Relationships Across Political Texts: Text Reuse and Issue Attention in Federal Appropriations

Devin Judge-Lord

Department of Political Science, University of Wisconsin-Madison

Text Reuse and Topic Models

Observed text reuse is one way to bring semantic structure into bag-of-words models. Copied text fragments (1) arise from a distinct generative process and (2) contain token-level information about relationships among texts. Combining text reuse and topic modeling approaches may help to:

- Map coalitions or networks
- Identify who received their desired policy language
- Measure change across versions of policy texts
- Trace issue attention over time, despite similar word-use.

Text reuse and topic modeling methods are complementary approaches for tracing policy ideas. Whereas topics (or change in topic) may capture broad issue attention and framing, text reuse may capture specific alignment. For example, Brookhart and Tahk (2015) use dynamic LDA to identify the origins of policy issues and Wilkerson et al. (2015) use text reuse to trace specific bill provisions. Using both LDA and text reuse, I aim to identify the distribution of topics across copied and edited provisions of policy texts, simultaneously identifying document-level similarities and differences in issue framing or emphasis as well as the topics of specific alignments.

Examples

Edits may introduce new ideas in otherwise aligned text:

> FY2017: "When the Subcommittee heard from the FDA it focused on preventing burdensome regulations for producers and the American people in addition to ongoing discussions of how the FDA is implementing the Food Safety Modernization Act FSMA and the XXXXXXXXX motivation XXXXXX for XXXXX the generic drug labeling rule and regulation of tobacco products"

> FY2018: "When the Subcommittee heard from the FDA it focused on preventing burdensome regulations for producers and the American people in addition to ongoing discussions of how the FDA is implementing the Food Safety Modernization Act FSMA XXX XXX addressing XXXXXXXXX opioid XXX abuse XXX XXXXXXX XXXX XXXXXXXX XXXX and regulation of tobacco products"

When similar words mean different things, edits are informative:

> FY2010: "CLIMATE CHANGE The Committee continues to support the Administration's efforts to address climate change. Funding for its voluntary climate change programs are continued through this bill."

> FY2012: "CLIMATE CHANGE This Committee remains skeptical of the Administration's efforts to re-package existing programs and to fund new ones in the name of climate change."

Question

Who drives issue attention in appropriations?

- Party leadership
- Committee chairs
- White House
- Federal Agencies

Measuring Change Across Versions

1. Percent new words, ngrams, sentences
2. Percent unaligned (e.g. Smith-Waterman Local Alignment)
3. Global Alignment Scores
4. Change in Topic Proportions ($\sum(\theta_d - \theta_{d'})$)

Text Reuse Across Annual Appropriations Subcommittee Reports

Figure: 1: Agriculture, Rural Development, Food and Drug Administration, and Related Agencies Appropriations Subcommittees
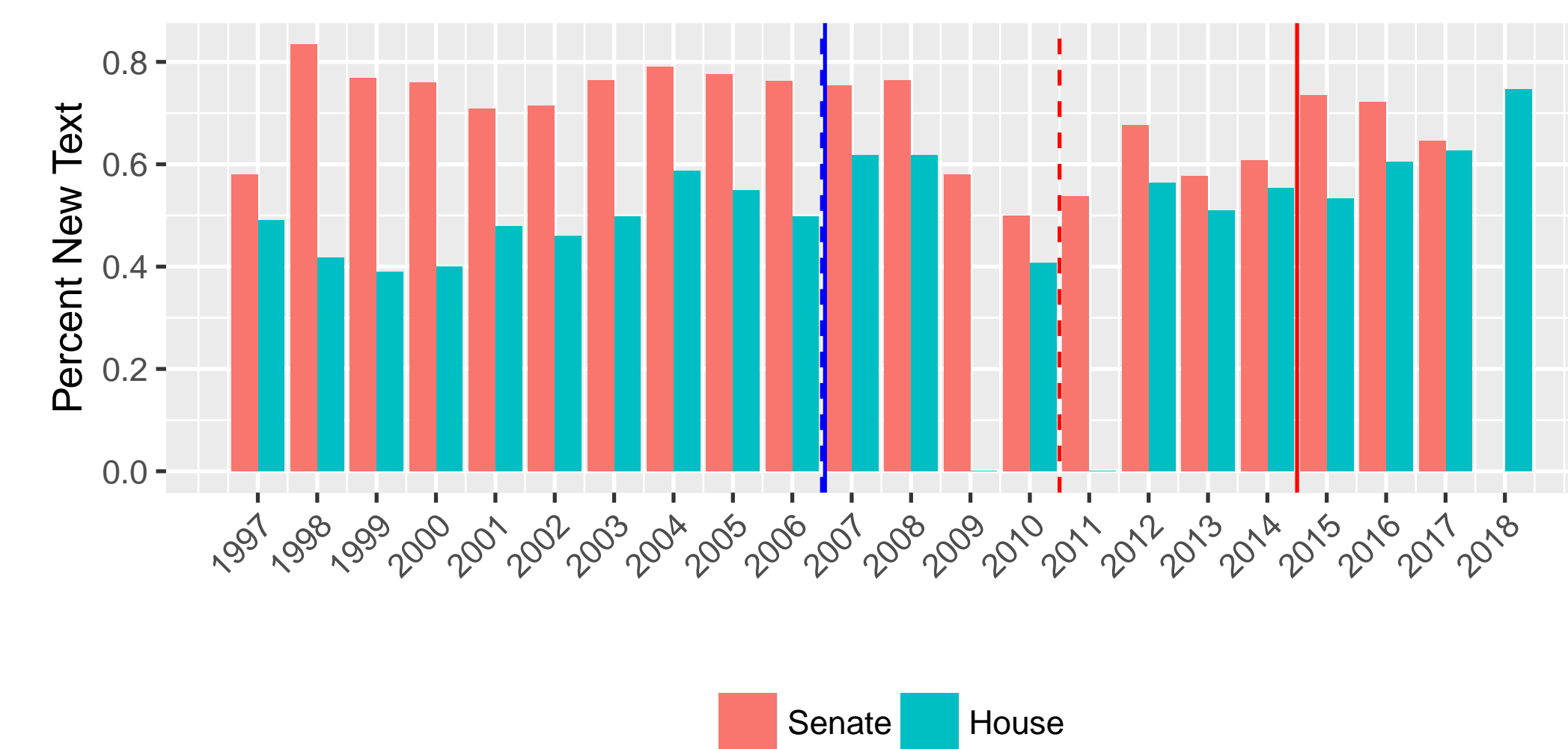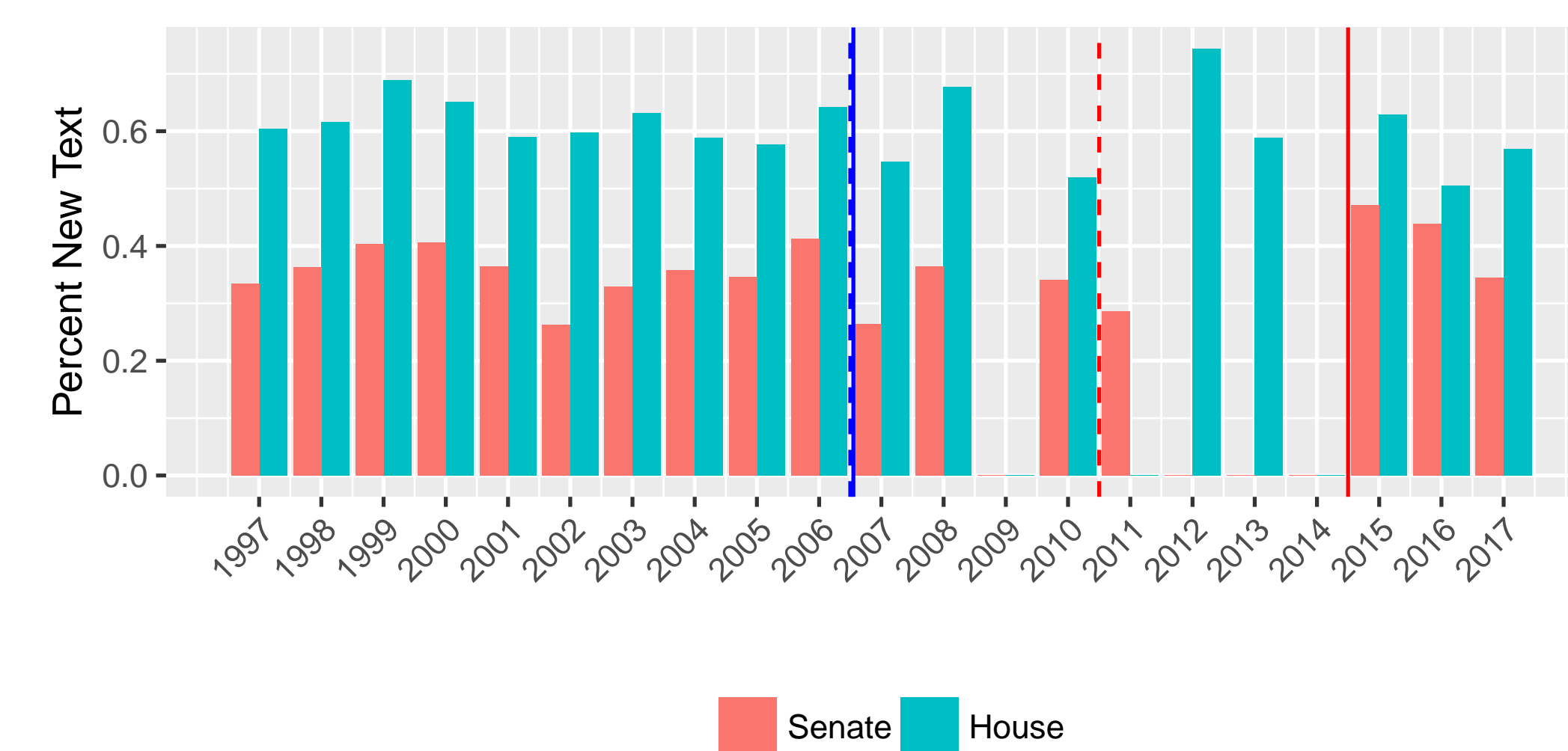

Figure: 2: Interior, Environment, and Related Agencies Appropriations Subcommittees

*Dashed vertical lines indicate party control change in the House, solid lines, in the Senate.
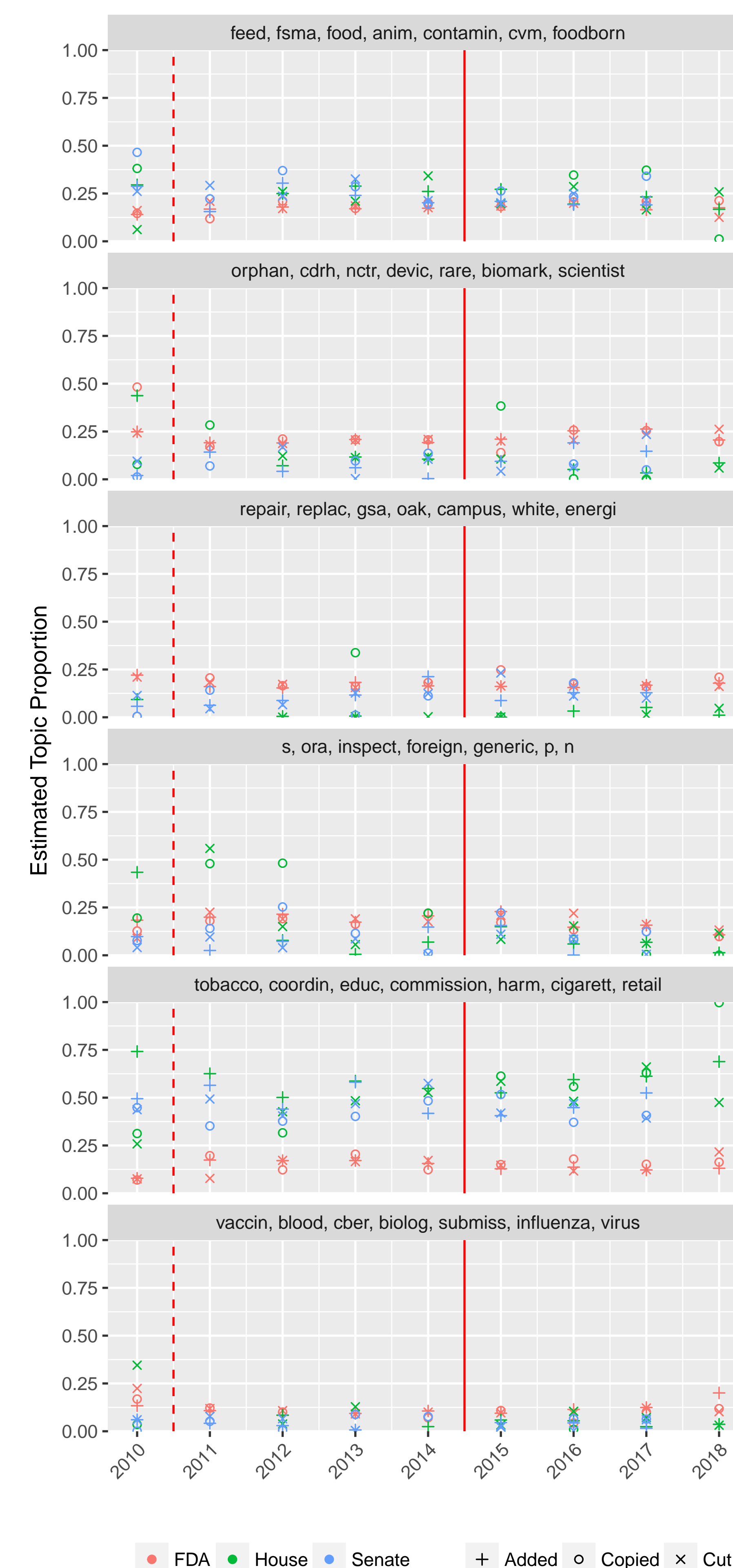
Data

Over 150,000 pages discussing policy priorities, goals, praise, and sanction

- Annual budget requests for 70 Federal Agencies
  - FY2010-2018, most earlier
  - 200-1200 pages describing how requested funds will be spent
- House and Senate Appropriations Subcommittee Reports
  - FY1996-2018
  - 100-200 pages describing how appropriated funds are to be spent
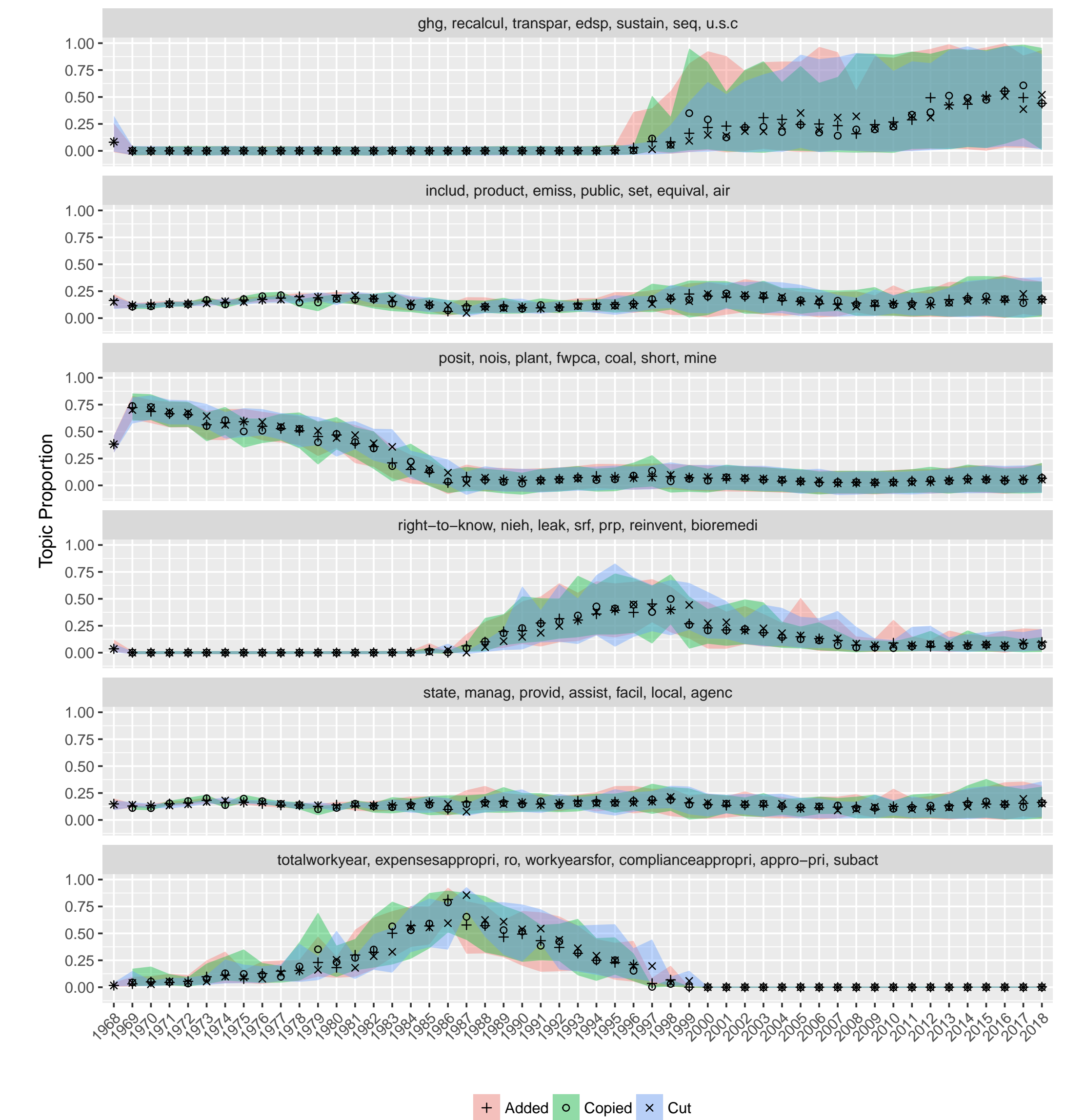 - Parsed by the agency addressed

Issue Attention Across Venues

Figure: 5. Issue Attention in Food and Drug Administration Appropriations

Text Added, Cut, and Copied Over Time

Figure: 6. Environmental Protection Agency Budget Justification Topics

Next Steps

1. Better LDA with text reuse:
   1. Drawing on Chang and Blei (2009), *document-level* relationships, i.e. network edges, can be inferred from patterns of text reuse.
   2. Text reuse identifies *token-level* relationships. Tokens belonging to copied phrases should be assigned to the same topic.

2. Model relationship between issue attention and appropriations:
Hypothesis: $\Delta Budget_{Agency} \approx \Delta\theta_{Congress} | \Delta\theta_{Agency}$

References

Brookhart, Jennifer L, and Alexander Tahk. 2015. "The Origin of Ideas."

Chang, Jonathan, and David M Blei. 2009. "Relational Topic Models for Document Networks." In *12th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 81–89. Clearwater Beach, FL.

Wilkerson, John, David Smith, and Nicholas Stramp. 2015. "Tracing the Flow of Policy Ideas in Legislatures: A Text Reuse Approach." *American Journal of Political Science* 59 (4): 943–956.