

Permutation-Based Testing of Topic Co-occurrence: A Network Analysis of Reddit Debates on DOGE, Tariffs, and the Big Beautiful Bill

Benjamin Forman-Barzilai¹, Ines Levin¹

¹University of California, Irvine
bformanb@uci.edu, inelevin@uci.edu

Abstract

We study how policy arguments are structured in Reddit discussions of three initiatives from the early period of Trump’s second presidency: the Department of Government Efficiency, the Liberation Day Tariffs, and the Big Beautiful Bill. We identify recurring policy considerations using a dictionary-based approach that is validated through manual coding and applied to full-text submissions. To distinguish meaningful associations from chance, we use a permutation-based test that compares observed co-occurrence counts for each pair of considerations to counts from randomly shuffled data with the same overall frequencies. From the pairs that pass this test, we build networks for each policy and describe their main features, including central topics, clusters of related considerations, and bridge topics that link clusters. The results show that the three debates have different argument structures and illustrate a simple way to add statistical testing to topic co-occurrence analysis.

Introduction

Online social media platforms have become major arenas for policy debates. Reddit, with its unique structure of subreddits dedicated to specific topics, has emerged as a popular destination for users seeking focused and in-depth discussions on public affairs. These conversations often span the ideological spectrum, taking place in subreddits open to diverse viewpoints or targeting users with specific ideological leanings. This makes Reddit a valuable resource for researchers, especially given its accessible data through an API and software libraries like the `RedditExtractor` R package, which we used in our analysis.¹

Previous research has shown that political discourse tends to become semantically and temporally compressed around major political events, meaning that people often discuss these events using similar language and within a narrow time frame (King, Pan, and Roberts 2017; Desiderio et al. 2025). In our analysis, we focus on three major events from the early part of Trump’s second presidency: Trump’s Liberation Day Tariffs (LDT), the Department of Government Efficiency (DOGE), and the Big Beautiful Bill (BBB). These

three policy initiatives served as primary narratives through which the media covered the early period of Trump’s second administration, becoming major factors shaping public perceptions of his return to office (Pew Research Center 2025):

- **Liberation Day Tariffs (LDT)**, implemented on April 2, 2025 (“Liberation Day”), established 10% baseline tariffs and country-specific duties as large as 50% across sectors including steel, automobiles, electronics, and consumer goods. Despite legal challenges and temporary exemptions for key trading partners, the administration reinstated the Liberation Day Tariffs in August 2025 with additional escalations in semiconductors and EV batteries. Discussion reflects concerns about consumer price increases alongside arguments about benefits for U.S. manufacturing and supply chain resilience (Pew Research Center 2025).
- **The Department of Government Efficiency (DOGE)**, established in January 2025 and initially led by Elon Musk, pursued aggressive federal restructuring including workforce reductions, weekly employee reporting requirements, and dismantling agencies like USAID (Megerian 2025). The initiative sparked multiple lawsuits alleging violations of federal advisory rules and constitutional concerns over spending terminations without congressional approval. While supporters praised modernization efforts and improved taxpayer returns, critics raised concerns about government capacity and legality. Musk stepped down in May 2025 amid cooling White House relations. By November 2025, DOGE had disbanded as a centralized entity with its functions absorbed by other agencies (Rozen 2025).
- **The Big Beautiful Bill (BBB)**, officially the One Big Beautiful Bill Act of 2025, was signed into law on July 4, 2025, combining tax cuts for individuals and small businesses with increased defense and border spending while reducing social programs including Medicaid and SNAP (Freking and Mascaro 2025). Supporters touted it as essential for economic growth and national security, while critics warned of steep deficit increases and reduced support for vulnerable populations, with debate centering on short-term stimulus benefits versus long-term fiscal risks.

By studying the discourse surrounding these policies, our research seeks to understand the structure of public dialogue

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹All replication materials are available at:

<https://doi.org/10.17605/OSF.IO/MHUCR>

and identify the types of considerations that may have bolstered or undermined public support.

Two common goals in text analysis applications within the social sciences are: (1) examining user interaction networks, where nodes represent users and links indicate interactions such as replies and following decisions (Barberá et al. 2015), and (2) analyzing topic prevalence, often through techniques like topic modeling (Roberts et al. 2014). While these approaches provide valuable insights, they do not reveal how political discourse is structured at the level of argumentation. The central question we address is not who participates in discussions or what topics are discussed, but rather how people assemble considerations—the specific concerns, values, and arguments they raise—to build their narratives around the three policy themes. When discussing DOGE, for example, do users who mention “government efficiency” also tend to invoke “job losses”? Do those discussing BBB’s tax cuts consistently pair them with concerns about the “national deficit”? This focus on how considerations co-occur aligns with research that treats ideas and concepts as the foundational elements of political argumentation (Sagarzazu and Klüver 2017; Spaier et al. 2017).

Our primary contribution is an empirical comparison of how policy considerations are combined in Reddit debates about DOGE, the Liberation Day Tariffs (LDT), and the Big Beautiful Bill (BBB). Unlike common hashtag network studies that rely on limited metadata (Garimella et al. 2018; Chen, Deb, and Ferrara 2022), we analyze the full text of Reddit submissions and construct co-occurrence networks that reveal which considerations tend to be central, which ones cluster together, and which serve as bridges between clusters. To ensure that these network patterns are not artifacts of highly frequent considerations or chance co-mentions, we apply a simple permutation-based procedure that compares the observed co-occurrence count for each pair to a null distribution generated by randomly shuffling considerations across submissions while holding their marginal frequencies fixed. This procedure offers a straightforward way to build co-occurrence networks that highlight substantive links between considerations while filtering out patterns that can be explained by chance or overall frequency.

Having established our research goals, we next review related approaches to semantic network analysis and situate our empirical approach and permutation-based test within this literature.

Background and Related Work

Topic co-occurrence networks are a type of network in which connections between topics, represented as nodes, are established when those topics systematically appear together across a collection of texts. A foundational approach developed by computer scientists is the semantic network, where nodes represent words or concepts and links denote semantic or substantive relationships (Bollacker et al. 2008). The resulting networks often exhibit community structure, with densely connected clusters of related concepts that can be identified through algorithms developed in network science (Fortunato 2010; Newman 2006). Community detection in

text-derived networks has become central to computational linguistics, from identifying discourse communities in social media (Zhang et al. 2017) to discovering semantic clusters in word co-occurrence networks (Ferrer i Cancho and Solé 2001). Recent NLP work has shown that community structure in text networks often reflects underlying linguistic and social phenomena (Bryden, Funk, and Jansen 2013).

A significant challenge, both for semantic networks and network analysis in general, is that links often capture multiple, overlapping types of relationships. For example, in an analysis of European Parliament members, connections could represent actual political alliances, but also the temporal coincidence of tweets or trending topics (Van Vliet, Törnberg, and Uitermark 2020). In our analysis, we focus not on the co-occurrence of individual words but of “considerations,” topics in people’s minds when expressing opinions about broader political themes. By focusing on these higher-level concepts, we aim to ensure that the identified connections reflect reasoning patterns rather than simple semantic similarity or stylistic conventions.

Political scientists have been at the forefront of developing and applying methods for treating text as data (Grimmer and Stewart 2013). In the analysis of topic networks, applications within the social sciences predominantly follow one of two primary methodologies: model-based approaches and dictionary-based approaches.

Model-based approaches, such as Structural Topic Modeling (STM), allow researchers to identify topics in a text corpus while incorporating document-level covariates to understand how topic prevalence and content vary across different contexts (Roberts et al. 2013). The method estimates correlations between topics conditional on these covariates, and readily available software enables researchers to create network representations based on these estimates (Roberts, Stewart, and Tingley 2019). A key difference between STM and simple co-occurrence analysis is that STM establishes relationships using a model that learns the probability of joint occurrence conditional on covariates, while simple co-occurrence analysis calculates unconditional frequencies of topic pairs without accounting for document-level characteristics. Our analysis follows the unconditional frequency approach and does not incorporate document-level covariates. Consequently, observed co-occurrence patterns may partly reflect the composition of Reddit users discussing each policy rather than inherent conceptual connections. However, this is not a major limitation given our goal, which is simply to characterize the overall incidence of topic co-occurrence in Reddit discussions of the three policies.

In contrast to STM, which involves topic discovery, dictionary-based approaches rely on pre-defined lexicons to identify considerations. This creates another important difference: while model-based approaches like STM simultaneously learn about the topics in the texts and model their relationships, dictionary methods depend on external definitions. Traditional dictionary-based procedures have several noted disadvantages: they can rely on potentially arbitrary and subjective topic definitions, classification is typically done deterministically (based on whether words associated with a topic are present or not) rather than probabilistically,

and constructing the dictionary can be labor-intensive.

For our research goals, the dictionary approach has several desirable properties. Because considerations are defined externally, they can be applied consistently across the three policies, enabling straightforward comparability. For example, Jones et al. (2018) employ a dictionary-based method to identify the political values expressed by members of Congress in their Twitter posts by counting words associated with each value. While their procedure does not investigate the co-occurrence of values, such an analysis would be a plausible extension. Furthermore, dictionaries can be carefully developed to ensure that all considerations are unambiguous and have clear interpretations. Such a level of clarity is not always achievable with topics discovered by unsupervised models.

Our methodology addresses dictionary-based limitations through LLM-assisted dictionary creation and, more importantly, a permutation-based procedure to establish statistical significance for co-occurrences rather than using arbitrary frequency thresholds. The permutation framework itself follows standard randomization inference, similar to approaches used in network analysis for validating community structure (Lancichinetti et al. 2011). Here we show how such a test can be integrated with community detection methods from network science (Blondel et al. 2008; Fortunato 2010) for dictionary-based co-occurrence analysis of specific policy debates.

This approach complements alternative topic modeling methods like BERTopic (Grootendorst 2022), which have seen recent social science applications (e.g., (Kloo, Cruickshank, and Carley 2024), who analyze social media discourse on the 2022 Russian invasion of Ukraine). While BERTopic can be used to generate network-style visualizations based on how closely topics cluster, our approach focuses on a pre-specified set of considerations and provides statistical validation through explicit significance testing against a null expectation of random co-occurrence. This helps distinguish thematic associations that are unlikely to arise by chance from patterns that can be explained by overall frequency alone.

Our work also advances upon frequency-based co-occurrence studies of social media hashtags. For example, Garimella et al. (2018) and Chen, Deb, and Ferrara (2022) both analyze hashtag co-occurrence to explore patterns in political discourse, with Garimella et al. (2018) using count thresholds to identify clusters that signal political coalitions. Building on this, other researchers have explored ideological framing over time. Tuters, Willaert, and Meyer (2023), for instance, created a hashtag map showing how the use of #mRNA in discussions about the COVID-19 vaccine became associated with right-wing conspiracy theories. Other studies have explored the diffusion of social movements; Xiong, Cho, and Boatwright (2019), for example, examines the #MeToo movement by identifying frequently used terms like “women” and “survivor,” but treats individual hashtags as isolated units rather than examining how they co-occur within posts. Our approach differs from these hashtag-based methods by using the full text of submissions and applying a statistical test to identify thematic groupings, rather than

relying on pre-defined tags and count thresholds.

Methodology

Data Collection and Processing

We focus on the `r/PoliticalDiscussion` subreddit because, with more than 100 thousand weekly visitors and more than 4 thousand weekly contributors, it serves as a central venue for political and policy debates among users of diverse ideological orientations. Discussions in this subreddit offer a useful benchmark for understanding how discourse unfolds in a relatively ideologically neutral environment. While there are multiple political subreddits to choose from, such as `r/Politics` and `r/News`, `r/PoliticalDiscussion` encourages longer, opinionated submissions, unconstrained by any specific topics. This allows co-occurrences to appear spontaneously, instead of a specific co-occurrence showing up frequently in the responses to a viral news article, for example. `r/PoliticalDiscussion` moderators also remove “memes” and “links substituting for explanation,” ensuring that discussion occurs through textual submissions on the subreddit rather than images or other websites.²

The timing of our data collection varied depending on policy. For the Department of Government Efficiency (DOGE), we collected data from January 20 to June 4, 2025, covering the period from the announcement of the new entity to a week after Musk departed from Trump’s administration. For Trump’s LDT, we gathered posts from April 2 (“Liberation Day”) to August 20, 2025, spanning from the initial announcement to the reinstatement of the new tariffs in August. We started data collection on Liberation Day to ensure that this specific event would be the primary subject of tariff discussion, not tariffs in other contexts or as a concept. For BBB, we collected data from January 20 to July 11, 2025, a period that includes its passage in the House and extends to a week after it passed the Senate in early July.

We collected posts using the R package `RedditExtractoR`, which connects to the Reddit API and enables the retrieval of posts from specific subreddits that mention designated keywords. For each of the three policies, we used the sets of keywords {“DOGE,” “Department of Government Efficiency”}, {“tariffs”}, and {“BBB,” “OBBBA,” and “Big Beautiful Bill”}, respectively.³ For each policy, we first identified threads that mentioned at least one of the relevant keywords, in either the original post or the comments below, and then downloaded the entire thread, including the original post and all its comments. This

²The `r/PoliticalDiscussion` subreddit maintains strict content moderation policies (reproduced in Appendix B). The rules reference a wiki guide titled “Tips on Writing a Successful Political Discussion Post” available at <https://www.reddit.com/r/PoliticalDiscussion/wiki/posts> (last accessed: January 09, 2026).

³For LDT, we used “tariffs” as our search keyword because it was the predominant term in Reddit discussions during our collection period and captured variations like “Trump’s tariffs” and “Liberation Day Tariffs.” We excluded the acronym “LDT” as it was rarely used conversationally and could refer to unrelated concepts. (We use “LDT” throughout this paper for brevity.)

process resulted in datasets containing 2,181, 3,981, and 977 submissions for DOGE, LDT, and BBB, respectively. For clarity, we use ‘submission’ to refer to an individual unit of text (either an original post or a comment), and ‘thread’ to refer to an original post plus all its associated comments. Our unit of analysis is the individual submission, not the thread.⁴

We matched submissions to dictionary entries using a two-stage approach: first identifying multi-word phrases through exact string matching, then applying Porter stemming to match morphological variants of single-word terms. This sequential approach preserves multi-word phrase integrity while capturing lexical variations (e.g., ‘consider,’ ‘considers,’ ‘considering’).

Dictionary Creation and Validation

Our analysis uses a unified dictionary created via a two-stage hybrid approach combining a large language model with human-in-the-loop curation. In the generation stage, we supplied a long, structured prompt to the Gemini 2.5 Pro model (accessed on August 8, 2025). We prompted the LLM with detailed policy descriptions, seed concepts, and instructions. For each policy, the model was tasked with generating thematically distinct considerations and associated search terms designed to capture informal, everyday language (including synonyms and slang), with a balance of positive, negative, and neutral terms. This process resulted in three initial policy-specific dictionaries.⁵

We selected Gemini 2.5 Pro for generating the initial dictionaries primarily for its large context window, which allowed us to supply long prompts including the full policy narrative together with detailed, standardized instructions in a single policy-specific prompt. The prompts themselves were drafted by the researchers before examining any Reddit submissions, based on our substantive knowledge of the issues and using limited LLM assistance to brainstorm example considerations and terms. A detailed description of the prompt structures is available in Appendix A.

In the refinement stage, we pooled the three policy-specific dictionaries (totaling 79 initial considerations) and systematically consolidated them into a unified dictionary of 61 considerations, with an average of 24.5 search terms each (see Table 1 for an example). The consolidation process involved three steps: (1) Merging semantically identical considerations across

⁴To protect user privacy while maintaining the analytical integrity of the dataset, we anonymized the author field in all Reddit data files included in the replication package using a hashing algorithm augmented with a private cryptographic salt.

⁵Appendix A provides a detailed description of the prompts. The replication package includes: (1) the full text of all prompts used; (2) the three initial policy-specific dictionaries generated by the LLM; (3) the final unified dictionary used in our analysis, allowing readers to trace the complete dictionary development process; (4) the complete R code; and (5) de-identified Reddit data sufficient to reproduce all tables and figures in the main text and Appendix. All analyses were conducted in R on a MacBook Air with an Apple M3 chip (8-core CPU, 10-core GPU, 16-core Neural Engine) and 24 GB of unified memory.

Sample Considerations	Sample Terms
Household Costs	cost of living, inflation, groceries, paycheck to paycheck
Labor Market	brain drain, jobs, layoff, hiring freeze, american workers
Fraud/Corruption	mismanagement, fraud, boondoggle, pork barrel, kickback
Trade War	trade war, retaliatory tariffs, tit-for-tat, trade conflict
National Security	national security, strategic sectors, defense industry, self-sufficiency
Executive Authority	executive order, presidential power, checks and balances, executive overreach
Social Safety Net	medicaid, food stamps, social security, welfare, safety net
Environment/Energy	climate change, emissions, clean energy, energy independence
Partisan Conflict	partisan, bipartisan, political theater, gridlock, deep state

Note: This table shows a sample of 9 considerations from the full unified dictionary of 61 considerations and an average of 24.5 search terms per consideration. The complete dictionary is available in the replication materials.

Table 1: Structure of the unified conceptual dictionary used for the analysis.

dictionaries. For example, `workforce_jobs` (DOGE), `labor_market_outcomes` (tariffs), and elements of `EconomicGrowthJobCreation` (BBB) were merged into a single `labor_market` consideration as they contained overlapping terms like ‘jobs,’ ‘lay-offs,’ ‘unemployment,’ and ‘american workers.’ (2) Consolidating near-duplicate considerations that addressed the same concept from different angles. For instance, `household_cost_of_living` (tariffs) and `HouseholdEconomicsCostofLiving` (BBB) became `household_costs`, while maintaining distinct categories for conceptually different aspects like `fiscal_savings` versus `tax_revenue`. (3) Splitting overly specific considerations into broader categories while creating new cross-cutting considerations. For example, Musk-related terms were separated into `musk_companies` (specific entities) and `external_influence` (broader conflict-of-interest concerns). This process, guided by our domain expertise with limited additional LLM assistance for identifying synonyms, reduced redundancy while ensuring comprehensive coverage across all three policy contexts.

To validate the assignment of terms into considerations in our dictionary, two coders (the two authors of this paper) independently completed three complementary validation tasks:

- **Task 1 (Term Review).** The coders see 100 randomly sampled terms, each paired with its actual assigned consideration. For each term, the coders judge whether the

term fits its assigned consideration (Yes/No/Uncertain). This task directly tests the face validity of the existing assignments.

- **Task 2 (Blind Classification).** The coders see 100 randomly sampled terms without their assigned considerations. For each term, the coders select the consideration that best fits. This task tests how intuitive and “recoverable” the categories are when coders do not know the original labels and must choose from the full set of considerations.
- **Task 3 (Signal Detection).** The coders see 200 term-consideration pairings, some correct and some intentionally incorrect. For each pairing, the coders decide whether the term belongs in that consideration (Yes/No). This task evaluates how easily coders can discriminate between correct and incorrect pairings. That is, whether specific term-consideration combinations “feel right” or “wrong” in practice.

The results of these validation exercises are summarized in Table 2.⁶ In Task 1, both coders simultaneously accepted the existing classification for 84% of terms (with individual acceptance rates of 95% for Coder 1 and 87% for Coder 2). Overall agreement on accept vs. non-accept decisions was 86%, and Gwet’s AC1 indicated high reliability (0.85).⁷ These results suggest that the vast majority of term-consideration assignments look appropriate to both coders.

Task 2 is intentionally more demanding, as coders must choose a label from the full set of considerations without seeing the original assignment. In this blind classification setting, both coders selected the *correct* consideration for the same term 48% of the time, although their individual accuracies were higher (55% for Coder 1 and 75% for Coder 2). The two coders agreed on their chosen consideration 55% of the time, with Cohen’s κ indicating moderate agreement (0.54). This pattern is consistent with a moderately difficult labeling task: the categories are recoverable above chance, but not trivial to infer from terms alone.

In Task 3, where coders simply judge whether a given term-consideration pairing is correct, performance is again very strong. Both coders were simultaneously accurate 90% of the time, with individual accuracies of 94% (Coder 1) and 95% (Coder 2). Agreement on Yes/No decisions was 90.5%, and Cohen’s κ indicated high reliability (0.81). Signal detection analyses further show excellent discriminability between correct and incorrect pairings (average $d' \approx 3.3$ across coders), indicating that coders can reliably distinguish valid from invalid assignments when evaluating specific pairings.

While these validation tasks provide reassurance about the quality of term-consideration pairings, it remains possible that the dictionary does not fully capture all considerations that arise in the context of the three policies. To

⁶More detailed validation statistics are available in Appendix Tables C.1–C.3.

⁷For Task 1 we report Gwet’s AC1 instead of Cohen’s κ to address the κ paradox, which arises here because both coders have very high acceptance rates, leading to highly unbalanced marginals.

Metric	Task 1 <i>Term Review</i>	Task 2 <i>Blind Classification</i>	Task 3 <i>Signal Detection</i>
Sample size	100 terms	100 terms	200 pairings
Performance ^b	84% accept	48% correct	90% accurate
Agreement	86%	55%	91%
Reliability ^a	0.85	0.54	0.81
Sensitivity ^c	–	–	$d' = 3.3$

^a Task 1: Gwet’s AC1; Tasks 2–3: Cohen’s κ .

^b Task 1: Both coders accept; Task 2: Both classify correctly; Task 3: Both accurate.

^c Signal detection sensitivity (d'), averaged across coders.

Table 2: Dictionary Validation Summary

address this and other dictionary-related concerns, we conducted extensive robustness testing after our main analysis. Specifically, we repeatedly re-estimated the three topic co-occurrence networks under a large set of random dictionary perturbations. This procedure, and its results, are described in Section *Robustness to Dictionary Perturbations*.

Estimation Method

If two considerations appear in the same submission, it does not necessarily mean that they tend to co-occur in a meaningful way. Even if considerations are raised independently, some co-occurrence will happen simply by chance. The more frequently a consideration is mentioned overall, the more likely it is to appear alongside other considerations, regardless of any substantive connection. For instance, if nearly every user refers to Trump’s role, it becomes impossible to determine whether this consideration is genuinely associated with others or just commonly mentioned. Meaningful associations can only be identified for considerations that are mentioned often enough to allow for analysis, but not so frequently that their presence becomes trivial or automatic.

To determine if the association between any pair of considerations was statistically significant, we conducted a permutation-based procedure to assess whether their co-occurrences happen more frequently than expected by chance alone. While permutation-based tests are commonly used in machine learning to evaluate variable importance in supervised learning algorithms (Levin 2022; Strobl et al. 2008) and in network science to validate community structure (Lancichinetti et al. 2011), to the best of our knowledge, they have not been applied to test for significant topic co-occurrence. The procedure involves the following steps:

1. **Observed Data Calculation:** We first constructed a binary submission-consideration matrix (SCM), with submissions as rows and considerations as columns. From this, we calculated:

- The frequency vector \mathbf{f} , where each element f_j is the

total number of submissions mentioning consideration j (the column sums of \mathbf{SCM}).

- The observed co-occurrence matrix (\mathbf{C}) by multiplying the transpose of the \mathbf{SCM} by itself ($\mathbf{C} = \mathbf{SCM}^T \cdot \mathbf{SCM}$). An element C_{jk} in this matrix provides the actual count of times considerations j and k appeared together.

2. **Null Expectation Simulation:** We then ran a loop for 1,000 iterations to create a null distribution. In each iteration:

- A simulated matrix (\mathbf{SCM}_s) was generated. This matrix preserved the exact frequency of each consideration (f_j) from the real data but randomly distributed their occurrences across all submissions (permutation step).
- The permutation breaks any genuine thematic links, creating a scenario where co-occurrence is driven only by chance.
- A synthetic co-occurrence matrix was calculated for this permuted data ($\mathbf{C}_s = \mathbf{SCM}_s^T \cdot \mathbf{SCM}_s$).

3. **Significance Testing:** After 1,000 simulations, we established a significance threshold for each pair of considerations by calculating the 97.5th percentile of their simulated co-occurrence counts.

- An observed co-occurrence C_{jk} from the real data was deemed statistically significant if it exceeded this 97.5th percentile threshold. This indicates that the two considerations are linked together in the discourse more frequently than can be explained by random chance alone.

After testing every pair of considerations for meaningful co-occurrence, we created network visualizations for each one of the three themes, where nodes represent considerations, the size of each node is proportional to the number of submissions mentioning the consideration (f_j), and a link is drawn between nodes if the observed co-occurrence exceeds the upper bound according to our simulation procedure (97.5th percentile of \mathbf{C}_s over all simulations), suggesting a statistically significant association.

We carried out the network analysis for each theme using the complete significant network, with one node per consideration in the dictionary. For the purpose of creating network visualizations, we filtered out the nodes (considerations) with relative frequency of less than 2.5% across submissions, to simplify the interpretation.

Our analysis reveals that a network's shape is highly informative, revealing several key features of the discourse for each major theme:

- **Central Hubs:** The existence of highly connected considerations that are frequently brought up in conjunction with many other considerations.
- **Thematic Clusters:** The formation of distinct communities of considerations that tend to co-occur, representing specific sub-arguments or narratives.

- **Conceptual Bridges:** The presence of “bridge” considerations that link otherwise disparate thematic clusters, tying different parts of the conversation together.
- **Core-Periphery Structure:** A clear division between a core of central, heavily debated considerations and a periphery of more marginal considerations.

We quantify these features of the network numerically by calculating standard measures from social network analysis. Each metric provides a different lens through which to understand the structure of the discourse:

- **Degree Centrality:** Identifies “thematic hubs,” which are high-frequency considerations discussed in conjunction with a wide variety of other issues. We report normalized values ranging from 0 to 1, where 0 indicates no connections and 1 indicates maximum possible connections. Values above 0.3 typically indicate highly connected nodes.
- **Community Detection (Louvain Method):** Detects “thematic clusters” which are groups of considerations more densely connected to each other than to the rest of the network, representing specific sub-arguments. We identify clusters using the `cluster_louvain()` function from the `igraph` package.
- **Betweenness Centrality:** Pinpoints “conceptual bridges,” which are considerations that link otherwise disparate thematic clusters. We report normalized values from 0 to 1, measuring the proportion of shortest paths passing through each node. Values above 0.1 suggest important bridging roles between different parts of the network.
- **Eigenvector Centrality:** Distinguishes the “core vs. periphery,” identifying influential considerations at the center of the discourse versus more marginal or idiosyncratic ones. We calculate eigenvector centrality using `eigen_centrality()`\$vector. These scores are scaled from 0 to 1, where higher values indicate connection to other well-connected nodes. The highest-scoring node in each network receives a value of 1.0, with others scaled proportionally.

We also quantify modularity, degree centralization, and clustering coefficient (transitivity) for the overall network structure. Modularity usually ranges from -0.5 to 1, with values above 0.3 indicating strong community structure and values above 0.5 suggesting distinct thematic clusters. Degree centralization ranges from 0 to 1, where values closer to 1 indicate a star-like structure with a few highly central nodes, while values closer to 0 suggest more distributed connectivity. The clustering coefficient ranges from 0 to 1, measuring the tendency for nodes to form triangles. Higher values indicate more densely interconnected local neighborhoods of considerations.

Simulation Study

Before analyzing the Reddit data we collected, we conducted a simulation study to evaluate whether our technique for identifying significant co-occurrence can recover true relationships between considerations and network structures

Metric	Int	Pol	Frag	C-P
<i>Basic Properties</i>				
Total Submissions	1,500	1,500	1,500	1,500
Nodes in Network	57	24	50	43
Edges in Network	177	141	86	151
<i>Structural Characteristics</i>				
Number of Communities	6	2	7	8
Modularity	0.463	0.422	0.666	0.192
Degree Centralization	0.157	0.359	0.113	0.261
Clustering Coefficient	0.422	0.898	0.674	0.757

Note: Int = Integrated, Pol = Polarized, Frag = Fragmented, C-P = Center-Periphery

Table 3: Network Structure Across Simulated Discourse Patterns

in simulated data with known properties. We generated synthetic discourse data using the same 61 policy considerations from our unified dictionary that we applied to the Reddit data. Rather than generating actual text, we directly simulated submission-consideration matrices where each row represents a hypothetical discussion submission and each column represents one of the 61 considerations.

We simulated four archetypal discourse structures that we might expect to find in real policy discussions:

- **Integrated Discourse:** Multiple overlapping mini-clusters with substantial background noise, representing broad cross-cutting discussion. Created through 25+ small overlapping clusters (2-3 considerations each) with higher baseline co-occurrence probability.
- **Polarized Discourse:** Two opposing ideological camps with minimal cross-communication. One cluster included market-oriented considerations, while the other included social equity considerations.
- **Fragmented Discourse:** Multiple isolated topical silos with minimal bridging. Separate clusters for technology, economics, and governance.
- **Center-Periphery Discourse:** A central cluster of interconnected core topics surrounded by peripheral topics that primarily connect through the core rather than to each other.

We then applied our complete analytical pipeline, including permutation-based significance testing (1,000 iterations at the 97.5th percentile threshold), to each synthetic matrix. Table 3 presents the resulting network metrics.

The validation reveals distinct structural signatures for each discourse pattern. Fragmented discourse has the highest modularity (0.666) with seven communities, indicating siloed conversations. Polarized discourse shows moderate modularity (0.422) with exactly two communities and very high clustering (0.898), reflecting tight within-group connections. Integrated discourse displays intermediate modularity (0.463) with broad participation (57 nodes) and substantial cross-cutting ties (177 edges), while the center-periphery structure has the lowest modularity (0.192), con-

Metric	LDT	DOGE	BBB
<i>Basic Properties</i>			
Total Submissions	3,981	2,181	977
Unique Users	1,910	868	529
Considerations Mentioned (> 2.5%)	13	15	10
<i>Structural Characteristics</i>			
Nodes in Significant Network	57	55	43
Edges in Significant Network	294	198	98
Number of Communities	6	5	5
Modularity	0.219	0.295	0.411
Degree Centralization	0.352	0.311	0.225
Clustering Coefficient	0.386	0.279	0.313

Table 4: Overall Network Structure Across Themes

firming overall integration despite eight communities. In particular, the high clustering in the Center-Periphery structure (0.757) reflects a tight central cluster through which peripheral topics connect, while the high degree centralization in the Polarized network (0.359) indicates a small set of bridge considerations spanning the two camps.

Appendix Figures D.1–D.4 visually recover these known network structures—overlapping clusters for the Integrated network, two blocs for the Polarized network, multiple loosely connected components for the Fragmented network, and a dense core with a sparse periphery for the Center-Periphery network—illustrating that the permutation-based edge selection preserves the global shape of the underlying discourse and distinguishes between archetypal network structures.

Results

We first examine the overall structures of the three topic co-occurrence networks (see Table 4). Of the 61 considerations in the unified dictionary, only 10–15 appeared in more than 2.5% of submissions across the three themes, though most (43–57) appeared at least once, yielding 98, 198, and 294 statistically significant co-occurrence pairs for BBB, DOGE, and LDT, respectively. These structural differences should be interpreted in light of the varying data volumes, with LDT generating over four times as many submissions as BBB (3,981 vs 977), potentially contributing to denser network connectivity for LDT.

The networks reveal distinct discourse structures across policies. While our Simulation Study demonstrated that the method can successfully recover known network patterns, the empirical networks likely represent more complex, hybrid forms not fully captured by our stylized simulations. BBB shows the most compartmentalized discussion, with high modularity (0.411) and 5 distinct thematic communities, suggesting conversations organized around separate, specialized considerations with limited cross-cutting discussion. LDT exhibits the opposite pattern: lower modularity (0.219) but higher clustering (0.386) and 6 communities, indicating a more interconnected discourse where considerations frequently appear together across different contexts. DOGE falls between these patterns with moderate modular-

ity (0.295) and 5 communities. DOGE’s clustering coefficient (0.279) is notably low, suggesting a sparse connectivity pattern distinct from the denser structures we explored in our simulations. The variation in community numbers (5–6) reflects different organizational logics: BBB’s five communities suggest more fragmented, specialized discussions around distinct policy provisions, while LDT’s six communities with lower modularity indicate broader thematic groupings with more cross-cutting discourse. None of the policy debates are characterized by a single dominant consideration, though LDT discourse is more organized around a few central themes (centralization 0.352) compared to the more evenly distributed discussions in BBB (0.225).

One notable finding is the absence of the high modularity combined with high clustering seen in our simulated polarized discourse (0.422 modularity, 0.898 clustering). While BBB shows moderate modularity (0.411) suggesting some community structure, it has low clustering (0.313), indicating these communities are not densely interconnected echo chambers. The other networks show either low modularity with moderate clustering (LDT: 0.219, 0.386) or moderate modularity with low clustering (DOGE: 0.295, 0.279). This pattern (never observing high values for both metrics simultaneously) suggests that $r/\text{PoliticalDiscussion}$ avoids the dense, isolated camps characteristic of polarized discourse.

Below, Tables 5–7 show the most frequently mentioned and most connected considerations based on our full network analysis. The accompanying network visualizations (Figures 1–3), however, are filtered to display only those considerations with an incidence above 2.5%. This threshold is applied to simplify the visualization and make the core argumentative structure easier to interpret, though it comes at the cost of omitting the network’s less frequent, peripheral nodes.

Liberation Day Tariffs

The discourse on Liberation Day Tariffs is organized around a clear core-periphery structure. The combination of low modularity (0.219) with moderate clustering and high centralization suggests a discourse organized around central themes, though the specific configuration differs from our stylized simulations. At the center are Trade Ideology, Labor Market, and Federal Agency Impact, which are both frequently mentioned and have high degree centrality (Table 5). Nearly every other consideration in the network connects to at least one of these nodes, establishing them as the foundation of the debate. In contrast, considerations like Partisan Conflict and Big Tech were peripheral.

Beyond this general structure, a closer look reveals the nuanced roles different considerations play. For instance, Federal Agency Impact has a high betweenness centrality (0.127), showing it acts as a conceptual bridge that links otherwise separate conversational clusters. The analysis also highlights the difference between a topic’s frequency and its structural influence. While Trade Ideology (7.16% incidence) and Partisan Conflict (7.99% incidence) were mentioned with similar frequency, their eigen-

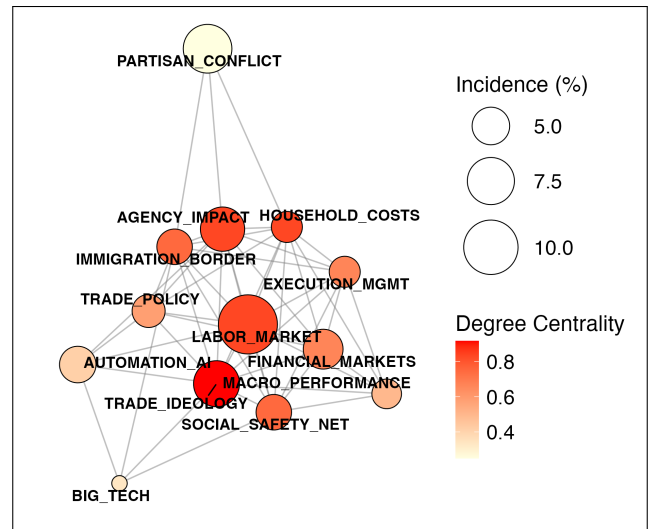


Figure 1: Co-occurrence network for LDT discourse. Nodes represent policy considerations, with size proportional to mention frequency and color indicating degree centrality. Edges represent statistically significant co-occurrences ($p < 0.025$).

Consideration	Inc. (%)	Deg.	Bet.	Eig.
<i>Panel A: Most Frequently Mentioned</i>				
Labor Market	11.88	0.482	0.106	0.875
Partisan Conflict	7.99	0.196	0.015	0.295
Trade Ideology	7.16	0.536	0.100	1.000
Federal Agencies	6.66	0.482	0.127	0.835
Financial Markets	5.50	0.304	0.024	0.630
<i>Panel B: Most Connected Hubs</i>				
Trade Ideology	7.16	0.536	0.100	1.000
Labor Market	11.88	0.482	0.106	0.875
Social Safety Net	4.57	0.482	0.096	0.867
Federal Agencies	6.66	0.482	0.127	0.835
Immigration	4.57	0.482	0.089	0.856

Table 5: Top Considerations in LDT Discourse

vector centralities are notably different (1.000 vs. 0.295). This demonstrates that while partisan talking points were common, they remained isolated from other key themes. Discussions about trade ideology, however, were deeply integrated into the broader debate, making that concept structurally essential to the arguments being made.

Department of Government Efficiency

Discourse on the Department of Government Efficiency (DOGE) is anchored by Federal Agency Impact and the Labor Market, which are the network’s central nodes with degree centralities of 0.444 and 0.370, respectively (Table 6). A key feature of this network is its low cluster-

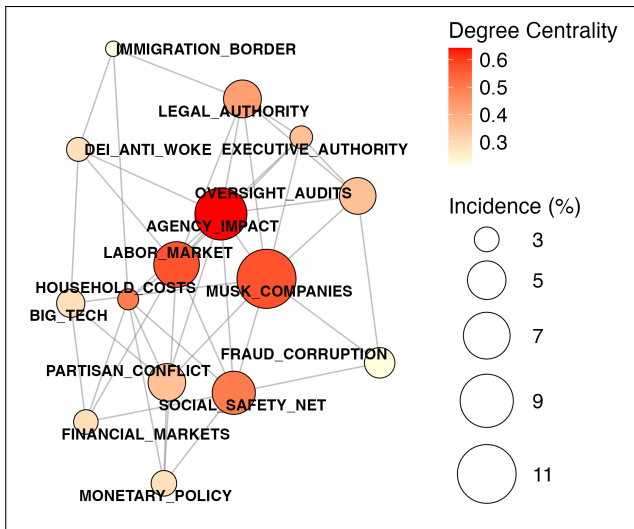


Figure 2: Co-occurrence network for DOGE discourse. Nodes represent policy considerations, with size proportional to mention frequency and color indicating degree centrality. Edges represent statistically significant co-occurrences ($p < 0.025$).

ing coefficient (0.279), the lowest we observed across all three policies. While our simulations explored various discourse patterns, this particularly sparse local connectivity suggests a distinct structure where related considerations rarely form triangular connections, perhaps reflecting the diverse and disconnected nature of agency-specific reform discussions. Despite having nearly as many topics as the LDT debate, the considerations are less interconnected, perhaps because users created distinct threads for niche aspects of the policy. This fragmentation often follows ideological lines: the most isolated nodes represent asymmetric partisan concerns such as DEI and Anti-Woke versus Fraud and Corruption. Conversely, topics centered on practical impacts like Federal Agency Impact and Musk Companies were more interconnected, indicating they were areas of broader, cross-ideological concern.

A deeper analysis of node roles highlights the structural importance of these central themes. Both Federal Agency Impact and Labor Market function as bridges connecting different parts of the conversation, as shown by their high betweenness centralities (0.155 and 0.132). Furthermore, the network reveals a key difference between topic prevalence and influence. Musk Companies was the most frequently mentioned consideration (11.14% incidence) but the discussion was often self-contained (moderate eigenvector centrality of 0.627). In contrast, Federal Agency Impact was mentioned less often (8.62%) but connected to and influenced the entire structure of the debate (maximum eigenvector score of 1.000).

Consideration	Inc. (%)	Deg.	Bet.	Eig.
<i>Panel A: Most Frequently Mentioned</i>				
Musk Companies	11.14	0.278	0.063	0.627
Federal Agencies	8.62	0.444	0.155	1.000
Labor Market	6.69	0.370	0.132	0.808
Social Safety Net	6.05	0.222	0.080	0.542
Net				
Legal Authority	4.86	0.259	0.071	0.570
<i>Panel B: Most Connected Hubs</i>				
Federal Agencies	8.62	0.444	0.155	1.000
Labor Market	6.69	0.370	0.132	0.808
Household Costs	2.75	0.315	0.073	0.744
Musk Companies	11.14	0.278	0.063	0.627
Legal Authority	4.86	0.259	0.071	0.570

Table 6: Top Considerations in DOGE Discourse

The One Big Beautiful Bill Act

The discourse network for BBB is uniquely defined by a thematic divide between social and economic issues. With modularity (0.411) indicating moderate compartmentalization, the network structure suggests parallel conversations about different aspects of this omnibus legislation, a pattern consistent with how complex, multi-faceted bills might naturally generate distinct discussion threads. The conversation is dominated by a large, central cluster focusing on the Social Safety Net, the Labor Market, Immigration, and Federal Agency Impact; considerations that are both frequently mentioned and highly connected (Table 7). This stands in contrast to a smaller, peripheral cluster focused on economic topics like Musk Companies and Financial Markets.

Within this divided structure, certain nodes play important roles in holding the conversation together. The Labor Market is the single most important conceptual bridge, with notably high betweenness centrality (0.348) that far surpasses any other consideration. This analysis also reinforces the distinction between a topic’s popularity and its structural influence. While Social Safety Net (10.44% incidence) and Partisan Conflict (9.52% incidence) were mentioned almost equally, their roles were substantially different. Social Safety Net was a highly influential theme, deeply integrated into the core policy debate (eigenvector centrality of 0.897), while discussions of Partisan Conflict were far more isolated and peripheral to the main arguments.

Robustness to Dictionary Perturbations

Dictionary-based text analysis requires subjective choices about conceptual boundaries and term coverage. To evaluate whether our network findings reflect robust discourse patterns or artifacts of these choices, we conducted systematic sensitivity testing across three perturbation strategies, generating 100 modified dictionaries per strategy for each policy (300 total dictionary variants per policy).

Each perturbation strategy addresses a distinct methodological concern:

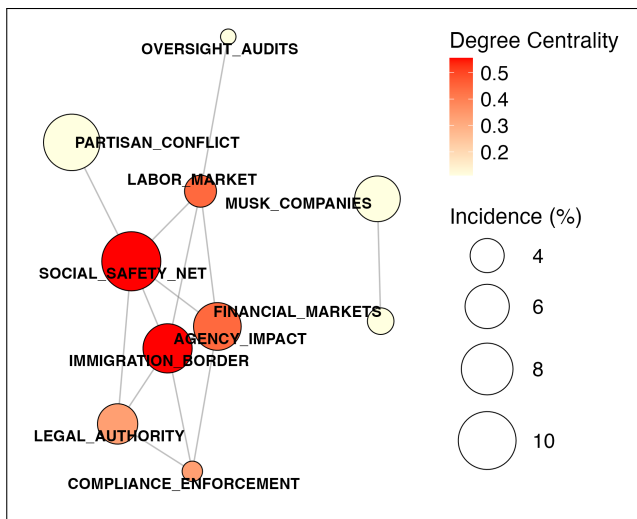


Figure 3: Co-occurrence network for BBB discourse. Nodes represent policy considerations, with size proportional to mention frequency and color indicating degree centrality. Edges represent statistically significant co-occurrences ($p < 0.025$).

Consideration	Inc. (%)	Deg.	Bet.	Eig.
<i>Panel A: Most Frequently Mentioned</i>				
Social Safety Net	10.44	0.310	0.189	0.897
Partisan Conflict	9.52	0.095	0.028	0.318
Immigration	7.27	0.310	0.158	0.942
Federal Agencies	6.86	0.214	0.100	0.717
Musk Companies	6.35	0.119	0.048	0.122
<i>Panel B: Most Connected Hubs</i>				
Labor Market	3.68	0.333	0.348	1.000
Social Safety Net	10.44	0.310	0.189	0.897
Immigration	7.27	0.310	0.158	0.942
Federal Agencies	6.86	0.214	0.100	0.717
Legal Authority	5.12	0.190	0.072	0.517

Table 7: Top Considerations in BBB Discourse

- **Near-Miss Replacement:** Tests sensitivity to conceptual boundaries. We identified 10 plausible “near-miss” considerations excluded from our original dictionary (e.g., *gig_economy* as alternative to *labor_market*, *supply_chain_policy* as alternative to *trade_policy*). For each of 100 runs, we randomly replaced one original consideration with one near-miss alternative. Each replacement introduced 15–20 new terms with verified zero overlap with existing dictionary terms. This addresses whether equally defensible alternative categorizations yield different network structures.
- **Within-Cluster Scrambling:** Tests sensitivity to term assignment. After grouping our 61 considerations into

nine semantic clusters (economic, labor, government, trade, social, political, cultural, technology, environment), we generated 100 variants by randomly reassigning 20–30% of terms between considerations within each cluster. For example, “union dues” might shift from *unions* to *worker_rights* (both in the labor cluster). This simulates realistic boundary ambiguity where human coders might reasonably disagree about term assignments while preserving broad semantic coherence.

- **Coverage Reduction:** Tests sensitivity to term comprehensiveness. We created 100 variants by randomly removing 20–40% of terms from each consideration (removal percentage drawn uniformly per variant). This directly evaluates robustness to incomplete coverage; a fundamental limitation since no dictionary can exhaustively enumerate all relevant expressions. Results show how the estimated network degrades as terms are removed.

For each dictionary variant, we replicated our complete analysis. Table E.1 in Appendix E shows mean percentage changes from baseline across perturbations and policies. Three key findings emerge:

First, network structure proves remarkably robust to alternative conceptualizations (near-miss perturbation). Replacing considerations with plausible alternatives causes minimal structural change: edges vary by only 1–5% and most other metrics remain within $\pm 10\%$. While community counts show moderate variation ($\pm 10\text{--}20\%$), core network properties (density, centralization, and clustering) remain stable.

Second, scrambling terms within clusters suggests our original assignments capture genuine semantic relationships. Within-cluster scrambling increases edge counts by 20–33% across all policies, with BBB showing particularly high sensitivity in community detection (+50%). This counterintuitive increase in connectivity when terms are randomly redistributed validates our original term-to-consideration mappings: the scrambling creates artificial connections between concepts that rarely co-occur naturally.

Third, comprehensive term coverage is essential for accurate network reconstruction. Removing 20–40% of dictionary terms causes substantial structural degradation across all policies: edge counts dropped by 29–40%, centralization drops by 13–39%, while modularity increases by up to 47% (LDT). This pattern (sparser networks with stronger community separation but weaker overall integration) indicates that missing terms fragment the network artificially. The consistency of this effect across all three policy discourses validates our investment in comprehensive dictionary development.

Detailed results including policy-specific findings are provided in Appendix Tables E.2–E.4.

Discussion and Conclusion

This paper uses a permutation-based topic co-occurrence test to map the structure of policy arguments in Reddit discussions of three key policies from Trump’s second administration. Our primary contribution is empirical: we show that each policy debate exhibits a distinct network shape and

identify which considerations occupy central, bridging, and peripheral positions in the argument structure. For instance, the LDT discussion centers on *Trade Ideology* and the *Labor Market*, whereas the BBB discussion is organized around the *Social Safety Net* and *Immigration*. The permutation procedure serves as a simple device to distinguish systematic co-occurrence from chance, ensuring that the observed clusters and bridges reflect substantive argumentative links rather than artifacts of overall frequency. In this way, we extend dictionary-based semantic network analysis by integrating a straightforward randomization-based significance test into topic co-occurrence networks for these specific policy debates, complementing prior work on semantic community detection (Bryden, Funk, and Jansen 2013).

Several limitations qualify these findings. The analysis relies on data from a single subreddit, *r/PoliticalDiscussion*, over a relatively short window in early 2025. The demographic and ideological composition of users participating in these discussions is unlikely to represent the broader population, a common limitation in computational studies of political discourse (Ruths and Pfeffer 2014). As a result, the observed argument structures may not generalize to other platforms, populations, or time periods. In addition, our dictionary-based approach is static and less adaptable than topic-discovery methods: it cannot capture newly emerging slang or considerations that fall outside the predefined lexicon. The co-occurrence procedure itself assumes that joint mentions reflect genuine argumentative structure, but user characteristics and Reddit thread dynamics may also shape which considerations appear together. Some of the links we identify may therefore reflect unobserved user attributes or platform-specific interaction patterns rather than shared conceptual frameworks.

Despite these limitations, our findings show that online political debates follow predictable patterns rather than being random collections of ideas. Topics consistently group together in logical clusters, suggesting these discussions are highly organized, reflecting the structured nature of political belief systems (Converse 2006) and partisan issue bundling (Baldassarri and Gelman 2008). Our network analysis reveals that certain key topics, like the *Labor Market*, are core concepts that connect many different lines of argument, resembling the bridging topics identified in cross-partisan discourse (Garimella et al. 2018). The fact that the conversational structure was so different for each of the three policies suggests that the nature of a policy shapes how the public discusses it, aligning with issue framing theory (Chong and Druckman 2007) and agenda-setting effects on public perceptions (Guo and McCombs 2011). This work contributes to theories of political communication by offering a data-driven method to map the structure of these debates, bridging computational community detection methods with substantive political science questions. By identifying which ideas are central and how they cluster, we can better understand how political narratives are built online.

Our empirical framework can be extended to study the structure and dynamics of political discourse in a wide

range of settings. Comparative analyses could apply similar dictionary-based coding and permutation-tested co-occurrence networks to political subreddits with distinct ideological orientations to quantify how partisanship shapes argumentative structures. Extending the analysis to other social media platforms, such as X (formerly Twitter) and Facebook, would allow examination of platform-specific discourse patterns, building on work showing how different platforms produce distinct political discussion patterns (Halpern and Gibbs 2013). Cross-national comparisons of discourse on substantively similar policies could reveal cultural differences in how considerations are combined. Longitudinal designs could track how co-occurrence networks evolve over a policy's lifecycle, and incorporating user-level data may clarify how individual characteristics influence consideration selection and co-occurrence patterns. Beyond electoral politics, the same empirical strategy could be used to map argument structures in domains such as public health and environmental policy.

As tools capable of generating human-like text become more powerful and accessible, empirical insights about how policy arguments are structured can be misused to craft more persuasive or manipulative messages. Although our analysis is descriptive and does not generate content, the patterns we document—and the simple procedures we use to uncover them—could in principle guide generative systems toward more realistic political messaging. We acknowledge this risk and view explicit discussion of potential misuse as an important ethical context for further applications of this empirical approach.

References

- Baldassarri, D.; and Gelman, A. 2008. Partisans without Constraint: Political Polarization and Trends in American Public Opinion. *American Journal of Sociology*, 114(2): 408–446.
- Barberá, P.; Jost, J. T.; Nagler, J.; Tucker, J. A.; and Bonneau, R. 2015. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10): 1531–1542.
- Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10008.
- Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; and Taylor, J. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, 1247–1250.
- Bryden, J.; Funk, S.; and Jansen, V. A. 2013. Word usage mirrors community structure in the online social network Twitter. *EPJ Data Science*, 2(1): 1–9.
- Chen, E.; Deb, A.; and Ferrara, E. 2022. #Election2020: the first public Twitter dataset on the 2020 US Presidential election. *Journal of Computational Social Science*, 5(1): 1–18.
- Chong, D.; and Druckman, J. N. 2007. Framing Theory. *Annual Review of Political Science*, 10(1): 103–126.

- Converse, P. E. 2006. The Nature of Belief Systems in Mass Publics. *Critical Review*, 18: 1–74. Originally Published 1964.
- Desiderio, A.; Mancini, A.; Cimini, G.; and Di Clemente, R. 2025. Highly engaging events reveal semantic and temporal compression in online community discourse. *PNAS Nexus*, 4(3): pgaf056.
- Ferrer i Cancho, R.; and Solé, R. V. 2001. The small world of human language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1482): 2261–2265.
- FORCE11. 2020. The FAIR Data principles. <https://force11.org/info/the-fair-data-principles/>.
- Fortunato, S. 2010. Community detection in graphs. *Physics Reports*, 486(3-5): 75–174.
- Freking, K.; and Mascaro, L. 2025. What’s in the tax and spending bill that Trump has signed into law. *Associated Press*.
- Garimella, K.; De Francisci Morales, G.; Gionis, A.; and Mathioudakis, M. 2018. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of the 2018 World Wide Web Conference*, 913–922.
- Geburu, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.
- Grimmer, J.; and Stewart, B. M. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3): 267–297.
- Grootendorst, M. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv:2203.05794.
- Guo, L.; and McCombs, M. 2011. Network Agenda Setting: A Third Level of Media Effects. In *Annual Conference of the International Communication Association*. Boston, MA.
- Halpern, D.; and Gibbs, J. 2013. Social Media as a Catalyst for Online Deliberation? Exploring the Affordances of Facebook and YouTube for Political Expression. *Computers in Human Behavior*, 29(3): 1159–1168.
- Jones, K. L.; Noorbaloochi, S.; Jost, J. T.; Bonneau, R.; Nagler, J.; and Tucker, J. A. 2018. Liberal and conservative values: What we can learn from congressional tweets. *Political Psychology*, 39(2): 423–443.
- King, G.; Pan, J.; and Roberts, M. E. 2017. How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review*, 111(3): 484–501.
- Kloo, I.; Cruickshank, I. J.; and Carley, K. M. 2024. A cross-platform topic analysis of the nazi narrative on Twitter and Telegram during the 2022 russian invasion of Ukraine. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 839–850.
- Lancichinetti, A.; Radicchi, F.; Ramasco, J. J.; and Fortunato, S. 2011. Finding Statistically Significant Communities in Networks. *PLoS ONE*, 6(4): e18961.
- Levin, I. 2022. Learning about spatial and temporal proximity using tree-based methods. *Statistics, Politics and Policy*, 13(1): 73–95.
- Megerian, C. 2025. Trump 100 days: DOGE. *Associated Press*.
- Newman, M. E. J. 2006. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23): 8577–8582.
- Pew Research Center. 2025. Trump’s Job Rating Drops, Key Policies Draw Majority Disapproval as He Nears 100 Days. Technical report, Pew Research Center.
- Roberts, M. E.; Stewart, B. M.; and Tingley, D. 2019. stm: An R package for structural topic models. *Journal of Statistical Software*, 91: 1–40.
- Roberts, M. E.; Stewart, B. M.; Tingley, D.; and Airoidi, E. M. 2013. The structural topic model and applied social science. In *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation*, volume 4, 1–20.
- Roberts, M. E.; Stewart, B. M.; Tingley, D.; Lucas, C.; Leder-Luis, J.; Gadarian, S. K.; Albertson, B.; and Rand, D. G. 2014. Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4): 1064–1082.
- Rozen, C. 2025. Exclusive: DOGE ‘doesn’t exist’ with eight months left on its charter. *Reuters*.
- Ruths, D.; and Pfeffer, J. 2014. Social media for large studies of behavior. *Science*, 346(6213): 1063–1064.
- Sagarzazu, I. n.; and Klüver, H. 2017. Coalition governments and party competition: Political communication strategies of coalition parties. *Political Science Research and Methods*, 5(2): 333–349.
- Spaiser, V.; Chadefaux, T.; Donnay, K.; Russmann, F.; and Helbing, D. 2017. Communication power struggles on social media: A case study of the 2011–12 Russian protests. *Journal of Information Technology & Politics*, 14(2): 132–153.
- Strobl, C.; Boulesteix, A.-L.; Kneib, T.; Augustin, T.; and Zeileis, A. 2008. Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1): 307.
- Tuters, M.; Willaert, T.; and Meyer, T. 2023. How science gets drawn into global conspiracy narratives. *Issues in Science and Technology*, 39(3): 32–36.
- Van Vliet, L.; Törnberg, P.; and Uitermark, J. 2020. The Twitter parliamentary database: Analyzing Twitter politics across 26 countries. *PLoS one*, 15(9): e0237073.
- Xiong, Y.; Cho, M.; and Boatwright, B. 2019. Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of Twitter during the #MeToo movement. *Public Relations Review*, 45(1): 10–23.
- Zhang, J.; Hamilton, W. L.; Danescu-Niculescu-Mizil, C.; Jurafsky, D.; and Leskovec, J. 2017. Community identity and user engagement in a multi-community landscape. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, 377–386.

Paper Checklist

1. For most authors...

- (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? **Yes.** The paper analyzes publicly available data from Reddit to understand the structure of political discourse. The final section includes a discussion of ethical implications and the potential for misuse.
- (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? **Yes.** The abstract and introduction say that the paper proposes a permutation-based method to identify significant topic co-occurrences and an application of this method to analyze Reddit discussions. The paper delivers on this by detailing the methodology and presenting network analysis results for three distinct policy debates.
- (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? **Yes.** The paper justifies the use of a permutation-based statistical test to move beyond simple co-occurrence counts and identify statistically meaningful relationships. The "Background and Related Work" section contrasts this approach with alternatives like STM and hashtag analysis, arguing for its suitability for the research question.
- (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? **Yes.** The limitations section explicitly states that the data is from a single subreddit, `r/PoliticalDiscussion`, and that its user base is not representative of the general population, which may influence the observed discourse patterns.
- (e) Did you describe the limitations of your work? **Yes.** A dedicated paragraph in the "Discussion and Conclusion" section outlines several limitations, including the non-representative nature of the Reddit data, the static nature of the dictionary-based approach, and the inability to control for user-level characteristics.
- (f) Did you discuss any potential negative societal impacts of your work? **Yes.** The final paragraph addresses this directly, noting that the analytical techniques can serve as a strategic guide for generative AI to create more believable messaging.
- (g) Did you discuss any potential misuse of your work? **Yes.** The final paragraph explicitly discusses how the method's insights could be used to create text that exploits known thematic connections, which would be a form of manipulation.
- (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? **Yes.** This is described in a footnote in

the Methodology section. We detail our approach to data anonymization and the release of an anonymized dataset to ensure reproducibility.

- (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? **Yes.** The paper includes a comprehensive discussion of limitations and ethical considerations, including potential negative societal impacts and misuse, which aligns with ethics review guidelines.
- ### 2. Additionally, if your study involves hypotheses testing...
- (a) Did you clearly state the assumptions underlying all theoretical results? **Yes.** The paper explains that the permutation test operates under the null hypothesis that topic co-occurrences are random (i.e., independent). The simulation procedure is designed to generate a null distribution based on this assumption.
 - (b) Have you provided justifications for all theoretical results? **N/A.** The paper presents empirical findings rather than formal theoretical results or proofs.
 - (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? **Yes.** The paper discusses alternative methodological approaches (e.g., STM, hashtag analysis) and also considers alternative explanations for the observed co-occurrence patterns in the limitations section, such as linguistic conventions or platform dynamics.
 - (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? **Yes.** The limitations section proposes that user characteristics and Reddit thread dynamics may influence the choice of considerations and contribute to the observed co-occurrence patterns.
 - (e) Did you address potential biases or limitations in your theoretical framework? **N/A.** The paper does not present a formal theoretical framework.
 - (f) Have you related your theoretical results to the existing literature in social science? **N/A.** The paper has no theoretical results, but it connects its methodology and empirical findings to the social science literature on text and network analysis.
 - (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? **N/A.** The paper does not have theoretical results, but the "Discussion and Conclusion" section details the implications of its empirical findings for understanding political discourse and outlines numerous avenues for future research.
- ### 3. Additionally, if you are including theoretical proofs...
- (a) Did you state the full set of assumptions of all theoretical results? **N/A.** The paper does not contain theoretical proofs.
 - (b) Did you include complete proofs of all theoretical results? **N/A.** The paper does not contain theoretical proofs.
- ### 4. Additionally, if you ran machine learning experiments...

- (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? **Yes. All replication materials including R code, anonymized data, LLM prompts, and dictionaries are available in the replication package.**
- (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? **Yes. For the permutation-based procedure, the paper specifies the key hyperparameters: 1,000 simulation iterations and a 97.5th percentile significance threshold. It also identifies the LLM used for dictionary creation.**
- (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? **The paper reports network metrics as point estimates.**
- (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? **Yes. The analysis was performed on a standard laptop; a MacBook Air with an Apple M3 chip (8-core CPU, 10-core GPU, 16-core Neural Engine) and 24GB of unified memory.**
- (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? **Yes. The paper’s evaluation framework rests on applying the statistical method to real-world data and demonstrating that the resulting networks are interpretable and reveal distinct structures, which is appropriate for a paper introducing a new methodology for discourse analysis.**
- (f) Do you discuss what is “the cost” of misclassification and fault (in)tolerance? **The paper implicitly sets a statistical tolerance for false positives (a Type I error rate of 2.5%).**
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity...**
- (a) If your work uses existing assets, did you cite the creators? **Yes. The paper names the key R package used (“RedditExtractoR”) and cites the creators of related methodologies like STM and BERTopic in the literature review.**
- (b) Did you mention the license of the assets? **No. The licenses of the software packages used are not mentioned.**
- (c) Did you include any new assets in the supplemental material or as a URL? **Yes. All replication materials including the initial LLM prompts, full unified dictionary, and R code are available in the replication package.**
- (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? **No. The paper uses public data from Reddit and, in line with common practice for such research, does not discuss user consent.**
- (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? **No. The paper does not explicitly discuss the presence of PII (except for usernames in the author field, which we anonymized using a hashing algorithm augmented with a private cryptographic salt) or potentially offensive content in the Reddit data.**
- (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? **N/A. The paper does not indicate an intention to release a new dataset.**
- (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? **N/A. The paper does not indicate an intention to release a new dataset.**
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity...**
- (a) Did you include the full text of instructions given to participants and screenshots? **N/A. The study analyzed public social media data and did not involve direct interaction with human subjects.**
- (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? **N/A. The study did not involve human subjects in a way that would typically require IRB approval.**
- (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? **N/A. No participants were compensated.**
- (d) Did you discuss how data is stored, shared, and deidentified? **N/A. The study did not involve human subjects in a way that required data deidentification procedures.**

Appendix to “Permutation-Based Testing of Topic Co-occurrence: A Network Analysis of Reddit Debates on DOGE, Tariffs, and the Big Beautiful Bill” by Benjamin Forman-Barzilai and Ines Levin

Appendix A: Prompt Structure for Dictionary Generation

For each of the three policy episodes (DOGE, Liberation Day Tariffs, and the Big Beautiful Bill), we constructed prompts with a common structure and only the policy-specific details varied. Each prompt contained the following elements:

- 1. General purpose and task framing.** All prompts opened with the same explanation that we were creating a thematic dictionary to code people’s opinions about a specific political or policy event based on social media content. The model was instructed to:
 - (a) Identify a comprehensive set of distinct “considerations” that could influence whether someone supports or opposes the policy; and
 - (b) For each consideration, list 20–30 keywords or short phrases that people might use in online discussion (e.g., tweets, Facebook posts, Reddit comments).
- 2. Constraints on considerations and terms.** We imposed identical stylistic and structural rules across policies:
 - Considerations must be broad enough to represent a dimension of opinion, but distinct enough to avoid overlap with other considerations.
 - Terms must be single words or short multi-word expressions (not full sentences), formulated as reusable phrases (e.g., “low morale” rather than “morale is low”).
 - Hashtags were allowed only if already in documented use; the model was instructed not to invent new hashtags or expressions.
 - Terms should include commonly used synonyms, slang, abbreviations, and emotionally charged language, reflecting ordinary, non-expert ways of talking about the topic.
 - Within each consideration, terms should span positive, negative, and neutral sentiment.
 - Overlapping terms across considerations were to be avoided, and style was to be kept consistent across categories.
- 3. Cross-policy Trump framing cues.** Each prompt included a shared “additional instruction” section directing the model to account for recurring themes that Donald Trump often invokes when framing policy debates (e.g., DEI controversies, immigration and border security, trade and tariffs, skepticism of official statistics, religion and cultural values, law-and-order messaging, executive authority, government waste and efficiency, federal research funding, campus antisemitism and free speech, cryptocurrency). The model was told to include such considerations when relevant and to keep them broad and applicable across policy contexts.
- 4. Policy-specific event description.** For each policy, we provided a detailed narrative describing the timing, legal basis, implementation, and political debate surrounding the policy episode. These descriptions highlighted likely implications (e.g., for employment, prices, deficits, social programs, national security) to anchor the considerations in the substantive context of each policy.
- 5. Policy-specific suggested considerations.** Each prompt listed a set of suggested considerations tailored to that specific policy (e.g., workforce effects, service delivery, fiscal savings for DOGE; prices, retaliation, supply chains for tariffs; growth, deficits, social programs, climate for the Big Beautiful Bill). The model was explicitly allowed to adapt, substitute, or expand this list and to supplement it with cross-policy Trump themes where appropriate.
- 6. Standardized output schema.** Finally, all prompts specified an identical R output format: each dictionary was to be returned as an R `list` where each consideration is a named element with three fields:
 - `label`: a unique, lowercase, snake_case identifier;
 - `name`: a human-readable descriptive title for the consideration; and
 - `terms`: a vector of 20–30 lowercase terms or short phrases reflecting everyday language.The prompts included a small R code example (e.g., `doge_dict <- list(...)`), and the same schema was used for all three policy dictionaries.

Appendix B: r/PoliticalDiscussion Subreddit Rules

The following rules are reproduced verbatim from the r/PoliticalDiscussion subreddit as of January 09, 2026, with hyperlinks rendered as plain text (not clickable). These guidelines illustrate the community's emphasis on substantive, civil political discourse.

[Beginning of verbatim reproduction]

SUBMISSION RULES

New submissions will not appear until approved by a moderator.

Wiki Guide: Tips On Writing a Successful Political Discussion Post

Please observe the following rules:

1. Submissions should be an impartial discussion prompt + questions.
 - Keep it civil, no political name-calling.
 - Do not ask loaded or rhetorical questions.
 - No personal opinions/proposals -> r/politicalopinions for those.
2. Provide some background and context. Offer substantive avenues for discussion.
 - Avoid highly speculative posts, all scenarios should be within the realm of reasonable possibility.
 - Do not request users help you with an argument, educate you, or perform research for you.
 - No posts that boil down to: DAE, ELI5, CMV, TIL, AskA, ``Thoughts?``, ``Discuss!``, or ``How does this affect the election?``
3. Everything in the post should be directly related to a political issue.
 - No meta discussion about reddit, subreddits, or redditors.
 - We are not a link subreddit. Don't just post links to news, blogs, surveys, videos, etc.
 - Potentially non-politics: Law, sociology, philosophy, celebrities, news, etc.
4. Formatting and housekeeping things:
 - The title should match the post. Don't use tags like [Serious]
 - Check to make sure another recent post doesn't already cover that topic.
 - Don't use all-caps. Format for readability: paragraphs, punctuation, and link containers.

COMMENT RULES

Keep this subreddit HIGH QUALITY by observing Reddiquette and our comment rules:

- **Keep it civil.** Do not personally insult other Redditors, or post racist, sexist, homophobic, or otherwise discriminatory content. Constructive debate is good; mockery, taunting, and name calling are not.
- **Do not submit low investment content.** This subreddit is for genuine discussion. Low effort content, including memes, links substituting for explanation, sarcasm, and non-substantive contributions will be removed per moderator discretion.
- **No meta discussion.** This is not a subreddit for discussing what's going on on reddit; conversation should be focused on the topic at hand. Meta content includes talking about reddit, other subreddits, redditors, and moderators. All meta content will be removed.
- **Warnings.** The rules are intended to maintain the high quality of the subreddit, and garden-variety violations will be met with a reminder from the moderators. If you would like to have your comment reinstated, please edit the offending material and let the moderators know via modmail. Users who demonstrate an inability to consistently follow our rules will however be banned at moderator discretion. Please also note that severe violations of the civility rule may result in an immediate permanent ban.

Further details regarding these rules are available here, and recent clarifications to the civility rule regarding taunting and mockery can be found here.

[End of verbatim reproduction]

Appendix C: Dictionary Validation

The following tables present detailed results from our dictionary validation exercise, in which two coders independently evaluated the quality of term-to-consideration assignments. We employed three complementary validation tasks to assess different aspects of the dictionary’s reliability and validity:

Table C.1 reports inter-rater reliability when coders directly evaluated whether 100 randomly sampled terms were appropriately assigned to their designated considerations. Table C.2 presents results from a more demanding blind classification task, where coders independently assigned the same 100 terms to considerations without knowing the original labels. Table C.3 shows signal detection performance when coders distinguished between correct and intentionally incorrect term-consideration pairings (200 total pairings), testing their ability to discriminate valid from invalid assignments.

These validation results demonstrate strong reliability of the dictionary assignments. Task 1 shows high face validity, with both coders accepting 84% of term-consideration pairings simultaneously (Gwet’s AC1 = 0.85). Task 2’s lower agreement (48% both correct) is expected given the demanding nature of blind classification from 61 possible considerations; the 55–75% individual accuracy rates substantially exceed chance (1.6%). Task 3’s strong signal detection performance ($d' > 3.0$) confirms that coders can reliably distinguish valid from invalid pairings, suggesting the dictionary captures meaningful semantic distinctions rather than arbitrary groupings.

Metric	Value
Total terms evaluated	100
Perfect agreement	86%
Both accept	84%
Coder 1 accepts	95%
Coder 2 accepts	87%
Gwet’s AC1	0.847

Note: Coders independently evaluated whether 100 randomly sampled terms were appropriately assigned to their designated considerations, indicating “accept” if the pairing was valid. Perfect agreement indicates both coders gave the same response (both accept or both reject). “Both accept” shows the percentage where both coders endorsed the pairing simultaneously. Gwet’s AC1 is a chance-corrected agreement coefficient robust to prevalence effects; values > 0.80 indicate strong agreement.

Table C.1: Task 1: Inter-Rater Reliability for Term-Consideration Assignments

Metric	Value
Terms classified	100
Coder 1 accuracy	55%
Coder 2 accuracy	75%
Both correct	48%
Classification agreement	55%
Cohen’s κ	0.540

Note: Coders independently assigned 100 terms to considerations from all 61 possible categories without knowing the original labels (blind classification). Accuracy measures match with original assignments. “Both correct” indicates terms where both coders matched the original assignment. “Classification agreement” shows cases where coders chose the same consideration (regardless of whether it matched the original). Cohen’s κ is a chance-corrected agreement coefficient; values of 0.41–0.60 indicate moderate agreement, > 0.60 substantial agreement.

Table C.2: Task 2: Blind Classification Validation Results

Metric	Value
Pairings evaluated	200
Coder 1 accuracy	94%
Coder 2 accuracy	95%
Coder 1 d'	3.03
Coder 2 d'	3.67
Agreement	91%
Cohen’s κ	0.810
Gwet’s AC1	0.810

Note: Coders evaluated 200 term-consideration pairings (100 correct, 100 intentionally mismatched) to distinguish valid from invalid assignments. Accuracy indicates correct identification of valid/invalid pairings. d' (d -prime) measures discriminability in signal detection; values of 1.0 indicate moderate discrimination, 2.0 good discrimination, and > 3.0 excellent discrimination. Agreement shows percentage of pairings where coders gave the same judgment. Cohen’s κ and Gwet’s AC1 are chance-corrected agreement coefficients; values > 0.80 indicate strong agreement.

Table C.3: Task 3: Signal Detection Performance

Appendix D: Method Validation

Tables D.1–D.4 present network metrics for the four simulated discourse patterns used in our validation exercise. For each pattern (Integrated, Polarized, Fragmented, and Center-Periphery) we report the most frequently mentioned considerations (Panel A) and the most connected hubs (Panel B), along with their network metrics.

The simulation results confirm that our method recovers distinct structural signatures in each hypothetical discourse pattern (Tables D.1–D.4). The Polarized pattern features a dominant bridge node (`transparency`) with high betweenness centrality (0.456), connecting two ideological clusters. The Fragmented pattern exhibits lower degree centralities overall and high modularity (0.666), indicating isolated topical silos. The Center-Periphery structure shows concentration around `household_costs` (degree 0.405) and high eigenvector centrality among core topics, while the Integrated pattern displays moderate connectivity across multiple nodes without dominant hubs. Taken together, these simulated patterns provide reference points for interpreting the network structures observed in the Reddit data, while acknowledging that real discourse may exhibit more complex hybrid configurations.

Consideration	Inc (%)	Deg	Bet	Eig
Panel A: Top 5 Most Frequently Mentioned				
<code>financial_markets</code>	20.20	0.196	0.058	0.835
<code>public_trust</code>	19.73	0.161	0.060	0.274
<code>household_costs</code>	19.00	0.268	0.067	0.990
<code>tax_revenue</code>	17.93	0.250	0.029	1.000
<code>worker_rights</code>	17.87	0.214	0.080	0.625
Panel B: Top 5 Most Connected Hubs				
<code>household_costs</code>	19.00	0.268	0.067	0.990
<code>tax_revenue</code>	17.93	0.250	0.029	1.000
<code>macro_performance</code>	15.40	0.232	0.056	0.878
<code>labor_market</code>	16.13	0.214	0.069	0.806
<code>worker_rights</code>	17.87	0.214	0.080	0.625

Note: Inc = Incidence; Deg = Degree centrality; Bet = Betweenness centrality; Eig = Eigenvector centrality.

Table D.1: Top Considerations in the Integrated Discourse

Consideration	Inc (%)	Deg	Bet	Eig
Panel A: Top 5 Most Frequently Mentioned				
<code>household_costs</code>	78.73	0.478	0.040	0.463
<code>transparency</code>	72.87	0.870	0.456	1.000
<code>worker_rights</code>	38.33	0.435	0.000	0.355
<code>tax_revenue</code>	37.93	0.522	0.007	0.779
<code>fraud_corruption</code>	37.53	0.522	0.007	0.779
Panel B: Top 5 Most Connected Hubs				
<code>transparency</code>	72.87	0.870	0.456	1.000
<code>compliance_enforce</code>	37.33	0.565	0.050	0.814
<code>small_business</code>	36.73	0.522	0.007	0.779
<code>fiscal_savings</code>	35.87	0.522	0.007	0.779
<code>fraud_corruption</code>	37.53	0.522	0.007	0.779

Note: Inc = Incidence; Deg = Degree centrality; Bet = Betweenness centrality; Eig = Eigenvector centrality.

Table D.2: Top Considerations in the Polarized Discourse

Consideration	Inc (%)	Deg	Bet	Eig
Panel A: Top 5 Most Frequently Mentioned				
<code>fiscal_savings</code>	19.33	0.184	0.009	1.000
<code>national_security</code>	18.00	0.122	0.009	0.000
<code>tax_revenue</code>	17.33	0.143	0.001	0.874
<code>macro_performance</code>	16.73	0.184	0.037	0.970
<code>export_access</code>	16.60	0.122	0.004	0.000
Panel B: Top 5 Most Connected Hubs				
<code>macro_performance</code>	16.73	0.184	0.037	0.970
<code>fiscal_savings</code>	19.33	0.184	0.009	1.000
<code>data_governance</code>	12.73	0.163	0.014	0.904
<code>tech_modernization</code>	11.87	0.163	0.004	0.952
<code>equity_vulnerable</code>	11.00	0.163	0.011	0.841

Note: Inc = Incidence; Deg = Degree centrality; Bet = Betweenness centrality; Eig = Eigenvector centrality.

Table D.3: Top Considerations in the Fragmented Discourse

Consideration	Inc (%)	Deg	Bet	Eig
Panel A: Top 5 Most Frequently Mentioned				
household_costs	18.53	0.405	0.022	0.995
environment_energy	12.40	0.333	0.001	0.890
transparency	11.07	0.286	0.000	0.791
social_safety_net	10.93	0.405	0.144	0.900
public_trust	10.80	0.381	0.113	0.947
Panel B: Top 5 Most Connected Hubs				
fiscal_savings	9.40	0.429	0.065	1.000
student_loans	9.47	0.429	0.084	0.962
household_costs	18.53	0.405	0.022	0.995
tax_revenue	8.93	0.405	0.032	0.963
social_safety_net	10.93	0.405	0.144	0.900

Note: Inc = Incidence; Deg = Degree centrality; Bet = Betweenness centrality; Eig = Eigenvector centrality.

Table D.4: Top Considerations in the Center-Periphery Discourse

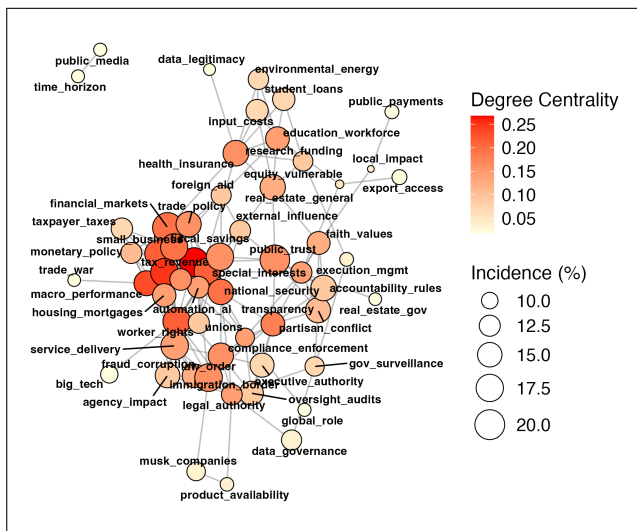


Figure D.1: Integrated discourse network showing overlapping clusters with substantial cross-cutting ties, representing deliberative discourse where multiple considerations regularly co-occur.

Figures D.1–D.4 visualize the network structures generated from our hypothetical discourse simulations. Nodes are sized by mention frequency and colored by degree centrality, with edges representing statistically significant co-occurrences above significance testing thresholds ($p < 0.025$).

The network visualizations confirm the statistical metrics. The Integrated network (Figure D.1) shows overlapping clusters with extensive cross-cutting ties. The Polarized network (Figure D.2) displays clear bifurcation with minimal bridging. The Fragmented network (Figure D.3) exhibits distinct isolated clusters. The Center-Periphery network (Figure D.4) shows a dominant hub structure. These visual patterns align with the quantitative metrics, validat-

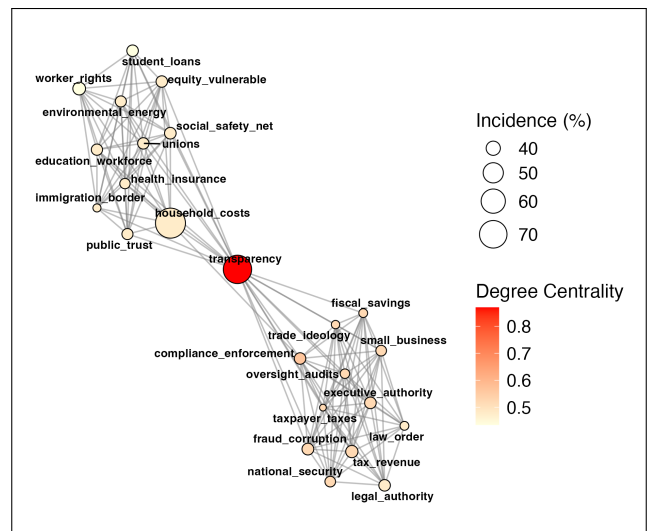


Figure D.2: Polarized discourse network exhibiting two distinct clusters with limited bridging connections, characteristic of ideologically divided discussion.

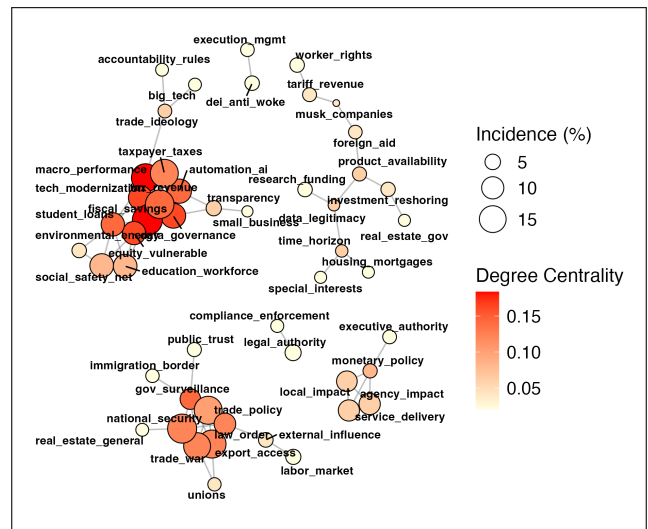


Figure D.3: Fragmented discourse network displaying isolated topic silos with minimal inter-cluster connections, indicating compartmentalized discussion.

ing our method's ability to distinguish meaningful discourse structures.

Appendix E: Sensitivity to Dictionary Perturbations

Table E.1 shows mean percentage changes from baseline across perturbations and policies. These results are discussed in the *Robustness to Dictionary Perturbations* section in the main text.

Metric	Perturbation	LDT	BBB	DOGE
<i>Structural Properties</i>				
Nodes	Near-miss	-1.5	-1.9	-0.5
	Scrambling	+5.3	+22.0	+9.0
	Coverage	-7.9	-7.0	-9.1
Edges	Near-miss	-1.0	-4.5	-4.8
	Scrambling	+32.9^a	+26.7^a	+20.4
	Coverage	-40.1^a	-31.6^a	-28.8^a
<i>Community Structure</i>				
Communities	Near-miss	-19.2	+20.0	+10.0
	Scrambling	-16.5	+49.6^a	+4.8
	Coverage	-0.3	+59.6^a	+0.2
Modularity	Near-miss	+0.8	-4.0	+2.5
	Scrambling	-3.1	-4.4	+5.6
	Coverage	+46.9^a	+2.4	+28.9^a
<i>Network Integration</i>				
Centralization	Near-miss	+2.7	-3.2	-2.2
	Scrambling	+5.9	+1.7	-9.3
	Coverage	-28.9^a	-12.9	-38.6^a
Clustering	Near-miss	-3.0	-6.3	-3.2
	Scrambling	+5.4	+14.2	+16.1
	Coverage	-22.9	-10.3	-1.5

Note: Values show mean % change from baseline across 100 runs per perturbation. Near-miss = replacing one consideration with plausible alternative; Scrambling = reassigning 20–30% of terms within semantic clusters; Coverage = removing 20–40% of terms from all considerations.

^a Absolute change >25%.

Table E.1: Network Sensitivity to Dictionary Perturbations

Tables E.2–E.4 present sensitivity analyses examining how our network metrics respond to three types of dictionary perturbations. These tests assess whether our findings are robust to potential measurement error in topic identification.

We implemented three perturbation strategies: (P1) Near-miss replacement, which substitutes one policy consideration with a semantically plausible alternative; (P2) Within-cluster scrambling, which randomly reassigns 20–30% of terms within their semantic clusters; and (P3) Coverage reduction, which removes 20–40% of terms from all considerations. For each perturbation, we generated 100 runs and report mean values with standard deviations.

The results indicate that our core network structures remain stable under modest dictionary modifications. While P3 (coverage reduction) produces the largest changes (particularly in edge counts and modularity), the overall network patterns persist. Changes exceeding 25% are marked for readability.

Metric	Value	% change
<i>Baseline</i>		
Nodes	57	–
Edges	294	–
Communities	6	–
Modularity	0.219	–
Centralization	0.352	–
Clustering	0.386	–
<i>P1: Near-miss replacement (n=100)</i>		
Nodes	56.2 (0.4)	-1.5
Edges	291.2 (8.9)	-1.0
Communities	4.8 (0.8)	-19.2
Modularity	0.221 (0.012)	+0.8
Centralization	0.361 (0.010)	+2.7
Clustering	0.375 (0.007)	-3.0
<i>P2: Within-cluster scrambling (n=100)</i>		
Nodes	60.0 (0.1)	+5.3
Edges	390.7 (2.9)	+32.9^a
Communities	5.0 (0.1)	-16.5
Modularity	0.212 (0.001)	-3.1
Centralization	0.372 (0.000)	+5.9
Clustering	0.407 (0.002)	+5.4
<i>P3: Coverage reduction (n=100)</i>		
Nodes	52.5 (0.5)	-7.9
Edges	176.0 (5.2)	-40.1^a
Communities	6.0 (0.1)	-0.3
Modularity	0.322 (0.015)	+46.9^a
Centralization	0.250 (0.017)	-28.9^a
Clustering	0.298 (0.002)	-22.9

Note: P1 replaces one consideration with a plausible alternative; P2 reassigns 20–30% of terms within semantic clusters; P3 removes 20–40% of terms from all considerations. Entries in the Value column for P1–P3 are mean (standard deviation) across 100 runs; baseline rows report observed values in the empirical data.

^a Absolute change >25%.

Table E.2: Sensitivity Analysis: Liberation Day Tariffs

The sensitivity analyses reveal differential robustness across perturbation types. Near-miss replacement (P1) produces minimal structural changes (typically <5% for edges), confirming that our findings are not artifacts of specific categorical boundaries. Within-cluster scrambling (P2) increases edge counts by 20–33%, validating our original term assignments (random redistribution creates spurious connections absent in the correctly specified dictionary). Coverage reduction (P3) causes the most substantial changes, with edge counts dropping 29–40% and modularity increasing up to 47%, indicating that comprehensive term coverage is essential for accurate network reconstruction.

The consistent pattern across all three policies (stability under P1, edge inflation under P2, and fragmentation under P3) demonstrates that our main findings reflect robust discourse structures rather than dictionary artifacts, provided the dictionary maintains comprehensive coverage within each consideration.

Metric	Value	% change
<i>Baseline</i>		
Nodes	43	–
Edges	98	–
Communities	5	–
Modularity	0.411	–
Centralization	0.225	–
Clustering	0.313	–
<i>P1: Near-miss replacement (n=100)</i>		
Nodes	42.2 (0.4)	-1.9
Edges	93.6 (3.4)	-4.5
Communities	6.0 (0.6)	+20.0
Modularity	0.394 (0.021)	-4.0
Centralization	0.218 (0.011)	-3.2
Clustering	0.294 (0.018)	-6.3
<i>P2: Within-cluster scrambling (n=100)</i>		
Nodes	52.5 (1.5)	+22.0
Edges	124.2 (1.5)	+26.7^a
Communities	7.5 (0.5)	+49.6^a
Modularity	0.393 (0.005)	-4.4
Centralization	0.229 (0.013)	+1.7
Clustering	0.358 (0.015)	+14.2
<i>P3: Coverage reduction (n=100)</i>		
Nodes	40.0 (0.1)	-7.0
Edges	67.0 (0.3)	-31.6^a
Communities	8.0 (0.2)	+59.6^a
Modularity	0.420 (0.006)	+2.4
Centralization	0.196 (0.003)	-12.9
Clustering	0.281 (0.002)	-10.3

Note: P1 replaces one consideration with a plausible alternative; P2 reassigns 20–30% of terms within semantic clusters; P3 removes 20–40% of terms from all considerations. Entries in the Value column for P1–P3 are mean (standard deviation) across 100 runs; baseline rows report observed values in the empirical data.

^a Absolute change >25%.

Table E.3: Sensitivity Analysis: Big Beautiful Bill

Metric	Value	% change
<i>Baseline</i>		
Nodes	55	–
Edges	198	–
Communities	5	–
Modularity	0.295	–
Centralization	0.311	–
Clustering	0.279	–
<i>P1: Near-miss replacement (n=100)</i>		
Nodes	54.8 (0.8)	-0.5
Edges	188.5 (2.1)	-4.8
Communities	5.5 (0.5)	+10.0
Modularity	0.303 (0.011)	+2.5
Centralization	0.304 (0.007)	-2.2
Clustering	0.271 (0.007)	-3.2
<i>P2: Within-cluster scrambling (n=100)</i>		
Nodes	60.0 (0.7)	+9.0
Edges	238.4 (12.0)	+20.4
Communities	5.2 (0.8)	+4.8
Modularity	0.312 (0.015)	+5.6
Centralization	0.282 (0.039)	-9.3
Clustering	0.324 (0.039)	+16.1
<i>P3: Coverage reduction (n=100)</i>		
Nodes	50.0 (0.1)	-9.1
Edges	140.9 (0.7)	-28.8^a
Communities	5.0 (0.1)	+0.2
Modularity	0.381 (0.001)	+28.9^a
Centralization	0.191 (0.001)	-38.6^a
Clustering	0.275 (0.001)	-1.5

Note: P1 replaces one consideration with a plausible alternative; P2 reassigns 20–30% of terms within semantic clusters; P3 removes 20–40% of terms from all considerations. Entries in the Value column for P1–P3 are mean (standard deviation) across 100 runs; baseline rows report observed values in the empirical data.

^a Absolute change >25%.

Table E.4: Sensitivity Analysis: DOGE