller percentage of their press releases to credit claiming than in previous years ... while democratic credit claiming spiked in 2009." and "Instead of claiming credit, Republicans amplified criticism of the Obama administration and Democratic policies".

**f) Think of yourself as an academic referee. Give two suggestions to the author of your assigned paper of things the author might do to improve their results or strengthen their evidence for the answer to the question.**

The results of this paper seem kind of trivial: the democrats are taking credit when democratic policies are getting passed while the republicans are abandoning credit. It would be more interesting to expand this analysis to include data from before 2005. Is this credit claiming/vs abandoning a new phenomenon in the Obama administration or a common tactic used by opposition parties? Furthermore, this could be a tactic that is specific to the period studied and not have any more inference. For example, when President Obama got re-elected the Republican Minority Leader Mitch McConnell said it was the number one priority to make President Obama a one-term president. To improve his results, I would want to see if this theory is more generally applicable.

What a politician says is less interesting to me than how they vote. Is it possible to tie what the politician says are their priorities with how they vote on the issues. From this analysis is it possible to determine when a politician is being a hypocrite (ie. Saying one thing and doing another). What is the political climate that produces hypocrites. Another potentially more interesting study would be to match the priorities of the district (by employment industry data or survey methods) that the congressperson represents with the congressman's stated priorities. To improve his results, the author should look at voting records and representation in addition to press releases.