Network Science (2026), **14**, e27, 1–20
doi:10.1017/nws.2026.10046

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

The structure of reasoning: Inferring conceptual networks from text

Sarah Shugars and Xinfeng Gu

Department of Communication, Rutgers University, New Brunswick, NJ, USA
Corresponding author: Sarah Shugars; Email: sarah.shugars@rutgers.edu

Abstract

For decades, public opinion scholars have argued for the need to go beyond measuring isolated political preferences to more richly examine how individuals reason about and justify the interconnections between their preferences. While early efforts used interviews and hand-coding to elicit the network structure of subjects' political reasoning, this manual approach can be time-consuming and difficult. Thankfully, modern computational methods and computing power hold the potential to revolutionize this space and allow for network inference at scale. In this paper, we therefore present a computational, text-based approach for inferring the network structure of individuals' political reasoning. This method identifies the key concepts a person raises and examines the implicit connections between those concepts—what ideas do they connect to which other ideas? This structural approach is theoretically justified in both the cognitive and linguistic literatures, which repeatedly suggest that humans store, retrieve, and interpret information through network structures. We show that this approach reveals meaningful individual variation that is correlated with known behavioral traits. The ability to measure and interrogate individuals' expressions of political reasoning holds the potential to shed new light on the dynamics of public opinion and political behavior. Questions of persuasion, ideological fracturing, and conversation quality all rely upon understanding individual styles of political expression. These dynamics are driven not just by what someone says but by how they say it.

Keywords: Socio-semantic networks; conceptual networks; cognitive networks; reasoning; deliberation; natural language processing

1. Introduction

The individually distinctive ways in which people express their preferences holds the potential to reveal broader variations in political behavior. A population's agreement on a given policy position, for example, may elide deeper divisions in the motivation behind that position. Similarly, discussants with opposing preferences may, under some circumstances, be able to find ways to engage productively across their differences. While the bulk of public opinion literature has reasonably focused on the output of what people believe—that is, individuals' discrete policy preferences—a robust understanding of the full deliberative system (Mansbridge, 1999) further requires analysis of how people express these preferences and interpret the preferences of others. While political preference models are invaluable for capturing trends in public opinion and predicting policy outcomes, they are not designed to analyze networks of preferences or interactions between individuals' preferences. What arguments can lead to opinion change and under what circumstances? What factors drive a political conversation to be productive or divisive? How can a society function democratically in the face of increasing levels of affective polarization? If we

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

hope to answer such critical questions of public opinion, we need network models of individuals' political reasoning and expression.

The call for such models is not new, but the computational tools needed to develop them are. A notable line of classic public opinion research (Axelrod, 1976; Campbell et al., 1960; Lane, 1962) used semi-structured interviews or hand-coding of texts in an effort to capture individual variation in the articulation of political preferences. More recent work has similarly argued that political ideology functions by binding together seemingly unrelated positions, such as racial and economic liberalism, into stable coalitions (Green, 2024; Noel, 2012). Such ideological bundling is not necessarily logical, but rather reflects a structural sense of "what goes with what" for a given ideology.

From a normative standpoint, studying individual variation in reasoning structures acknowledges democratic ideals of citizen voice and opens the door for examining strategies of moving toward this ideal. From a practical stand point, this variation holds the potential to reflect the success and failures of elite messaging: even if average citizens primarily repeat elite talking points, it is still worth examining which talking points they find themselves repeating. While early efforts at modeling individual-level reasoning were often abandoned as too arduous and time-consuming, modern computational methods hold the promise to meaningfully revive these efforts.

This paper presents an initial step in such a revival and seeks to demonstrate that conceptual network structures (1) can be computationally inferred from text and (2) meaningfully reflect individuals' personal traits. This work is largely exploratory; recovering a lost tradition and shedding insight into conflicting behavioral priors. The core argument is that individuals talk about politics in subtly different and unique ways. Examining the *structure* of expressed reasoning, separate from the *content* of those reasons, can therefore shed insight into variations in political behavior. This paper aims to capture the structure of how ordinary citizens connect political ideas, above and beyond the specific content of those ideas. This, of course, is not to argue that the content of political preferences are not themselves deeply important, merely that the structure of expressed reasons is an under-studied and meaningful dimension of political behavior as well.

In pursuit of these joint goals, this paper presents a computational, text-based approach for inferring conceptual networks from text. Through the application of this method to free-response text taken from survey data, the paper demonstrates the potential for these network structures to generate behavioral insights. The method proposed here, described in Section 4, leverages the grammatical structure of text to infer the implicitly encoded connections between mentioned concepts. Furthermore, the method uses word embeddings (Mikolov et al., 2013) to identify words which represent the same concept.

This method is applied to data taken from the 2016 American National Election Studies (ANES). Described in detail in Section 3, the survey solicited multiple text-based responses from a total of 4,270 subjects. Respondents further reported several key personal characteristics, including party identification and big-five personality traits measured by the Ten-Item Personality Inventory (TIPI) (Gosling et al., 2003). These personality traits are known to be correlated with political ideology (Haidt, 2012; McCrae and Costa Jr, 1999) and therefore bolster external validation for the method while simultaneously providing behavioral insights of their own. These measures and their hypothesized correlations are discussed further in Section 3. Furthermore, since respondents provided text across multiple prompts, we can examine the ways in which people express themselves across multiple topics and compare that to how different people approach the same topic.

This paper finds that individuals do appear to structure their political expressions in individually distinctive ways which are indicative of a personal style beyond mere ideological talking points. This "reasoning fingerprint" holds the potential to provide new behavioral insights, especially in regards to deliberation and conversation quality. This suggests that computational approaches for inferring conceptual network structure may be a fruitful line of research for

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

public opinion and political behavior. If we hope to understand the deliberative system or design deliberative interventions, we must not only measure what people say but also how they say it.

2. Related work

Cognitive processes and linguistic expression are both known to be phenomena with network structures (Quillian, 1967; Shavelson, 1974; Toulmin, 1958; Walton, 1996). Studies of reasoning (Axelrod, 1976; Carley, 1993; Toulmin, 1958), arguing (Toulmin, 1958; Walton, 1996), remembering (Collins and Loftus, 1975; Quillian, 1967), and learning (Shaffer et al., 2009; Shavelson, 1974) all suggest that individuals express and interpret beliefs in interconnected, structured ways. People store and retrieve information not as isolated packets of information, but as complex networks of interconnected concepts. When speaking with others, we raise ideas that seem related to what they said; when thinking to ourselves, we move from idea to idea via their connections; and when assessing a complex issue, we weigh the pros and cons as well as their interconnections in order to arrive at a final judgment. Network interpretations of the cognitive organization of knowledge are bolstered by behavioral observation of arguments, deliberation, written texts, and self-reports that repeatedly suggest that individuals perceive their ideas to be connected to each other in complex networks of support or contradiction.

Furthermore, cognitive and linguistic processes are inexorably linked: the conceptual networks which cognitively store information (Collins and Loftus, 1975; Dorsey et al., 1999; Quillian, 1967) cannot be directly observed and must be inferred primarily through language (Shugars and González-Bailón, 2024). This inference process has generally proceeded from two directions: a psychological approach which begins with theories of cognition and attempts to recover these structures through experimentation, observed behavior, and collaborative knowledge-building; and a linguistic approach which seeks to explain semantic patterns, meanings, and grammars using network structure. These two strains of study often converge on similar types of models, though they reflect the varied disciplines targeting this shared problem. Additionally, work in moral philosophy has aimed to normatively assess individual conceptual network structure, leaving aside issues of measuring that structure. Finally, popular behavioral approaches focus exclusively on clusters of latent traits as drivers of behavior, neglecting any network structure. This paper builds upon all these literatures, seeking to develop and validate an integrated approach for understanding individual-level conceptual network structure which can bring new insight to behavioral understandings.

Psychological models argue that human memory search is made possible by storing information as a network in which concepts, represented as nodes, are connected by relational links to other conceptual nodes (Collins and Loftus, 1975; Quillian, 1967). In Quillian's (1967) theory of semantic memory, for example, a node provides a shallow understanding of a given concept and is represented by a single word or phrase. A "concept" more deeply considered, then, contains indefinitely large amounts of information and is properly expressed as the entire network accessible from a given concept node (Collins and Loftus, 1975). Such a knowledge structure allows a person to store a concept as a compressed object (node) while simultaneously allowing access to a richer understanding through the network structure (Quillian, 1967).

These psychological theories have been applied in a range of settings. Semantic network libraries such as BabelNet (Navigli and Ponzetto, 2012), ConceptNet (Speer and Havasi, 2012), and SNePS (Shapiro and Rapaport, 1987) rely on the core psychological intuition that a concept, encoded as a word, can be best described though its associated concepts, which themselves are encoded as words. Education scholars have similarly leveraged psychological theories to argue that knowledge itself has a network structure and that "learning" can therefore be considered as a process of developing the right knowledge structures. In other words, the skill of applying existing knowledge to new situations relies upon developing an understanding of how relevant information is interconnected (Dorsey et al., 1999; Hong and Page, 2004; Shaffer et al., 2009; Shavelson,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

1974). Social scientists have further argued that conceptual networks can be used to examine how individuals reason and make choices between alternatives (Axelrod, 1976; Carley, 1993). In weighing possible outcomes, a person evaluates connected concepts and consequences; exploring paths within their conceptual network in order to determine the optimal choice. Political deliberation provides a natural venue to extend such models, as participants may enter conversation with differing views and must therefore attempt to share structured knowledge before reaching a decision.

Notably, the exchange of knowledge is most frequently done through language; leading to a separate stream of work engaging the structure of language as a proxy for the structure of knowledge. Perhaps the most well developed such models trace their roots back to Aristotelian efforts to define the structure of argumentation (Toulmin, 1958). Such structures may be relatively simple: a major premise connected to a minor premise leads inevitably to a logical conclusion; or it may be significantly more complex, such as in the two dozen schemes described by Walton (1996) or the Context Free Grammar introduced by Mochales and Moens (2011). But while theorists have differed in the specifics of the models they put forth, their approaches all begin with implicit acceptance of the network structure of arguments: the soundness of a conclusion rests not only upon the ideas supporting it, but on the ways in which those ideas are connected. In other words, arguments fundamentally have a coherent structure expressed through linguistic structure and defined by evidence relationships (Cohen, 1987). The search for these structures has given rise to a rich body of research known as argument mining, in which supervised and semi-supervised computational methods automate the search for the sorts of argument structures articulated by Aristotle or Toulmin (Mochales and Moens, 2011). The conceptual networks inferred via these methods tend to be more structured and hierarchical than those inferred from open-ended psychological approaches, but the basic structure of nodes and edges representing ideas and their interconnections remains.

While psychological and linguistic approaches aim to infer and examine conceptual network structure, an important line of work in philosophy has developed normative theories regarding the properties of these networks. These theories rely primarily on principles of coherence, considering a moral position valid insofar as it is coherent with other views (Christen and Ott, 2013; Dorsey, 2006; Rawls, 1993). What constitutes "coherence," however, differs between philosophers, leading to differing topological interpretations. In Henry Sidgwick's influential version of utilitarianism, for instance, "the current moral rules" such as "do not lie" are used to generate most of our actual judgments (Sidgwick, 1907), leading to topologies in which some ideas serve as central gatekeepers. In particularist moral theories, by contrast, each moral judgment is only linked to others by loose and local analogies (Dancy, 1993), implying that no ideas should enjoy disproportionate centrality in a person's network of moral ideas. McNaughton and Rawling (2000) argue that this is the flaw of particularism, because some concepts really are "central" to morality. This suggests a hybrid approach in which core ideas are central but do not dominate the reasoning structure. These varied definitions of "coherence" share an understanding that consistency between individual pairs of beliefs is too low of a standard for judging the validity of a moral position. On the one hand, individual beliefs may be consistent but unrelated, while on the other hand, expecting all pairs of beliefs to be directly connected is too stringent a standard since moral views range over a wide variety of topics. Several scholars have therefore explicitly argued for whole network approaches to coherence. Thagard (1998) proposes a theory involving literal network relations, though he overlooks many of the relevant formal features of networks. Berker (2015) posits that an individual's beliefs should be modeled as a network to reveal its degree of coherence and begins to explore the variety of forms that a network of moral values can take.

Finally, while this work's focus on the expression of political reasoning runs parallel to Zaller's (1992) examination of survey responses, Zaller provides a helpful framework through which to interpret the reasoning and articulation process. Zaller (1992) argue that survey responses can be modeled as a process of constrained stochastic sampling: individuals receive information through

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

external signals, selectively accept information which conforms to prior beliefs, and then sample from those available beliefs to generate an ideal-point estimation of their preference on the fly. This process is stochastic and will result in a single individual giving varied responses over time, but it is also heavily constrained—a subject may exhibit variability in how extreme their stated preference is, but is unlikely to spontaneously flip from one end of the political spectrum to the other.

While Zaller does not consider the network structure of political reasoning in his work, it is a natural extension to consider a similar process in this space. Following Zaller, we imagine that people receive and selectively accept external information. We then argue that this accepted information is stored as a latent conceptual network and represents the ideas and connections one has at their disposal. When expressing reasoning, individuals sample from this latent network to determine the precise topics they raise as well as the interconnected ways in which they aim to justify their beliefs.

3. Data

This study relies on data from the 2016 ANES (ANES, 2019), with a nationally representative sample of 4,270 US adults. The data features a series of open-ended questions, including eight prompts designed to elicit political opinions. On average, subjects responded to four of the eight open-ended prompts. Responses were generally short, with an average of 9.4 words and a standard deviation of 13.9 words. However, several responses were quite lengthy, with the longest response consisting of 204 words.

Open-ended questions asked respondents what they liked or disliked about political parties and presidential candidates. These questions were paired to include one question about the Democratic Party and its candidate and one question about the Republican Party and its candidate. For example, one pair of questions asked "If there is something that would make you vote for the [Democratic/Republican] presidential candidate." In the face-to-face version, interviewers were instructed to follow up with "What is that?" and "Anything else?" until the respondent says no. On the web, respondents were given a free-response box in which to reply. Other free-response questions asked, "Is there something that would make you vote against the [Democratic/Republican] candidate"; "If there is something you like about the [Democratic/Republican] party"; and "If there is something you dislike about the [Democratic/Republican] party."

Most respondents provided at least one answer to these questions, resulting in 4,066 subjects with at least one text responses. Of those, 28.06% were from surveys administered face-to-face while 71.94% were administered over the web. The mode distribution is comparable to that of the full sample (i.e., 27.63% from face-to-face mode and 72.37% from web mode), suggesting that mode did not systematically influence respondents' likelihood of providing text answers. Additionally, the survey also included measurements of Big-5 personality traits and party identification, which are key variables of interest in this study.

Party identification was measured with two questions. Respondents were first asked to indicate their party affiliation and then asked to what extent they identified with their partisanship. This was then transformed into a scale, ranging from 0 to 6 in ANES 2016 ($M = 2.86$, $SD = 2.15$), with 0 indicating a left-leaning ideology (strong democrat) and 6 right-wing ideology (strong republican). Respondents also indicated their age and gender.

Personality traits were measured using the TIPI (Gosling et al., 2003), a 10-item instrument measuring the Big-Five personality dimensions. The Big-Five framework is a widely accepted and validated model for understanding human personality traits (John and Srivastava, 1999). Many studies have successfully applied this framework to examine how individual-level variations relate to people's political discussion (Gerber et al., 2011; Hibbing et al., 2011) and semantic

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

traits (Kacmár and Körtvélyessy, 2023; Salsabila and Setiawan, 2021). It captures five key dimensions of human personality on a spectrum, including openness, conscientiousness, extraversion, agreeableness, and neuroticism. The TIPI scale was developed to provide a brief assessment of the Big-Five personality when long instruments (e.g., TDA (Goldberg, 1992), BFI (John and Srivastava, 1999)) are not feasible, making it especially suitable to be embedded in large-scale surveys like ANES. In the 2016 ANES, respondents were asked how well the following words describe them on a seven-point scale (1 = "Extremely poorly," 4 = "Neither poor nor well," 7 = "Extremely well"). Each pair of items was then recoded to measure one of the five dimensions of personality: openness ($M = 5.00$, $SD = 1.14$), conscientiousness ($M = 5.70$, $SD = 1.14$), extraversion ($M = 4.22$, $SD = 1.38$), agreeableness ($M = 5.20$, $SD = 1.14$), neuroticism ($M = 3.03$, $SD = 1.26$).

Using this data, we infer the network structure of respondents' free-response text and examine the following research questions:

RQ1: To what extent does the structure of an individual's conceptual network correlate to personality or demographic traits?

RQ2: Do subjects tend to generate conceptual networks with consistent structure, or do the prompts tend to elicit similar structures across subjects?

In exploring these questions, its important to note that all our data comes from a single, large-scale survey. While we can examine between- and within-subject variation in the context of this survey, we do not know if individuals' reasoning structures will be persistent across other contexts. In other words, the networks observed here can be best understood as reflecting how subjects express their reasoning on a survey, but that may be different from how they express their reasoning is other settings–such as informal conversation. Nevertheless, we argue that expressions of reasoning are an important aspect of the democratic process (Axelrod, 1976; Mansbridge, 1999). To better understand these expressions, the methods developed here can and should be tested across a range of settings.

4. Methods

This paper presents a method for inferring the latent conceptual network structure of text. While the literature suggests cognitive reasoning and linguistic expression are both best modeled through network structure, two important theoretical questions must be considered when developing such a model. First, what precisely is being connected, and second, what is the nature of those connections? In other words, in the resulting network model, what do the nodes represent and what do the edges represent?

In this section, we theoretically motivate the node and edge representations in a conceptual network, and describe a method for inferring these constructs. We next describe the challenges of network measurement and present a number of tools to measure and compare inferred network structures.

4.1 Inferring concepts (nodes)

A conceptual network is intended to represent the interconnections between concepts, which in turn requires the operationalization of what constitutes a "concept." In his classic work on semantic memory, Quillian (1967) argues that a "concept" can be understood as a compressed object which contains indefinitely large amounts of information. As a cognitive process, concepts serve as a heuristic guide to the boundaries of a topic which would otherwise require an arbitrarily large amount of resources to describe precisely. In this sense a "concept" is a recursive knowledge structure in which a meta-concept is itself comprised of a network of subconcepts. In practical terms,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

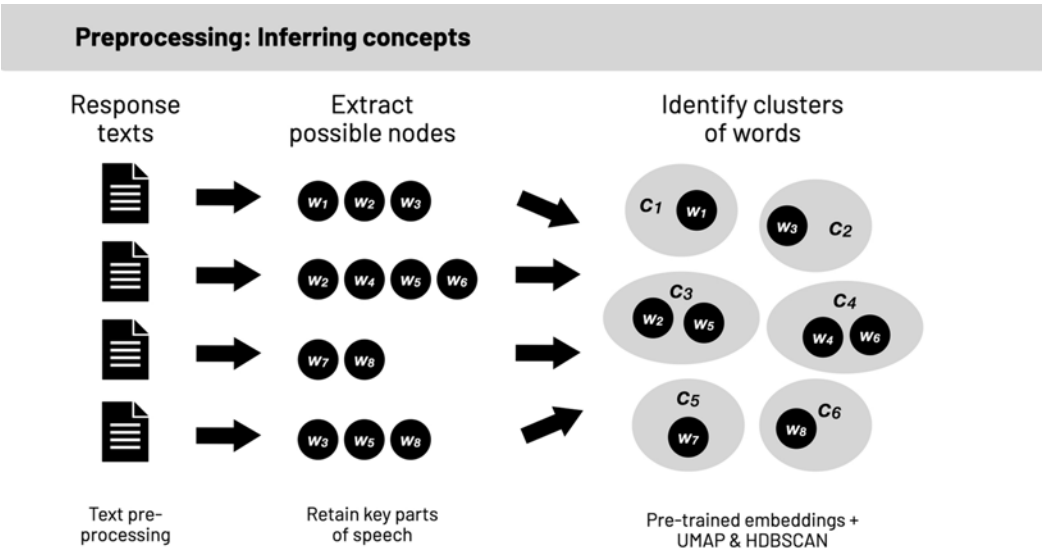


Figure 1. Pipeline for inferring concepts across the full corpus.

a "concept" is encoded as a word which carries heuristic meaning as to its network of associated subconcepts, which themselves are encoded as words. This is the core intuition behind classic semantic network libraries such as BabelNet (Navigli and Ponzetto, 2012), ConceptNet (Speer and Havasi, 2012), and SNePS (Shapiro and Rapaport, 1987).

Building on this intuition, we develop the following steps to infer concepts from a collection of texts. These steps are summarized in Figure 1. First, textual data is preprocessed using the BERT-based contextualSpellCheck package (Goel, 2021). This reduces the number of typos present in the text. Next, all words are converted to lower case and tagged with their part of speech using SpaCy (Honnibal et al., 2020). While grammatical structure is used to infer the *connections* between concepts—as described in Section 4.2—the *nodes* should only reflect the mentioned "concepts" themselves, not filler words or other grammatical elements. Therefore, only words which are likely to carry meaning are retained as possible nodes in the conceptual network. Specifically, we retain words that are any of the following parts of speech: adjective, adverb, noun, proper noun, or verb.

This provides an initial set of possible nodes, however each unique word does not necessarily reflect a unique concept. For example, the words "divide" and "split" may be reasonably understood to capture the same broad idea and therefore should be represented as a single node rather than as two distinct nodes. These words are not morphologically equivalent (e.g., do not share the same lemma or base form), but both evoke a shared concept of separation. In order to identify which words point to the same underlying concept, we therefore turn to word embeddings (Mikolov et al., 2013) as the next step in our data pipeline.

Trained on vast corpora of text, these models embed words in high-dimensional space representing the contexts in which those words appear. Words which are similar—or, more precisely, which occur in similar contexts—will then cluster together in this space. In line with previous work, similar words are assumed to point to the same concept (Firth, 1957). This allows subjects to have some reasonable lexical diversity without considering every unique word to indicate a unique concept. Specifically, once we have established a set of potential node words, each word is lemmatized and converted to an embedding vector. Since training word embeddings requires enormous quantities of data, we use a publicly available dataset of pretrained embeddings which have been found to be appropriate for most tasks (Spirling and Rodríguez, 2019). Specifically, we use embeddings pretrained on 100 billion words from the Google News corpus (Mikolov et al., 2013). This

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

dataset embeds words in 300-dimensional space by maximizing the average log probability:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c, j\neq 0}\log p(w_{t+j}|w_t) \tag{1}$$

For a sequence of training words $w_1, w_2, \ldots w_T$ and a context window of c. Once embedded, we use UMAP to reduce dimensionality of these vectors and cluster the embeddings using HDBSCAN.

While HDBSCAN does not require researchers to prespecify the number of k clusters, it does have several hyperparameters which inform algorithmic output. As described in Appendix A, we search this hyperparameter space, estimating models for 1440 possible hyperparameter combinations. For each resulting model, we calculate intra-cluster coherence using the cosine similarity of the full word embeddings. We penalize models for their least coherent cluster as well as for producing a small number of clusters. We take as our final model the one with the highest coherence score.

To demonstrate that these clusters are robust to random fluctuations in the model, we calculate the variation of information (Meila, 2003) between our final model and alternate iterations of the model. As shown in Appendix B, the word clusters identified through our data pipeline are relatively stable over multiple stochastic iterations. Finally, to ensure that the resulting clusters can reasonably be interpreted as representing cohesive "concepts" we manually inspect the lists of words that are clustered together by our selected model.

This process helps us go from extracting a list of unique words in our corpus to inferring a list of unique concepts. Those concepts are then taken to be the nodes in our conceptual networks. For example, one cluster includes the words: "divided," "splitting," "separation," and "breakup." Rather than be treated as independent nodes, each of these words is considered to represent the same concept and therefore the same node. If a respondent were to use multiple words from that list in their reply, they would be seen to be referencing the same concept (node) multiple times. On the other hand, if a respondent only used a single word from that list, there would be no functional effect on the inferred network structure, as they would only be referencing that specific concept (node) once. Additional examples of word clusters can be found in Appendix C. Across all text responses in our dataset, we find 2,149 total nodes, each representing a concept. Of those, 1,690 concepts are represented by a single words, while 459 concepts are represented by two or more words.

4.2 Inferring connections (edges)

The next challenge is determining both theoretically and operationally what constitutes the *connections* between inferred concepts. Building off work in linguistics, we argue that natural language is inherently structured—through grammar—to express the relations between concepts. We therefore leverage the grammatical structure of text in order to determine conceptual relations. Specifically, we determine the grammatical parse of a text by using SpaCy to identify each word's syntactic dependencies. This syntactic parse identifies the grammatical relations between words, for example, linking adjectives to the nouns they modify and subjects to their related objects. Importantly, these grammatical connections can be meaningfully interpreted—indeed, the very purpose of these grammatical rules is to serve as a tool to help humans encode and decode linguistic communication.

While the grammatical parse serves as the network's foundation, the final network structure should represent connections between *concepts*, not between individual words. While the process outlined in Section 4.1 provides a framework for translating word nodes to concept nodes, we next describe a process for modifying the parse tree of a given response in order to abstract up to connections between concepts. This process is summarized in Figure 2.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

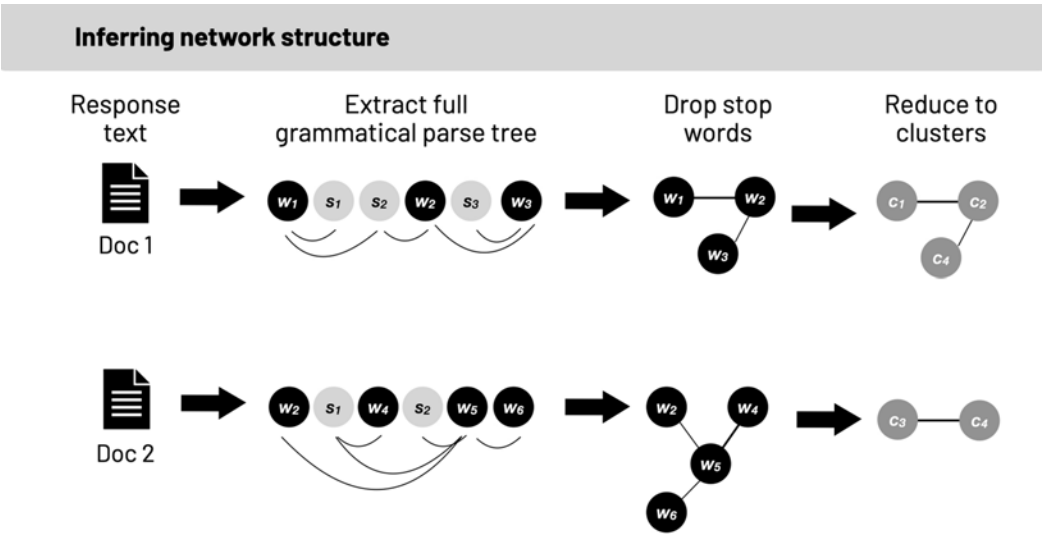


Figure 2. Pipeline for inferring network structure for a single response.

First, for each sentence in a subject's response, all child–parent relationships are determined. A word is considered to be a "child" of another word if it syntactically depends upon that word. Each word has at most 1 parent but can have any number of children. Next, any words that indicate negation—such as "not"—are flagged within the sentence. While these negation words are not retained as nodes themselves, they do reflect a *negative relationship* between concepts in the sentence. Ultimately, these negation words are therefore replaced with a negatively weighted edge between its grammatical parent and child.

Second, the subnetwork of retained words is extracted from the full parse tree of the sentence. This is done in an iterative fashion: for each retained word, we first check if its parent is also a word which should be retained based on its part of speech. If not, we iteratively check grandparents. An edge is recorded between this word and its first retained ancestor. If no ancestor is retained as a node, no connection is recorded. Each connection is recorded with a weight of $+1$ unless a negation word indicates a weight of -1.

Finally, the remaining words of the network are relabeled using the cluster labels determined through the process described in Section 4.1. This means that if a subject references the same concept multiple times—even if using different words—those words will now share the same node label. Nodes with the same label are then merged, reflecting the final "concepts" captured by that sentence. Alternatively, if a subject only references unique concepts, there will be no change to network structure. In this work, we are particularly focused on the resulting *structure* of reasoning, separate from the *content* of that reasoning. In this sense, the node labels are arbitrary. The key detail is identifying whether any of the included words reference the same concept and should therefore alter the final inferred structure. For each subject's response, this process is done at the sentence level and then aggregated over the full response text. This results in a weighted, signed network of conceptual interrelations—where heavier connections indicate that concepts were connected multiple times in a response. Illustrative response text along with inferred networks can be seen in Figure 3.

4.3 Assessing correlation between personality and structure

The first research question explores the degree to which personality and demographic characteristics correlate to the structure of conceptual networks inferred from a subject's free-response text.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

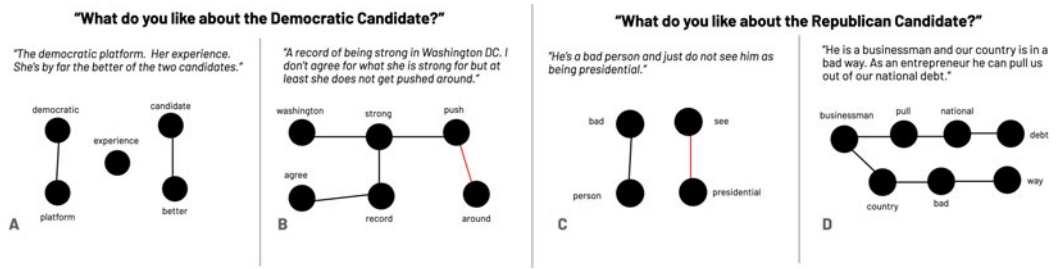


Figure 3. Example responses and networks. Black edges represent standard syntactic dependencies between concepts; red edges represent negation between concepts.

To examine this, we first calculate a number of theoretically motivated network statistics for each response and then estimate the correlation between each network statistic s and each personal trait p.

4.3.1 Network measures

In order to assess the different ways in which individuals' structure their responses, we describe the topology of inferred conceptual networks through seven network measures. These measures capture several key dimensions that holistically summarize network properties. Building off the moral philosophy literature, we engage measures which capture connectivity, complexity, and hierarchy—each representing different understandings of what good moral "coherence" should look like.

Connectivity serves as the baseline for coherence and is measured here through the percent of nodes which are in the giant component. Complexity is suggestive of the particularist view (Dancy, 1993) of coherence which advocates for richly interconnected and resilient networks. This approach is captured through the measures of average degree, clustering, and density. Finally, hierarchy reflects the view of utilitarianism as advanced by Sidgwick (1907) and covers measures of entropy, disassortativity, and standard deviation of degree. These measures are described in detail in Table 1.

While each measure captures a single feature of network structure, together they provide a holistic description of a network's local and global characteristics. For example, while the average degree—the number of connections nodes have on average—is a valuable piece of information, it alone does not provide detailed topological insight. From that single statistic, we cannot tell whether a network is heterogeneous (has nodes of differing degrees) or homogeneous (has nodes of similar degrees), whether it is connected or has multiple components, nor whether it is densely interconnected or sparse.

To provide a more intuitive sense of what these measures indicate, Figure 4 compares these statistics across three stylized example networks. Each network has a fixed number of nodes ($N = 16$) and edges ($E = 24$)—resulting in equal density ($d = 0.2$)—and are constructed to have equal average degree ($k = 3$). However, these networks display strikingly different topological properties, which are conveyed through our additional network statistics. In particular, we see that higher standard deviation and higher disassortivity are both indicative of heterogeneous, hub-and-spoke like structures. Entropy provides a weakly opposite indicator, with homogeneous networks having slightly higher entropy than heterogeneous networks. Given that entropy is calculated as $-\sum(p_k \times \log p_k)$, the minimal effect of variations in degree distribution suggests that higher entropy is more likely to be indicative of a higher node count. The final two network measures, giant component percent and clustering, each provide unique topological insight not captured by the other network measures. Specifically, the giant component percent indicates

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

Table 1. Measures of network structure

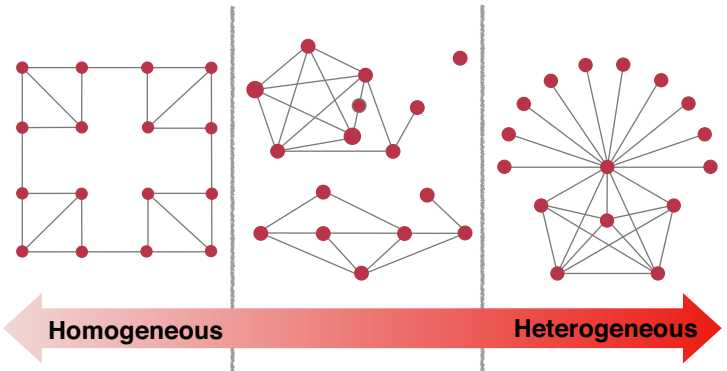
Connectivity (Baseline)	
Giant component percent	The percent of nodes in the largest component of the network, N_G/N. This measure indicates how cohesive the network is. A value of 1 indicates the network has a single component (e.g., a path exists between any two nodes), while lower values indicate that the network has multiple, disconnected components.
Complexity (Dancy, 1993)	
Average degree (k avg)	The average degree across all nodes in the network. Higher values indicate that nodes have more connections on average.
Clustering	A measure of how locally connected a network is. High values indicate triadic closure (Saramäki et al., 2007), while low values indicate a network that is locally tree-like.
Density	The ratio of existing edges to the total possible edges, $2E/(N(N-1))$. This is a measures the overall interconnectivity of a network with a value of 1 indicating that every idea is connected to every other idea and a value of 0 indicated that no concepts (nodes) are connected.
Hierarchy (Sidgwick, 1907)	
Entropy	Estimates the amount of information contained in the network's normalized degree distribution (p_k) (Shannon, 1948). This measure is dependent on both the length of the distribution (eg, N) and the heterogeneity of the distribution. Measured as $-\sum(p_k \times \log(p_k))$.
Disassortativity	Measured as the inverse of the Pearson correlation coefficient, $-r$, disassortativity captures the degree homophilly of the network (Newman, 2003). Values range from -1 to 1, with a value of 1 indicating that high degree nodes tend to connect to low-degree nodes, as in a star network. Note that for this study, we use disassortativity in order to have the same dimension and valence as standard deviation.
Standard deviation of degree (k std)	The standard deviation of the network's degree distribution. Lower numbers indicate that nodes are more homogeneous in their degree whereas larger values indicate greater difference between the lowest and highest degree nodes in the network. For the purposes of this study, k_{std} is normalized against a hub-and-spoke network with the same number of nodes N.

whether a network is connected or fractured into multiple components, while clustering indicates the presence of triangles—for example, the tendency of nodes which share a neighbor to themselves be connected.

It should also be noted that these network measures differ in how robust they are to noise. Statistics such as average degree, standard deviation of degree, and density are among the more robust measures, and will not change significantly with the random addition or removal of edges. Giant component is perhaps the least robust measure, as the random removal of a single edge could result in an isolated node and thus prevent a network from being complete connected.

Given these seven measures, we can then compare structural proprieties across networks, determining which networks are topologically similar and which are divergent. Furthermore, by examining the full set of metric-level comparisons, we can gain insight into the drivers of topological similarity or difference. Specifically, we are interested here in the degree to which each network statistic correlates, or not, to a respondent's personal characteristics.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046



density	0.2	0.2	0.2
k avg	3.0	3.0	3.0
clustering	0.2	0.4	0.3
giant component	1.0	0.4	1.0
entropy	0.0	1.5	1.0
disassortativity	-1.0	-0.1	0.7
k std	0.0	3.0	3.5

Figure 4. Comparison of network statistics across three stylized example networks. Each network has $N = 16$ and $E = 24$.

4.3.2 Estimating correlation

To estimate the correlation between each network statistic s and each personal trait p, we fit a series of multilevel models of the form:

$$s = \beta_0 + \beta_1 p + \beta_2 w + \alpha_q + \epsilon \tag{2}$$

Where s is the measured network statistic, p is a given personal trait, w controls for the number of words in a response, and α_q is a question-level random effect. For all models, we cluster standard errors at the respondent level. Importantly, in this exploratory work, we do not seek to measure the *effect* of personal trait p on network statistic, s. Rather, we are primarily interested in the strength and direction of the *association* between these variables. We use a multilevel model, rather than a more common correlation metric (e.g., Pearson's), because this approach allows us to control for key factors, such as word count and question-level interdependence. To capture the correlation between variables, we therefore focus our analysis on the t-statistic resulting from each model. Calculated as $\frac{\beta}{SE}$, the t-statistic effectively captures the magnitude and valence of variable relationships in a standardized, unitless measure which controls for possible confounding factors. Full regression results, including estimated effect sizes, can be found in Appendix D.

While our primary model *controls for* word count, one might reasonably ask what these network measures get us beyond what we could infer by simply *using* word count. Many network statistics are sensitive to the number of nodes and edges included, and those metrics, in turn, are going to be sensitive to the overall length of input text. Since personal traits may themselves be correlated to the verbosity of a respondent's answer—for example, subjects with higher education or political interest may systematically provide longer answers—it is worth considering a simpler model which uses only word count and does not rely on network inference at all. Therefore, we also consider the baseline model:

$$w = \beta_0 + \beta_1 p + \alpha_q + \epsilon \tag{3}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

In this model, we directly examine the correlations between the word count (w) of a response and each personal trait p. We again include question-level random effects (α_q) and cluster standard errors at the respondent level. Comparing this baseline model to our primary model shows whether network-based measures of reasoning structure provide information beyond what word count alone can offer.

While it is challenging to have *a priori* assumptions about what specific patterns might emerge in such a novel investigation, we expect to find meaningful correlations between network statistics and personality traits. Specifically, we hypothesize that subjects with higher education levels, political interest, and political participation will tend to generate more complex networks and may also tend to generate networks exhibiting more hierarchy—reflecting the additional thought and reasoning they have put into their positions. Subjects who are more open to new experiences, as well as those who are more extroverted, agreeable, and conscientious may similarly be more detailed in their responses, leading to more interconnected responses.

4.4 Assessing reasoning fingerprint

The second research question compares conceptual network structure within and between respondents. Network comparison is most frequently defined between networks on the same set of nodes. In such cases, we can understand the "difference" between the networks in terms of changes to how those nodes are connected. Conceptualizing comparison is more complex in this case because respondents very frequently do not bring up the same concepts—meaning that the networks inferred from their answers rarely share the same nodes. In this paper, we are primarily interested in examining variation in network *structure*, independent of variations in network *content*. In other words, from a conceptual standpoint, two networks should be similar if they exhibit similar topological properties, even if the specific concepts (nodes) are different. We therefore rely on portrait divergence (Bagrow and Bollt, 2019), which focuses on comparing the structural properties of arbitrary networks.

This approach defines a graph's portrait as an asymmetric matrix, B, in which the B_{kl} entry captures the number of nodes k which have path length l. This portrait contains both high- and low-dimensional information about the structure of the network. The network's degree distribution, for example, is captured by the first row ($B_{1,k}$), while the shortest path distribution is encoded as $\frac{1}{2}\sum_{k=0}^{N} kB_{l,k}$. This matrix is normalized to the row-wise cumulative distribution of B, and the similarity between two networks is calculated as the Kolmogorov–Smirnov test statistic K_l, for example, the maximum distance between the two matrices. A distance of 0 indicates networks which have identical topological properties, while a larger distance indicates networks which have less structural similarity.

For the 3,845 subjects who responded to at least two free-response questions, we then conduct two sets of network comparisons. Subjects' "self-similarity" is calculated by averaging the pair-wise portrait divergence between each network inferred from their free responses answers. The resulting measure captures how similar, on average, the structure of a subject's own responses tend to be across question prompts. We then compare this to the structures produced by different subjects. Specifically, for each subject, we first identify the subset of respondents who share their party ID. Since we might reasonably expect subjects' responses to these political questions to be ideologically driven, this creates a reasonably similar set of comparison subjects. Then, for each subjects' response, we calculate the portrait divergence between the subject's own conceptual network and the network inferred from all in-partisans who answered the same question. This presents a relatively high bar for comparison—we are looking specifically at the structure of responses from subjects who share a party ID and who are responding to the same prompt.

This approach gets at two potential drivers of conceptual network structure. If these reasoning structures are driven by personal style and characteristics, we would expect people to generate

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

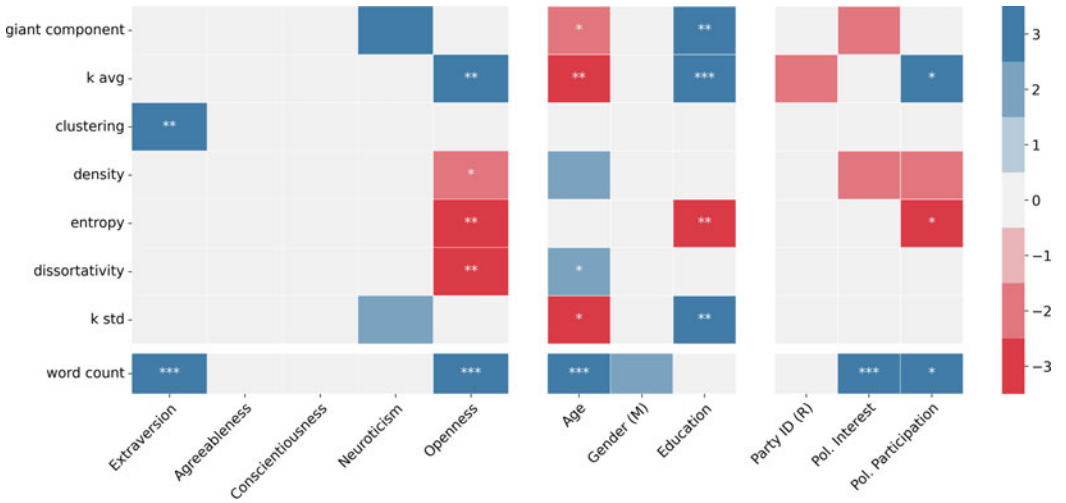


Figure 5. Correlations between network statistics and latent personality and demographic measures. Stars indicate significance after Benjamini–Hochberg FDR correction, *** ($p < 0.001$), ** ($p < 0.01$), * ($p < 0.05$).

similar network structures regardless of the specific question they are responding to. On the other hand, if different questions tend to inspire different types of reasoning, then we would expect networks inferred from the same question—across different subjects—to be more similar.

5. Results

5.1 Personality and reasoning

First, we examine the correlation between personal traits and reasoning structure. The results are shown in Figure 5. We display all correlations which our models find significant, and annotate the significance after using the Benjamini–Hochberg method to control the False Discovery Rate. Network measures are ordered following the theoretical expectations outlined in Table 1, examining variation by connectivity, complexity, and hierarchy. We draw on the examples in Figure 3 to illustrate how these typologies might manifest. While all network-measure models control for word count, the bottom row reflects our baseline model which only compares personal traits to word count. We find striking correlations between many personal characteristics and network measures, beyond what we would find by using word count alone. The findings of this exploratory work emphasize the potential value of network approaches and suggest possible pathways for future examination.

Subjects who have high levels of neuroticism and those who are more educated appear more likely to produce connected networks in which most concepts belong to the giant component (as in Figure 3, panels B and D). This suggests that these respondents may see—or feel the need to articulate—connections between the ideas they raise. Interestingly, subjects who are older or who have higher political interest are more likely to produce networks that are fragmented—for example, in which many of the concepts they bring up are not explicitly connected (as in Figure 3, panels A and C). This may reflect a gap between private reasoning and public expression. That is, subjects who produce fragmented networks may still see their ideas as connected, but they may take those connections for granted and not feel the need to spell them out in their expressions. Alternatively, this could be an artifact of the specific questions used in this dataset: these respondents may simply be more likely to generate a list of unconnected ideas when asked to share their political likes and dislikes.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

We further see notable variation with our measures of network complexity (k avg, clustering, and density), inspired by the particularist moral theories of Dancy (1993). These measures rarely have consistent patterns across subjects' demographic and personality traits. That is, there is no type of respondent who consistently produces "complex" reasoning networks as indicated by these measures. Interestingly, average degree and density appear to be opposite indicators, even after controlling for word count. The control is important here because longer responses are likely to generate larger networks, which could increase average degree while decreasing density. Indeed, we see in our baseline model that some traits which are significant predictors of word count (i.e., openness, political participation, age) are also associated with notable patterns in average degree and density. But the pattern of opposite indicators persists even after controlling for word count in the primary model, meaning that this association may be more than an outcome of response length. This also highlights how network models can provide insight beyond a word-count-only model: it is not just that these traits are associated with longer responses, rather these traits may be associated distinctive topological properties.

Specifically, subjects who have higher levels of openness and political participation tend to produce sparse networks with high average degree (as in Figure 3, panel D), while those who are older tend to produce dense networks with low average degree (Figure 3, panel C). Respondents with high education levels also tend to produce networks with higher average degree, but this characteristic does not have a significant correlation with network density. At the same time, extroverted respondents are the only ones who consistently produce highly clustered networks.

This suggests that different types of subjects may express the complexity of their reasoning in different ways. No respondents produce dense, highly connected networks in which most ideas are connected to most other ideas. Older respondents tend to produce networks in which concepts generally have few connections. Combined with our measure of connectivity (giant component), this suggests these respondents tend to generate collections of isolated dyads or open triads. In practical terms, this reinforces the idea that older subjects may be more likely to respond with list-like answers (Figure 3, panels A and C). Conversely, respondents with high openness and political participation tend to have central ideas loosely connected to additional concepts which are not interconnected themselves (Figure 3, panels B and D). Interestingly, as we will see in examining our final set of measures, these networks do not tend to be hierarchical, suggesting these respondents may express reasoning in line with particularist views.

Our measures of hierarchy (entropy, disassortativity, standard deviation of degree) reflect Sidgwick's (1907) utilitarian view that some concepts should be central, with all other ideas connecting in a hub-and-spoke manner (Figure 3, panel B). Here, we again see that no group of subjects consistently produce strongly hierarchal networks. In fact, respondents who are open to new ideas and experiences tend to produce more homogeneous networks. The negative correlation with entropy suggests that these networks have minimal degree variation while the negative correlation with disassortativity further suggests concepts tend to connect to other concepts which share a similar number of connections. Importantly, these networks still have low density—suggesting they are not fully homogeneous and some concepts may still enjoy more central roles. This is consistent with the line-like structure of Figure 3, panel D. There is some evidence for a similar pattern among subjects with high political participation. These respondents similarly tend to produce low-entropy but minimally dense networks with high average degree.

Users with high education levels appear to produce the most hierarchal networks, with high variation in the number of connections each concept has (e.g., high standard deviation of degree). Coupled with the fact these networks also tend to have a high average degree, this suggests hierarchal networks in which some concepts are highly connected (hubs) while others have few connections (spokes), as in Figure 3, panel B. Interestingly, these networks also tend to exhibit low entropy, suggesting consistent patterns in connectivity across nodes. Finally, as noted above, older subjects tend to produce disconnected networks, possibly indicating answers in the form of lists of

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

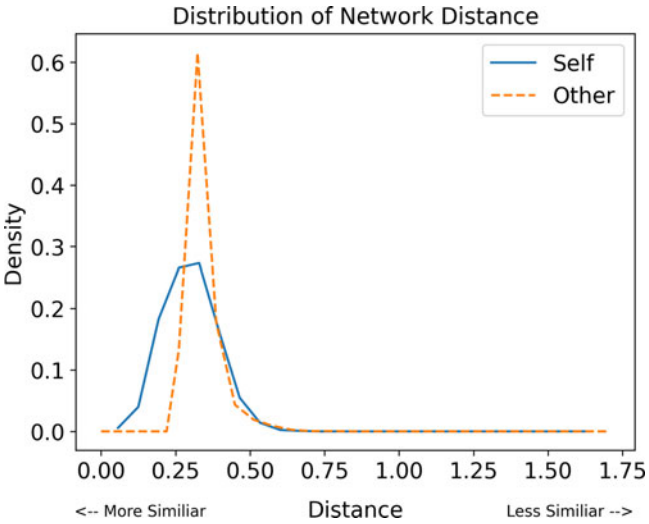


Figure 6. Distribution of network distance. "Self" captures average distance between a single subject's responses, while "Other" captures average distance with other respondents. Comparisons are limited to subjects with the same party ID responding to the same question.

unconnected concepts. The combination of high disassortativity and low standard deviation reinforce this finding, suggesting that there is little variation in how many connections concepts have, but that concepts tend to connect with dissimilar others. This is consistent with an open-triad pattern in which a central node has $k = 2$ and is connect to two dissimilar nodes of $k = 1$.

Interestingly, the network structures produced by respondents appears to be largely independent of party ID. Republicans may be slightly more likely to produce networks with lower average degree, but this finding is not significant after controlling the False Discovery Rate. Overall, this could be a promising finding for democratic outcomes—subjects do not appear to structure their reasoning based on partisanship, but rather in ways associated with their personalities, age, education, and interests.

All these findings are robust to variation in response length and could not be found in a model using word count alone. This suggests that there is value in advancing these network approaches to conceptual reasoning—only network measures can capture *how* people structure their responses, separate from the verbosity of those responses. While this analysis suggests meaningful correlations between personality and expressed structure, its important to note that it cannot disambiguate between possible effects. These same personality traits are believed to lead to ideological positions (Haidt and Joseph, 2008), making it unclear whether variations in structure are indeed an indication of personality or more generally a reflection of a given position's ideological talking points.

5.2 Reasoning fingerprint

As a first step toward such disambiguation, we further inspect the reasoning fingerprint associated with each subject in our data. Figure 6 shows the distributions of portrait divergence (Bagrow and Bollt, 2019) between subjects' own responses (Self) and the responses of in-partisans responding to the same question (Other). A distance of 0 indicates networks that are topologically identical, while higher distances indicate networks that are less similar in structure.

As we can see, networks inferred from a single individual are significantly likely to be more similar than networks inferred from different individuals. Many subjects produce responses that

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

are similarly structured regardless of the prompting question, and these structures are more likely to be different than structures produced by other respondents answering the same question. In fact, nearly a quarter (23%) of subjects produced networks structures that were more self-similar than the minimum observed similarity across users ($D = 0.23$). A t-test shows that this difference is significant ($p \ll 0.01$).

While there may be other individual correlates, such as education, driving this result, the difference in these distributions suggests that the structure of subjects' responses are not merely driven by partisanship or the specific question prompt. Rather, the inferred network structure may be capturing something about individual style or preference. Put differently, the structure of reasoning appears to be an individual characteristic rather than merely a topical or ideological one.

6. Discussion

Arguments for conceptual networks have been made in a variety of fields for decades. Within the public opinion literature, such models have been seen as a crucial tool for understanding political behavior and making sense of the deliberative system (Axelrod, 1976; Campbell et al., 1960; Lane, 1962). While recent work has begun to map the rhetorical structures through which political elites organize political concepts (Green, 2024), the present paper extends this logic to the mass public, asking whether ordinary citizens exhibit structural variation in political reasoning and how this may be behaviorally meaningful.

We argue that citizens express themselves in individually distinctive ways and that measuring those expressions has the potential for behavioral insight along a number of dimensions. From the standpoint of deliberative democracy, the ways in which individuals express themselves and understand each other is likely to influence the ability of people to successfully deliberate and collaborate around matters of common concern. Furthermore, even if we do not believe that citizens are rational creatures capable of formulating their own views, studying individual variation in expressions of reasoning still holds the potential to illuminate popular reflection of elite messaging.

This paper demonstrates that network approaches hold the potential to generate meaningful behavioral insight and a calls for a revival of these techniques for understanding political opinions. By demonstrating that conceptual network structures can be computationally inferred from text and meaningfully reflect individuals' personal traits, our findings underscore the value of revisiting questions around individual expressions of political reasoning. Due to the relative lack of modern work in this area, this paper is largely exploratory and demonstrates the value of considering the network structure of expressed opinion in the face of conflicting behavioral priors.

Specifically, this paper has presented a text-based method for inferring conceptual network structure. This method utilizes grammatical structure to capture the implicit connections between concepts and uses word embeddings (Mikolov et al., 2013) in order to identify which words point to the same concept. This approach captures the richness of natural language and the subtle ways in which individual expressions differ. Building off a long line of work in moral philosophy, inferred network structures are considered in terms of differing understandings of moral "coherence." A range of network measures capture the richness of these structures and are used to classify the argumentative quality of texts.

This method is applied to publicly available ANES data from 2016 (ANES, 2019), in which respondents provided between one and eight free-response answers as well as providing demographic and personality measures. We found several significant and substantive correlations between subject's personal characteristics and the network structures of their reasoning. Specifically, subjects with higher openness and who were more politically active were more likely to produce homogeneous networks in line with particularist moral philosophies (Dancy, 1993). Conversely, subjects who were more educated tended to produce more hierarchal networks, possibly reflective of utilitarian philosophies (Sidgwick, 1907). Older users tended to produce list-like

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

responses, resulting in disconnected, sparse networks. Furthermore, most subjects responded to multiple free-response questions covering their political likes and dislikes. This allowed us to compare conceptual network structure both within and between respondents. Specifically, examining whether subjects tended to produce similar networks across prompts or whether similar prompt tend to result in similar networks. We found that subjects tend to produce more self-similar network structures than topically or ideologically similar network structures. That is, a single subject's responses are likely to be more similar to each other than different subjects' responses to the same prompt.

Across all these exploratory findings, it is important to remember that we are only able to observe the structure of expressed political reasoning. While previous literature has suggested a connection between cognitive processes and linguistic expressions (Collins and Loftus, 1975; Quillian, 1967), our empirical data only reflects what respondents *say*, not necessarily what they *think*. This distinction is important, in part, because our analysis is restricted to a single dataset—a large-scale survey. The specific network structures we find here may therefore reflect how subjects tend to express themselves within that context and may not extend to other settings. For example, respondents who generate lists in response to survey questions may or may not actually structure their internal beliefs in this manner, and they may or may not choose to express themselves similarly under other circumstances, such as in an informal, face-to-face conversation. Nevertheless, we believe that citizens' expressions of political beliefs play an important role in shaping democratic outcomes such as perceptions of public opinion and everyday deliberation (Mansbridge, 1999). Therefore, while we hope to investigate other expressive contexts in future work, we also see inherent value in understanding the structure of those expressions.

This work ultimately aims to provide an approach which can provide insight into individual-level reasoning phenomena in order to better understand political behavior. The results of this study underscore that the network structure of subjects' expressions may be a meaningful tool for doing so. A classic line of public opinion research has long argued for the value of network approaches. This method values individuals' political expressions and takes seriously the challenge of expressing and discussing complex political thought. With modern computational techniques and data accessibility making network approaches tractable in this space, it is time to revive this classic line of research and to explore the ways in which network analysis can enhance our understanding of political behavior.

Acknowledgements. The authors would like to thank Nicholas Beauchamp and Peter Levine for their mentorship of this work, as well as the scholars of Visions in Methodology and Penn State's New Faces in Political Methodology for their invaluable feedback.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/nws.2026.10046.

Competing interests. None

Funding statement. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Data availability statement. The ANES data used in this study is publicly available at https://electionstudies.org. Replication code and analysis is available at https://github.com/sshugars/conceptual_networks

References

ANES (2019). ANES 2016 Time Series Study [dataset and documentation]. September 4, 2019 version. Available at https://www.electionstudies.org

Axelrod, R. (1976). *Structure of Decision: The Cognitive Maps of Political Elites*. Princeton University Press.

Bagrow, J. P., & Bollt, E. M. (2019). An information-theoretic, all-scales approach to comparing networks. *Applied Network Science*, *4*(1), 45.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

Berker, S. (2015). Coherentism via graphs. *Philosophical Issues*, *25*(1), 322–352.
Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). *The American Voter*. University of Chicago Press.
Carley, K. (1993). Coding choices for textual analysis: A comparison of content analysis and map analysis. *Sociological Methodology*, *23*, 75–126. doi: 10.2307/271007
Christen, M., & Ott, T. (2013). Quantified coherence of moral beliefs as predictive factor for moral agency. In Musschenga, B., & van Harskamp, A. (eds.), *What Makes Us Moral? On the Capacities and Conditions for Being Moral*. Springer, pp. 73–96.
Cohen, R. (1987). Analyzing the structure of argumentative discourse. *Computational Linguistics*, *13*(1–2), 11–24.
Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407.
Dancy, J. (1993). *Moral Reasons*. Cambridge: Blackwell.
Dorsey, D. (2006). A coherence theory of truth in ethics. *Philosophical Studies*, *127*(3), 493–523.
Dorsey, D. W., Campbell, G. E., Foster, L. L., & Miles, D. E. (1999). Assessing knowledge structures: Relations with experience and posttraining performance. *Human Performance*, *12*(1), 31–57.
Firth, J. R. (1957). *Studies in Linguistic Analysis*. Wiley-Blackwell.
Gerber, A. S., Huber, G. A., Doherty, D., & Dowling, C. M. (2011). The big five personality traits in the political arena. *Annual Review of Political Science*, *14*(1), 265–287.
Goel, R. (2021). Contextual Spell Check. doi: 10.5281/zenodo.4642379. Available at https://github.com/R1j1t/contextualSpellCheck
Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*(1), 26.
Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528.
Green, J. (2024). The rhetorical "What goes with what": Political Pundits and the discursive superstructure of ideology in US politics. *Public Opinion Quarterly*, *88*(4), 1166–1190.
Haidt, J. (2012). *The Righteous Mind: Why Good People are Divided by Politics and Religion*. Vintage.
Haidt, J., & Joseph, C. (2008). *The Moral Mind: How Five Sets of Innate Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules*, Vol. *3*. Oxford University Press. doi: 10.1093/acprof:oso/9780195332834.003.0019
Hibbing, M. V., Ritchie, M., & Anderson, M. R. (2011). Personality and political discussion. *Political Behavior*, *33*(4), 601–624.
Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. In Proceedings of the National Academy of Sciences of the United States of America, Vol. 101, pp. 16385–16389. doi: 10.1073/pnas.0403723101
Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). *spaCy: Industrial-strength Natural Language Processing in Python*. doi: 10.5281/zenodo.1212303.
John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research*, *2*(1999), 102–138.
Kacmár, P., & Körtvélyessy, L. (2023). Big-Five model of personality and word formation: Role of open-mindedness in semantic transparency and economy of expression. *Language and Cognition*, *15*(2), 217–244.
Lane, R. E. (1962). *Political Ideology: Why the American Common Man Believes What He Does*. Free Press of Glencoe.
Mansbridge, J. (1999). Everyday talk in the deliberative system. In Macedo, S. (ed.), *Deliberative Politics: Essays on Democracy and Disagreement*. Oxford University Press, pp. 1–211.
McCrae, R. R., & Costa, P. T.Jr (1999). A five-factor theory of personality. *Handbook of Personality: Theory and Research*, *2*(1999), 139–153.
McNaughton, D., & Rawling, P. (2000). Unprincipled ethics. *Hooker and Little*, *2000*, 256–275.
Meila, M. (2003). Comparing clusterings by the variation of information. In Schölkopf, B., & Warmuth, M. K. (eds.), *Learning Theory and Kernel Machines*. Berlin, Heidelberg: Springer, pp. 173–187.
Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
Mochales, R., & Moens, M. F. (2011). Argumentation mining. *Artificial Intelligence and Law*, *19*(1), 1–22.
Navigli, R., & Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, *193*, 217–250.
Newman, M. E. (2003). Mixing patterns in networks. *Physical Review E*, *67*(2), 026126.
Noel, H. (2012). The coalition merchants: The ideological roots of the civil rights realignment. *The Journal of Politics*, *74*(1), 156–173.
Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Systems Research and Behavioral Science*, *12*(5), 410–430.
Rawls, J. (1993). *Political Liberalism*. John Dewey essays in philosophy. Columbia University Press. Available at https://books.google.com/books?id=JHNPbt7pEeQC.
Salsabila, G. D., & Setiawan, E. B. (2021). Semantic approach for Big Five personality prediction on Twitter. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *5*(4), 680–687.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046

Saramäki, J., Kivelä, M., Onnela, J. P., Kaski, K., & Kertesz, J. (2007). Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E*, *75*(2), 027105.
Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A. A., & Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning. *International Journal of Learning and Media*, *1*(2), 33–53.
Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.
Shapiro, S. C., & Rapaport, W. J. (1987). SNePS considered as a fully intensional propositional semantic network. In Cercone, N., & McCalla, G. (eds.), *The Knowledge Frontier*. Springer, pp. 262–315.
Shavelson, R. J. (1974). Methods for examining representations of a subject-matter structure in a student's memory. *Journal of Research in Science Teaching*, *11*(3), 231–249.
Shugars, S., & González-Bailón, S. (2024). Semantic and cultural networks. In McLevey, J., Scott, J., & Carrington, P. J. (eds.), *The Sage Handbook of Social Network Analysis*. Sage Publications, pp. 202–211. doi: 10.4135/9781529614695.n14.
Sidgwick, H. (1907). *The Methods of Ethics*. Hackett Publishing.
Speer, R., & Havasi, C. (2012). Representing general relational knowledge in ConceptNet 5. In *LREC*. European Language Resources Association (ELRA), pp. 3679–3686.
Spirling, A., & Rodríguez, P. L. (2019). Word Embeddings: What works, what doesn't, and how to tell the difference for applied research. Working paper.
Thagard, P. (1998). Ethical coherence. *Philosophical Psychology*, *11*(4), 405–422.
Toulmin, S. E. (1958). *The Uses of Argument*. Cambridge University Press.
Walton, D. N. (1996). *Argumentation Schemes for Presumptive Reasoning*. L. Erlbaum Associates.
Zaller, J. R. (1992). *The Nature and Origins of Mass Opinion*. Cambridge University Press.

Cite this article: Shugars S. and Gu X. (2026). The structure of reasoning: Inferring conceptual networks from text. *Network Science* **14**, 1–20. https://doi.org/10.1017/nws.2026.10046

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 05:09:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/nws.2026.10046