PLSC 504 - Topics in Political Methodology
Exercise Eight
December 12, 2024

Introduction

This is an exercise about topic models, and (a little bit of) sentiment analysis. Note that, as we discussed in class, this exercise is *optional*.

Exercise

We'll start with a paper – available in the course github repository – which you should read:

> Spirling, Arthur. 2012. "U.S. Treaty Making with American Indians: Institutional Change and Relative Power, 1784-1911." *American Journal of Political Science* 56:84-97.

Spirling's focus is on the "harshness" of treaties between Native American tribes and the U.S. government. Our main focus, in contrast, is on the topical content of those treaties.

The "Exercises" folder in the course github repository contains a subfolder with the full texts of the 365 "operable" treaties in Spirling's analysis, as well as a `.csv` file containing some metadata about those treaties. Your assignment is as follows:

1. Describe the process by which you preprocess your data, including justifications for the decisions you made.

2. Fit an LDA model to the texts, and describing the topics the model extracts, in whatever manner you see fit.

3. Take steps to determine / assess / justify the number of topics you believe to be appropriate for this corpus. In the end, choose and defend a number of topics for your analysis. (Hint: `stm` is your friend here.)

4. Fit a CTM to the text, using the same number of topics, and compare it to the LDA results. Discuss differences, if any.

5. Assess whether and how the prevalence of topics varies (perhaps smoothly) over time during the period in Spirling's data.

6. Assess whether and to what extent treaties created / ratified in Washington, D.C. are of systematically different topics than those from elsewhere in the country.

7. Finally: Once you've settled on a reasonable number of topics, assign each treaty to a topic. Then assess whether the sentiment (positivity/negativity) of each treaty varies across the topics assigned.

As noted above, this homework is *optional*. As such, it has no formal due date.