Networks of Meaning:
Advancing Semantic Network Analysis for Organizational Scholarship

Derek Harmon
Michigan State University

Jamie Song
ESMT Berlin

ABSTRACT

Meaning is relational, residing not in individual concepts but in how they connect to one another. To measure meaning in text, organizational scholars increasingly use natural language processing (NLP) methods, but most methods capture how concepts relate in pairs, not how those relations are organized into a larger structure. Semantic network analysis (SemNA), an established method in cognitive science and communication, recovers this structure by representing text as a network of concepts and ties. We develop SemNA into a systematic framework that lets organizational scholars formulate and test the structural claims about meaning their theories rest on but that existing methods leave beyond reach. It does so through two innovations. First, it distinguishes between explicit usage networks (built from words actors observably use together) and latent meaning networks (inferred from distributional similarity in embeddings). Second, it translates established network measures into the semantic domain. Applying the framework to implicit gender bias in corporate earnings calls, we show how both network types expose bias as enacted in communication and embedded in latent meaning, a layered structure existing NLP methods cannot recover. We close by discussing how SemNA yields testable hypotheses about the structure of meaning across organizational settings.

Acknowledgements: The authors used an AI assistant (Claude, by Anthropic — Opus-family models, via the Claude Code interface) for writing and editing assistance (revising text, improving clarity, and flagging inconsistencies) and as a software tool to write the replication-archive code and run analysis scripts. All conceptualization, research design, and original analyses were carried out by the authors without AI. All AI-assisted outputs were verified by the authors, who are solely responsible for the content.

INTRODUCTION

Meaning is central to organizational life, shaping how strategies are communicated, how cultures are sustained, and how markets categorize their members. Underlying organizational scholarship's enduring concern with meaning is a more basic premise about where meaning resides, one with deep philosophical roots. Saussure (1916) argued that a word acquires significance only through its connections to other words, and Wittgenstein (1953) held that a word's meaning lies in how it is used, taking shape from the words that surround it rather than from the word itself. What "innovation" means in a given setting thus depends on what it is connected to, signaling something different when bound up with "disruption" and "risk" than when paired with "efficiency" and "compliance." Meaning, in this view, is fundamentally relational — a property not of individual words but of the relationships among them.

To capture this meaning, organizational scholars have increasingly turned to text and to the natural language processing (NLP) methods used to analyze it (Grimmer et al., 2022). The most widely used, dictionary-based counts, classify individual words into predefined categories, capturing meaning one word at a time and setting the relations among words aside (Short et al., 2018). More recent methods take the relational nature of meaning more seriously, with topic models linking words by their co-occurrence across documents and embeddings by the contexts they share (Aceves & Evans, 2024; Hannigan et al., 2019). In both, meaning is represented one pair at a time, as an association between two words. Knowing how two concepts relate, however, is not the same as knowing how those relations are organized. A topic tells you which words belong together, but not how those words are structured relative to one another. A cluster in embedding space tells you which words are nearby, but not whether they form a tight, self-reinforcing group or span otherwise-disconnected domains.

This organization — the pattern of *relationships among relationships* — is exactly what a wide range of organizational constructs are built on. Creativity (Amabile, 1983) and optimal distinctiveness (Taeuscher et al., 2021) both rest on concepts that bridge otherwise disconnected domains. Leader vision (Carton et al., 2014) and framing (Cornelissen & Werner, 2014) both turn on how the concepts a communicator invokes hang together as a cohesive whole. And organizational culture (Goldberg, 2011) and institutional logics (Thornton et al., 2012) both presuppose a densely interconnected core of values or principles, with peripheral concepts linked to this core but not as densely to one another. What unites these examples is that the meaning they depend on cannot be located in any single pair of concepts, but only in how the larger web of associations is structured. This topological layer of meaning is fundamental to how meaning is constructed, communicated, and interpreted in organizational life.

Studying meaning at this level requires a method that preserves the full structure of relationships among concepts, not just properties of individual words or word pairs. Semantic network analysis (SemNA) offers exactly this. By representing textual data as a network — where concepts are nodes and their relationships are ties — SemNA makes the topological organization of meaning not only visible, but also measurable. The method has deep roots in cognitive science, where Quillian (1967) and Collins and Loftus (1975) developed semantic networks to model the structure of human memory, and in communications, where they have mapped belief systems and public discourse (Carley, 1993; Carley & Kaufer, 1993; Danowski, 1993). Yet despite this foundation in adjacent disciplines, SemNA has not yet been developed into a systematic framework for organizational scholarship (see Figure 1). Dictionaries, topic models, and word embeddings have each become mainstream tools for analyzing text, but no comparable framework yet exists for the topological properties that network representation preserves. Without such a framework, the structural claims about meaning that pervade organizational theorizing remain difficult to formulate and to examine directly.

[Insert Figure 1 here]

This paper provides that framework. By developing SemNA into a systematic approach for constructing and interpreting semantic networks, we give organizational scholars a way to formulate and examine the structural claims about meaning that current methods leave out of reach. The framework rests on two innovations. First, we formalize the distinction between two types of semantic networks: explicit usage networks, in which ties represent words that actors observably use together in communication, and latent meaning networks, in which ties represent implicit associations inferred from the tendency of words to appear in similar linguistic contexts. This distinction mirrors a divide already present across NLP methods, where topic models rely on observable co-occurrence while embeddings capture latent associations. This delineation gives scholars a principled basis for capturing meaning as a resource mobilized through communication or as a latent cognitive structure. Second, we translate a suite of established network measures into the semantic domain. Each operationalizes a structural property that scholars invoke but rarely measure, from the bridging of distant domains to the cohesion of a communicator's message and the core-periphery organization of a belief system. To make the framework immediately usable, we release `semna` as an open-source Python package with worked examples on public-domain corpora so scholars can apply it to their own questions.

We then demonstrate the theoretical value of this framework through an empirical illustration that measures implicit gender bias in corporate earnings calls. We apply both network types to show how the topological properties SemNA reveals extend beyond the dyadic associations that serve as its foundation, exposing gender bias as both enacted in communication and embedded in latent meaning. We then showcase the reach of our framework in the discussion, where we review a range of research conversations across management and offer example hypotheses the framework makes possible for a selection of them. Together, this framework gives organizational scholars a way to formulate and test hypotheses about the structural organization of meaning across a range of literatures and settings.

The paper is structured as follows. First, we provide a brief history of SemNA, tracing its roots in philosophy and cognitive science to contextualize its current state in organizational scholarship. Second, we introduce our conceptual framework, which integrates both innovations to expand how organizational scholars might construct and interpret semantic networks. Third, we present the empirical illustration of gender bias in corporate earnings calls. Finally, we discuss how SemNA might advance theoretical work across multiple literatures, followed by the method's limitations and future extensions.

A BRIEF HISTORY OF SEMANTIC NETWORK ANALYSIS

The Relational Foundations of Meaning

Long before semantic network analysis was developed, philosophers established the idea that meaning exists within systems of relationships rather than in isolated words. Early structuralists, most notably Saussure (1916), argued that words derive significance only through their relationships with other words. Pragmatists like Dewey (1932) and Peirce (1934) built on this idea, noting that meaning is embedded in chains of relationships rather than one-to-one correspondences. This relational view was expanded by Wittgenstein (1953, §43), who famously asserted that "the meaning of a word is its use in the language," emphasizing that meaning is not a fixed definition, but derived from the surrounding semantic context (Quine, 1960).

The relational view of meaning has since informed diverse analytical traditions across the social sciences. But it was Quillian's (1966, 1967) groundbreaking doctoral dissertation on semantic memory that first formalized this view as a network structure, translating philosophical intuition into a concrete, implementable representation. This work proposed that human memory could be represented using relational networks. Quillian envisioned a structure where concepts in an actor's memory (e.g., "canary" and "bird") are represented as nodes connected by different types of relational links (e.g., "is a"), creating what he called a "semantic network." This conceptual model portrayed semantic memory as an interconnected web where meaning is distributed across a network rather than contained within individual concepts. Crucially, Quillian's model was not merely representational — it generated testable predictions about how people store and retrieve meaning, opening the door to empirical validation.

Quillian and his colleagues developed ways to experimentally test these ideas to see if the way people retrieved information from memory aligned with predictions grounded in semantic network structures. Collins and Quillian (1969), for example, showed that the greater the semantic distance between concepts (e.g., concepts located at a different abstraction levels), the more time participants needed to verify the relationship. Collins and Loftus (1975) built on this insight to demonstrate that concepts, once activated, flowed through semantic networks. For instance, when the concept of "ambulance" is activated, it spreads to different clusters of related concepts, such as "siren" (e.g., a feature of ambulances), "emergency" (e.g., what ambulances respond to), or "hospital" (e.g., where ambulances go). Taken together, these early papers offered preliminary evidence that human cognition and knowledge could be represented as a network with varying degrees of associative strength between concepts.

Development and Applications of Semantic Network Analysis

Building on these foundations, the interdisciplinary work of Carley and Danowski in the 1980s and 1990s laid the groundwork for SemNA as it is understood today. Conceptually, these scholars expanded the application of semantic networks beyond individual semantic memory, demonstrating how these networks could map shared mental models within teams (Carley, 1997), collective cultural beliefs (Carley, 1994), and public discourse (Carley & Kaufer, 1993; Danowski, 1993). They also developed techniques for extracting words from text, determining proximity between words, and mapping these relationships in a network structure to aid visualization (Carley, 1993; Danowski, 1982). These innovations transformed SemNA from a conceptual model into an analytical approach that would be picked up by other domains.

One such domain is cognitive science, where SemNA has become a powerful tool for understanding mental processes (for reviews, see Christensen & Kenett, 2023; Siew et al., 2019).[1] Developmental psychologists show how the structure of semantic networks evolves across one's lifespan (Dubossarsky et al., 2017; Wulff et al., 2022), and clinical psychologists have used semantic networks to conceptualize mental disorders as networks of related symptoms (Borsboom & Cramer, 2013; McNally et al., 2015). Extending Quillian's work on semantic memory, cognitive psychologists have also developed metrics for quantifying the distance between semantic concepts (Kenett et al., 2017) and applied these techniques to studying verbal fluency (Lerner et al., 2009; Kim et al., 2019) and measuring cognitive abilities (Rastelli et al., 2020), finding, for example, that creative individuals possess richer and more flexible associative networks than less creative individuals (Kenett et al., 2014; Kenett & Faust, 2019).

Organizational scholars have only just begun to engage with semantic network thinking. Early studies drew on network-related concepts without constructing networks directly, using network textual analysis to map how architectural movements mobilize symbolic codes (Jones et al., 2012) or semantic network theory to explain category label familiarity (Zunino et al., 2019). Since then, several scholars have constructed explicit usage networks from word co-occurrence

[1] Methodological work on SemNA has been developed in adjacent fields, most notably Christensen and Kenett (2023) in psychology. Our paper differs in that it distinguishes between explicit usage and latent meaning networks as well as translates established network measures into interpretations specific to organizational research questions.

to examine how meaning is organized in cultural and institutional contexts, from the prominence of stylistic elements in fashion markets (Claes & Godart, 2025; Godart & Claes, 2017; Godart & Galunic, 2019) to the structure of institutional logics and professional discourse (Giorgi et al., 2019; Höllerer et al., 2020; Jancsary et al., 2017; Santana & Kim, 2026). In addition, one recent paper has also begun to construct semantic networks from word embeddings to measure which words function as brokers between distinct domains of meaning (Song & Harmon, 2026).

Collectively, these independent efforts reflect a growing recognition among organizational scholars that the network structure of meaning carries theoretical weight, illuminating aspects of meaning that other methodological approaches do not directly offer. Yet they have advanced largely in isolation, without a shared account of how network structure maps onto properties of meaning. As a result, an approach that has proven powerful in adjacent fields remains underdeveloped for organizational research, its potential limited by the absence of a shared framework. This paper provides that framework, giving organizational scholars a systematic way to construct semantic networks and interpret the topological organization of meaning they reveal.

SEMANTIC NETWORK ANALYSIS: A CONCEPTUAL FRAMEWORK

Our conceptual framework has two components: network construction, which formalizes the distinction between explicit usage and latent meaning semantic networks, and network interpretation, which translates established network measures into the semantic domain, giving each measure a distinct interpretation depending on the network type.

Network Construction

Semantic networks are constructed with nodes and ties, which together form the topological structure through which meaning can be represented. Importantly, what a semantic network captures — and how it should be interpreted — is a function of how it is constructed. As such, this section lays out the components of a semantic network, the methodological choices that shape their construction, and the resulting distinction between two types of networks.

Nodes represent semantic concepts, such as an entity, attribute, or theme expressed in a text (Carley, 1993). In practice, semantic concepts are typically operationalized as individual words or phrases. For example, Kennedy (2008) used firm names mentioned in the media as nodes, whereas Godart and Galunic (2019) used stylistic elements mentioned by fashion houses in their collections. The selection of which semantic concepts to denote as nodes is a decision that shapes the resulting network's structure and interpretation.

Ties represent conceptual relationships between semantic concepts. Conceptually, a tie is based on the proximity (or distance) between two different concepts (Carley, 1993). Prior work has almost exclusively operationalized this proximity through *co-occurrence* (e.g., Danowski, 1993), treating two concepts as proximal when they appear together within a defined window of text. For example, "innovation" and "growth" frequently co-occur in shareholder letters, marking them as proximal under co-occurrence. We extend this tradition by integrating a second operationalization into our framework, *distributional similarity* (Veremyev et al., 2019), drawing on word embedding methods from natural language processing. This approach treats two concepts as proximal when they appear in similar linguistic contexts across a corpus, even when they do not directly co-occur. For example, "innovation" and "uncertainty" rarely appear together in the same text but occupy similar linguistic positions (each discussed alongside terms such as "risk" and "learning"), marking them as proximal under distributional similarity.

These two operationalizations of proximity give rise to two distinct types of semantic networks, as summarized in Table 1. *Explicit usage* networks, built on co-occurrence, capture the observable connections actors forge when communicating, representing the enactment of meaning in discourse. *Latent meaning* networks, built on distributional similarity, capture the implicit associations embedded across a corpus, representing the sedimented landscape of associations that actors rarely make explicit but navigate on a daily basis. Next, we walk through how each network is constructed, the methodological decisions involved, and the key distinctions that emerge between them.

[Insert Table 1 here]

Explicit Usage Semantic Network

Because explicit usage networks rely on co-occurrences in text, they are well-suited for studying how meaning is communicated, contested, and negotiated through discourse. They allow researchers to map how concepts are publicly connected, how new categories emerge through shifting language, and how rhetorical structures shape interpretation.

Input Corpus. The input corpus for constructing an explicit usage semantic network consists only of the text the researcher wishes to analyze. It need not be large. What matters is that it be targeted and theoretically aligned with the research question. Kennedy (2008), for example, analyzed a corpus of press releases and media coverage to study the emergence of a new market category. Giorgi, Maoret, and Zajac (2019) used firm annual reports to examine concepts surrounding “safety” in the U.S. automotive industry, enabling them to trace the evolving cultural meanings of legal compliance. In both cases, the corpus was delimited to the texts where relevant meaning construction unfolded.

Capturing Nodes. Nodes represent the semantic concepts that populate the network. For explicit usage semantic networks, their selection once again depends on the researcher’s theoretical focus. Kennedy (2008) defined nodes as firm names mentioned in the media (e.g., “Sun Microsystems,” “Apollo Computer”), treating each as a cognitive object within the emerging workstation market. Godart and Galunic (2019), in contrast, represented stylistic elements (e.g., “color,” “fabric”) as nodes in a semantic network of high-fashion design, enabling them to trace how certain cultural elements gained prominence over time. The decision of what counts as a node thus reflects the level of analysis at which the researcher seeks to infer meaning.

Capturing Ties. Ties in an explicit usage network are determined by the co-occurrence of nodes within a defined textual window. The researcher must specify the context window within which two concepts are considered to co-occur (e.g., sentence, paragraph, or entire document). This methodological choice shapes the granularity of meaning captured. Godart and Claes (2017) used a narrow four-word window to the left and right of focal brand names like "Rolex" to identify semantic ties (e.g., "Rolex"–"luxury") that revealed how audiences associated brands with cultural attributes in the luxury watch market. Kennedy (2008) instead defined his window at the level of news stories, treating firms mentioned together in the same article as linked. Once the context window is determined, researchers must translate co-occurrences into network ties, either by treating co-occurrence frequency as a continuous measure of tie strength or by establishing a tie when the frequency of co-occurrences exceeds a particular threshold.

Latent Meaning Semantic Network

Latent meaning networks capture a different kind of relatedness. Where co-occurrence picks up concepts that authors directly link in text, distributional similarity picks up concepts that are used in similar contexts across the corpus, even when authors never link them directly. The intuition is that two words with similar distributional profiles can be substituted for one another in many sentences without disrupting meaning, indicating that they share an underlying conceptual role. This makes latent meaning networks well-suited for studying the implicit associations and cognitive schemas that surface discourse cannot reveal.

Input Corpus. Constructing a latent meaning semantic network requires a corpus large enough to capture the background linguistic regularities that shape word meanings. Such a corpus provides the basis for estimating (i.e., training) the underlying semantic structure of language. Researchers can then analyze the semantic structure of the entire training corpus, or they can apply the semantic structure from the training corpus to a smaller target corpus that reflects a specific organizational context of interest. For example, Song and Harmon (2026) used as a training corpus all the language used by entrepreneurs and their audiences on the Product Hunt platform to identify ambiguous words, but then used that to measure the ambiguity of entrepreneurs' pitches on the platform as their target corpus.

Capturing Nodes. As with explicit networks, nodes represent semantic concepts, but here they are embedded in a multidimensional vector space rather than a two-dimensional co-occurrence matrix. Each word's vector encodes its contextual usage across the training corpus. These vectors are typically learned through NLP models such as Word2Vec, GloVe, or BERT (Devlin et al., 2019; Mikolov, Sutskever, et al., 2013; Pennington et al., 2014). The vectors learned by these models then serve as the nodes of the latent meaning network (e.g., Veremyev et al., 2019). The advantage of this approach is that it allows the nodes of interest to emerge from the data rather than being predefined by the researcher.

Capturing Ties. In a latent meaning network, ties represent semantic similarity rather than co-mention. The most common metric to capture ties is cosine similarity between two word vectors, which measures the angle between them in multidimensional space. A high cosine similarity indicates that the two words occur in similar linguistic contexts and thus share meaning. Researchers can use this as a measure of tie strength (i.e., the higher the similarity, the stronger the tie) or create ties based on a cosine similarity threshold (e.g., > 0.5 on a 0–1 scale).[2] Regardless of approach, the multidimensionality of using cosine similarity for tie creation allows researchers to observe subtle gradations of meaning that explicit networks cannot capture.

[2] A threshold approach affects network density (Veremyev et al., 2019) and should be supplemented with sensitivity analyses of nearby cosine similarity thresholds.

Network Interpretation

The second component of our framework is network interpretation, since the value of a semantic network depends on the measures used to interpret it. Drawing on the established network analytic tradition (Borgatti & Halgin, 2011; Uzzi & Spiro, 2005; Wasserman & Faust, 1994), we translate a comprehensive set of measures into the semantic domain, giving each a distinct interpretation depending on whether the network is explicit or latent. Table 2 summarizes these measures, organized into dyadic, node, and network levels. Across these levels, measures vary in how much of the topological organization of meaning they capture. Some aggregate only dyadic information that other NLP methods can partially approximate, while others capture how those relationships are organized, either among a concept's neighbors or across the network as a whole. The more a measure captures this topological structure, the more distinctive its analytical leverage beyond what dictionaries, topic models, and word embeddings can offer.

[Insert Table 2 here]

Dyad-Level Measures

At the foundation of any semantic network is the dyad, the relationship between two concepts. The key dyadic measure is the *existence of a tie*, indicating whether two nodes are connected. In explicit usage networks, a tie represents discursive co-occurrence. The frequent pairing of "innovation" and "growth" in shareholder letters, for example, signals that these ideas are linguistically coupled in narratives of success. In latent meaning networks, a tie represents semantic similarity. Even when "innovation" and "uncertainty" rarely appear together, they may sit close in semantic space because both occur alongside terms such as "risk" and "learning," revealing a deeper cognitive linkage. Because dyadic measures capture only pairwise relationships, they overlap with what existing NLP methods already provide. Their value in SemNA is as the raw material from which the higher-order measures below are built.

Node-Level Measures

Building on the foundation of dyadic ties, node-level measures reveal how individual concepts are positioned within the topological organization of meaning. These measures vary in how much of that organization they capture and, in turn, how distinct they are from existing methods. We organize these measures into three categories of increasing topological capture: nodal prominence, structural position, and neighborhood structure.

Nodal Prominence. Nodal prominence measures capture how visible or influential a concept is based on the accumulation of its immediate ties. Because they aggregate only local ties, they overlap with existing NLP approaches. Their value in SemNA is interpretive clarity, providing node-level scores that are directly comparable across focal concepts.

Degree centrality counts how many ties a concept has, identifying nodes that function as semantic hubs (Freeman, 1978). *Eigenvector centrality* (Bonacich, 1987) extends this by capturing recursive influence based on the centrality of a concept's neighbors, identifying anchors of meaning rather than simply frequent co-occurring terms. Both measures capture how prominent a concept is in the meaning system. In explicit usage networks, high-degree concepts appear across many different contexts in discourse, while high-eigenvector concepts anchor the core vocabulary by connecting to other widely-used terms. In latent meaning networks, high-degree concepts relate broadly to other concepts in shared meaning space, while high-eigenvector concepts anchor the most cohesive regions of meaning by connecting to other broadly-related terms.

Structural Position. Structural position measures capture how a concept is situated within the global structure of the network, requiring knowledge of connections across the entire system rather than just immediate ties. They have no genuine non-network analogue and offer distinct analytical leverage. Embedding models identify which words are similar to a focal term but do not directly determine whether that term mediates connections between other concepts elsewhere in the system. Topic models can identify co-occurrence across themes but do not directly specify which concept is structurally responsible for linking them.

Shortest path length measures the minimum number of intermediary concepts required to connect two nodes. A short path means two concepts are tightly connected, while a long path means they are linked only through chains of intermediaries. In explicit usage networks, this captures how directly two ideas are coupled in discourse. For example, if "innovation" connects to "accountability" only through "governance," speakers may use each of them alongside "governance" but rarely invoke them together, so the two ideas meet only through that shared bridging concept. In latent meaning networks, it captures how directly two concepts are connected in semantic association. For example, if "chairwoman" connects to "strategy" only through "board," the two are low in direct similarity but share an intermediary.

Betweenness centrality measures how often a concept lies on the shortest paths between other pairs of concepts (Freeman, 1977). High-betweenness concepts connect regions of the meaning system that are otherwise only weakly linked or joined primarily through them. In explicit usage networks, these bridges integrate distinct discursive domains. For instance, a term such as "sustainability" may connect discussions of financial performance with those of social responsibility, serving as the concept through which these domains are linguistically integrated. In latent meaning networks, these bridges integrate distinct regions of semantic association. For example, a concept such as "risk" may have high betweenness because it lies on the shortest associative path between financial concepts on one side and operational concepts on the other.

Core-periphery structure identifies whether a concept occupies a densely interconnected core of the meaning system or a structurally dependent periphery (Borgatti & Everett, 2000). Core concepts are extensively connected to other core concepts, while peripheral concepts connect primarily to the core but not to one another. In explicit usage networks, core concepts such as “strategy” or “leadership” appear across a wide range of corporate discussions, while peripheral concepts appear only in specialized discussions. In latent meaning networks, a concept such as “chairwoman” may connect to the leadership core but remain peripheral if it is not embedded within the dense network of associations that define that core.

Neighborhood Structure. Neighborhood structure measures capture the internal organization of a concept’s immediate semantic environment rather than its position in the global network. They offer SemNA’s most distinctive analytical leverage by revealing how a concept’s immediate neighbors are connected to one another, a structure of *relationships among relationships* that is invisible at the dyad level.

Ego density measures the proportion of realized ties among a node’s immediate neighbors (Hanneman & Riddle, 2005). High ego density means a concept’s semantic context is tightly interconnected, whereas low ego density implies a more diffuse, contested, or multidimensional meaning environment. In explicit usage networks, high ego density indicates that a word tends to appear in a stable, well-defined topical cluster. For example, “creativity” may consistently co-occur with “learning,” “experimentation,” and “failure,” which also frequently co-occur with one another, forming a coherent discourse of innovation. In latent meaning networks, high ego density suggests that the underlying meanings associated with a word are closely related, reflecting a tightly bounded conceptual domain. For example, “compliance” may cluster tightly with “rules,” “regulations,” and “audit,” reflecting a regulatory cognitive schema.

Brokerage, often captured by the inverse of Burt’s (1992) constraint measure (e.g., Kwon et al., 2020; Quintane & Carnabuci, 2016; Zhang, Aven, & Kleinbaum, 2024), assesses how much a concept bridges otherwise disconnected neighbors in the ego network. While ego density counts how many ties exist among a concept’s neighbors, brokerage reflects how those ties are arranged, and whether the neighbors are non-redundant, connecting to one another primarily through the concept rather than directly. In explicit usage networks, high-brokerage concepts connect vocabularies that are not otherwise used together, such as when the word "sustainability" co-occurs with a financial vocabulary and with an ethical one that rarely co-occur with each other. In latent meaning networks, high-brokerage concepts are semantically ambiguous, used in distinct ways across contexts. For example, the term "value" may be implicitly associated with both financial performance and organizational ethics systems of thought.

Network-Level Measures

At the highest level of analysis, network-level measures characterize the overall topological organization of meaning across an entire corpus. Some describe system-wide properties that other NLP methods can partially approximate through aggregation, while others require knowledge of the full relational topology and pathways that connect different regions of meaning. We organize these measures into three categories of increasing topological capture: aggregate connectivity, network segmentation, and network reach.

Aggregate Connectivity. Aggregate connectivity captures the pattern of connections in the semantic system, including how densely concepts are linked and how that density is distributed across the network. These properties are partially approximable using non-network approaches (e.g., frequency distributions, cosine similarity). However, the network formulation differs in two important ways. First, it imposes a threshold that distinguishes meaningful structural ties from weak associations. Second, it measures structural dominance based on a concept's position in the topology rather than raw frequency.

Density, defined as the proportion of all possible ties that are present in the network, provides a baseline measure of system-wide semantic integration (Wasserman & Faust, 1994). High density means a meaning system is tightly interconnected, while low density suggests a more dispersed structure. In explicit usage networks, high density indicates that communicators frequently link ideas together. For example, a single firm's annual letters tend to be denser than a cross-industry corpus, reflecting firm-specific consistency in language use. In latent meaning networks, high density signals conceptual integration. For example, the latent meaning network of a single discipline's publications tends to be denser than one spanning multiple fields, reflecting tighter conceptual integration within disciplinary boundaries.

Centralization captures whether a few hub concepts dominate the meaning system or connectivity is dispersed across many nodes (Freeman, 1978). In explicit usage networks, high centralization indicates that discourse is organized around a small number of dominant framing concepts through which most other ideas are linguistically connected. A CEO's annual letters might exhibit high centralization around signature strategic terms like "innovation," while news coverage of the firm is likely to show a more dispersed framing. In latent meaning networks, high centralization indicates that the cognitive architecture of the meaning system is anchored by a few core concepts, with most associative access flowing through a small number of structurally privileged nodes. Abstract terms like "value" or "performance" in corporate discourse might function as central anchors around which other concepts orient in semantic space.

Network Segmentation. Network segmentation captures the extent to which a semantic system divides into distinct and internally coherent domains of meaning. Topic models also partition language into clusters, but capture only the thematic grouping of words rather than the structure of relationships between those topics. The network formulation adds a critical dimension by preserving community boundaries while simultaneously measuring the presence and strength of ties that connect them. This allows researchers to determine whether domains are structurally isolated or connected, a distinction that cannot be recovered from topic assignments.

Modularity assesses how the network divides into internally cohesive but externally distinct clusters (Newman, 2006). High modularity means a system has clearly separated regions of meaning, whereas low modularity means meaning flows continuously across the system. In explicit usage networks, modularity identifies discursive communities or sets of words that tend to co-occur. For example, a corpus of CEO speeches might separate into "innovation and growth" and "ethics and responsibility" clusters. In latent meaning networks, modularity uncovers implicit schemas that structure collective cognition. For example, annual reports might separate into "profit and efficiency" and "community and welfare" clusters where economic and social logics do not overlap.

Network Reach. Network reach captures how concepts are structurally connected across the semantic system through chains of intermediaries, reflecting the global structure of pathways linking different regions of meaning. These measures rely on the full topology of the network and, therefore, offer distinct analytical leverage.

Average path length is the network-level extension of the shortest path measure, aggregating distances across all pairs to reveal the global navigability of the meaning system. A short average path indicates a broadly integrated system, whereas a long one signals fragmentation into topical domains that are structurally distant. In explicit usage networks this reflects how few discursive intermediaries separate ideas, and in latent meaning networks how many associative steps separate concepts. Discourse from a single firm or discipline tends toward short paths, whereas a corpus pooling many industries or disciplines exhibits a much longer one.

Small-worldliness extends this logic, capturing whether a network exhibits strong local clustering and short global path lengths at once (Watts & Strogatz, 1998). In explicit usage networks, this pattern marks well-developed discourse, where specialized subtopics stay distinct yet link readily through bridging terms, as when "AI" links "data" and "strategy." In latent meaning networks, it indicates collective cognition that is both differentiated and integrated, as when “innovation” bridges clustered creativity and strategy concepts in semantic space.

Summary

The framework pairs two components for analyzing the topological organization of meaning. Network construction distinguishes explicit usage networks from latent meaning networks, while network interpretation translates established network measures into the semantic domain, giving each a distinct reading by network type. Together they let organizational scholars treat meaning as both a communicative resource and an underlying cognitive structure.

PRACTICAL APPLICATION

Empirical Setting

We demonstrate the utility of the SemNA framework in the context of earnings calls, examining how gender is structurally organized in executive communication. Our aim is to show how SemNA reveals topological properties of gendered meaning — patterns in how gendered concepts are positioned relative to one another in the broader system of meaning — that existing NLP methods do not directly capture. To illustrate the framework’s range, we apply both network types. The explicit usage network traces how executives invoke gendered concepts alongside issues of leadership and strategy. The latent meaning network reveals how gendered terms are associated with implicit domains of meaning that reflect the architecture of collective cognition. The demonstration is meant to show what the framework reveals about the structure of gendered meaning, not to offer a comprehensive account of gender bias in corporate discourse.

Sample

Our sample comprises earnings call transcripts from all U.S. public firms in 2024, the most recent full year available at the time of analysis, obtained through the Financial Modeling Prep (FMP) API. The corpus contains transcripts for 3,195 companies across all four quarters.

Each call consists of a prepared presentation, where executives deliver scripted remarks, and a Q&A session, where analysts' questions often prompt unscripted responses, providing substantial variation in linguistic style, topical content, and firm context. We followed standard preprocessing procedures as described in detail in Appendix A.

Network Construction

Both networks draw on the same corpus and preprocessing pipeline but differ in how nodes and ties are defined, reflecting their distinct theoretical purposes.

Constructing the Explicit Usage Network

The explicit usage network is built from a researcher-selected lexicon of gender-, role-, and leadership-related terms, organized into six theoretically grounded categories: gender identity terms, family roles, professional titles, leadership traits, leader competence, and leader actions. Ties are defined by sentence-level co-occurrence at a t-score threshold of 2.5, a conventional cutoff for retaining co-occurrences that are statistically distinguishable from chance, which in this corpus corresponds to the top five percent of all word pairs. The lexicon determines which concepts enter the network as nodes, but all structural relationships among those nodes are empirically derived from co-occurrence patterns in the corpus rather than imposed by the lexicon design. Appendix B provides full details on lexicon construction, threshold selection, and robustness checks.

Constructing the Latent Meaning Network

The latent meaning network derives both nodes and ties from the corpus rather than relying on a researcher-specified set of terms. Nodes are 100-dimensional Word2Vec vectors learned over the full vocabulary of the preprocessed corpus, with semantically similar terms placed closer together in the embedding space based on the distributional contexts in which they appear. Ties are defined by cosine similarity above a threshold of 0.55, a level that reflects a substantively meaningful degree of semantic overlap based on inspection of in-sample word pairs at and around this cutoff. Appendix C provides full details on the embedding model choice (Word2Vec rather than contextual models such as BERT), the dimensionality and minimum-frequency parameters, the similarity threshold, and robustness checks.

Network Interpretation

For each network, we first examine dyadic relationships, followed by a representative set of semantic network measures from Table 2 to show how SemNA reveals topological structures that extend beyond the dyad. To isolate the gendered structure of meaning, we compare terms that differ only in the gender they convey. In the explicit usage network, we compare the formal role titles "chairman" and "chairwoman" to examine how gendered leadership terms are used in observable discourse. We examine the same pair in the latent meaning network to assess whether these terms are also embedded in different patterns of semantic association as well as the gendered pronouns "he" and "she" to further uncover how male and female actors are implicitly represented. We then integrate insights from the two networks to show how gender bias emerges at two distinct levels of meaning.

Insights from the Explicit Usage Network: How Gender Is Communicated

At the dyadic level, "chairman" co-occurs with a broad range of concepts (e.g., strategy, leadership, board, team, chair, chief, executive), whereas "chairwoman" co-occurs with only board, officer, and chief. At this level, "chairman" appears more broadly integrated. But dyadic co-occurrence alone cannot reveal whether those associated concepts are contained within a single domain or span multiple distinct regions of discourse, nor whether a concept merely participates in discourse or actively structures it. Answering these questions requires moving beyond dyadic counts to a concept's position within the broader topology, which we examine through two increasingly structural lenses.

We begin with nodal prominence. Beyond simply co-occurring with more terms, "chairman" is anchored to other central concepts in the discourse, with an eigenvector centrality more than an order of magnitude greater than "chairwoman's" (0.014 vs. 0.001). Although "chairwoman" attaches to prominent governance terms such as board and chief, its few ties leave it borrowing from the core leadership vocabulary without becoming part of it. "Chairman," in contrast, is itself one of the anchors of that vocabulary. This gap persists even when word frequency is held constant (see Appendix D). The implication is that the male-gendered title is not only used more often but occupies a more influential position in the meaning system, whereas the female-gendered title remains adjacent to prominence without attaining it.

To explore this further, we then turn to the internal organization of each concept's neighborhood, captured by brokerage. Brokerage is anchored in a concept's ego network (i.e., the concepts it directly co-occurs with and whether those concepts are themselves tied to one another) and is high when those neighbors fall into otherwise separate clusters that the concept bridges. As shown in Figure 2, the ego network of "chairman" reveals four topical clusters that it bridges: (i) governance and formal authority (e.g., board, director, counsel), (ii) executive identity (e.g., president, chief, executive), (iii) strategic direction (e.g., strategy, vision, leadership), and (iv) operational and relational (e.g., team, trust, morale). This partition is structural rather than merely visual, with modularity (0.20) suggesting that the four clusters are separable communities. The fact that these communities are disconnected from one another but are linked through the concept of "chairman" is confirmed by its high brokerage score (0.93).[3]

[Insert Figure 2 here]

[3] Brokerage is computed from Burt's constraint on unweighted ties (brokerage = 1 − constraint). The t-score threshold used to construct the network already performs the tie-selection role of the edge weights, so the structural measure treats each retained tie as a present bridge and is not driven by the frequency-inflated t-scores of high-frequency terms. The asymmetry is preserved if weighted by t-score rather than computed on unweighted ties.

"Chairwoman," in contrast, exhibits a more restricted profile. Its ego network contains only three neighbors (i.e., board, officer, chief), all of which belong to a single governance domain and are tightly interconnected. There is thus no internal segmentation to detect, leaving its modularity at zero with a single undivided cluster. Because there are no separate domains for chairwoman to bridge, the concept has a lower brokerage score (0.64). As with eigenvector centrality, this structural asymmetry surrounding brokerage is not an artifact of the relative frequency of "chairman" and "chairwoman" usage (see Appendix D).

Read together, the two lenses tell a consistent and compounding story. In explicit communication, "chairman" is prominent among central terms and brokers multiple domains of leadership meaning. "Chairwoman" is neither, in that it is peripheral in prominence and structurally enclosed within a narrow governance subdomain. Such asymmetry is not readily visible at the dyadic level, and it emerges only when we examine how a concept is positioned within the network and how its neighbors connect to one another.

Insights from the Latent Meaning Network: How Gender Is Implicitly Structured

We begin once again at the dyadic level with pairwise similarities, then turn to the topological structure those similarities form. At the dyadic level, "chairman" is strongly associated with a range of executive and leadership terms in the Word2Vec embedding, including CEO (0.80), president (0.80), executive (0.80), chair (0.78), CFO (0.76), and chief (0.76). "Chairwoman," by contrast, appears too infrequently in the corpus to produce a stable latent neighborhood. Indeed, latent meaning networks require that a term's corpus usage be both frequent enough to produce a stable vector estimate and concentrated enough to anchor it in a coherent neighborhood. These conditions are not just methodological, but are also substantive in that the repeated, patterned usage that stabilizes a term's vector is also what sediments it into a settled, shared meaning. Put differently, a term that does not anchor may be one whose meaning has not yet been conventionalized rather than simply one the method cannot capture.

When we examine the topological structure of how “chairman’s” neighbors are organized relative to one another, we see a slightly lower brokerage score in the latent network (0.832) than in the explicit (0.93), suggesting that its neighbors are themselves more interconnected in latent meaning space. Adding ego density, which captures how tightly a concept’s neighbors connect to each other, confirms this pattern. “Chairman’s” ego density (0.821) indicates that its associated terms form a self-reinforcing cluster. As shown in Figure 3, president, CEO, CFO, COO, and executive are strongly interconnected, anchoring “chairman” in a coherent core of executive leadership in implicit cognition.

[Insert Figure 3 here]

However, because of the limitations of not being able to compare “chairman” and “chairwoman” directly, we next turn to the gendered pronouns “he” and “she,” which open a window into gendered meanings that role titles cannot offer. Two features make them valuable here. First, both appear frequently, allowing the direct male–female comparison the role-title pair could not support. Second, and more fundamentally, pronouns do not denote substantive concepts of their own; they are referential markers whose meaning lies entirely in the contexts that surround them. Because speakers do not select pronouns to make a point the way they select substantive nouns, the associations that accumulate around “he” and “she” are largely un-curated. This makes them especially revealing of the implicit gender schemas that deliberate discourse tends to conceal (e.g., Cao, Koning, & Nanda, 2024; Kozlowski et al., 2019).

To capture each gendered referent as a single concept, we consolidate the pronoun forms during construction of the latent semantic network, collapsing his, him, and himself into the male term “he” and her, hers, and herself into one female term “she.” At the dyadic level, the female pronoun's nearest terms combine family and professional concepts (e.g., wife, career, daughter, husband), whereas the male pronoun's nearest terms are professional and leadership-adjacent (e.g., personally, career, mentor, CEO, appointment). While these dyadic similarities tell us which concepts each pronoun is close to, they don't tell us how those associated concepts are arranged relative to one another in a broader structure of meaning.

The ego networks of each concept make this internal organization visible (Figure 4). The neighborhood of "he" is large but structurally loose, spanning professional and leadership terms (e.g., career, CEO, leadership), relational terms (e.g., colleague, team, friendship), and evaluative courtesy terms (e.g., grateful, privilege, invaluable). These terms are only loosely connected to one another, giving "he" a low ego density (0.243) and a high brokerage (0.936). The neighborhood of "she" is smaller but internally organized. Its family terms are not only each close to "she" but are close to one another, with husband, wife, and daughter forming a closed, mutually reinforcing cluster that reflects domestic identity. This interconnection among the neighbors leads to a higher ego density (0.321) and a lower brokerage (0.838). The structural asymmetry is substantive. Women are encoded within a cohesive, self-referencing gendered cluster organized around domestic roles, whereas men, though associated with a broader range of terms, are not encoded within any comparable gendered cluster, occupying a diffuse contextual register that coheres into no role-specific structure.[4]

[Insert Figure 4 here]

Taken together, the latent analysis surfaces an implicit cognitive pattern that no single pair of concepts could show. The cognitive register of leadership in corporate language is

[4] This asymmetry is not an artifact of the chosen similarity threshold. Across cosine cutoffs from 0.50 to 0.60 (see Appendix C), "she" consistently occupies a denser, less-brokered, family-centered neighborhood than "he," with the ego-density and brokerage directions holding throughout.

gendered. It accommodates chairman and "he," whose neighborhoods reach executive terms such as CEO, but not "she," whose anchored neighborhood contains no leadership terms but instead includes only non-leadership, domestic registers.

Gender Bias in Use and in Meaning

Each of these networks stands on its own. The explicit network shows how gendered terms are positioned in discourse, and the latent network shows how they are implicitly structured in cognition. Although our framework does not require using both, the two layers can be read together to reveal additional insights.

"Chairman" is the simpler case. In the explicit network, it brokers across the breadth of leadership discourse, whereas in the latent network, it anchors a single, tightly cohesive cluster of executive terms. The title is thus rhetorically versatile in how it is used but concentrated in what it means. The female case reveals a different pattern. Here, bias appears in both layers at once. In the explicit network, "chairwoman" is communicatively marginal, confined to a narrow governance corner. The latent network shows the meaning structure that accompanies this. Because the title is itself too infrequent to anchor a neighborhood, we trace the female register through "she." This pronoun binds to family and domestic terms while "he" reaches professional and leadership ones. Marginal usage (in the explicit usage network) and domestic meaning (in the latent meaning network) point the same way, which suggests the bias is systematic rather than incidental. By building both networks and applying the same measures to each, SemNA makes this layered architecture observable in ways prior NLP methods cannot.

Beyond demonstrating the framework, this analysis reframes how bias in language can be studied. SemNA locates bias not in word counts, sentiment, or isolated associations, but in the topology of meaning. This view raises new questions, such as whether a female leadership term fails to anchor a semantic neighborhood merely because it is rare or because its meaning is not yet conventionalized (Eagly & Karau, 2002; Kozlowski et al., 2019). While answering such questions is beyond the scope of this paper, our framework allows analysis of how bias operates in both communication and cognition and how these two levels of meaning relate.

Limitations of This Application

This illustration carries several scope conditions. First, the corpus is limited to U.S. public-firm earnings calls in 2024, offering depth in one discursive setting but no view into how meaning structures evolve over time or across different types of text. Second, we highlight findings for a small set of theoretically central gendered terms (chairman, chairwoman, he, she), with the full lexicon listed in Appendix B. Both extensions await future applications.

Third, we concentrated on node-based measures (Table 2) by design. Substantively, the gender bias we study is a property of how a particular term is positioned relative to other meanings, not of the overall network, so the natural unit of analysis is a focal concept and the neighborhood it organizes. Methodologically, whole-network measures such as density, centralization, and modularity are comparative by nature, so their leverage emerges in the cross-corpus and longitudinal comparisons we reserve for future work. The one segmentation measure we report, modularity, is computed on a concept's ego network to characterize its neighborhood rather than the corpus.

Fourth, both networks rely on researcher-specified parameters. The explicit network uses a lexicon and t-score threshold, the latent an embedding model and cosine threshold. We report robustness checks across alternative t-score thresholds (Appendix B), alternative cosine thresholds (Appendix C), and a frequency-matched bootstrap of the chairman/chairwoman comparison (Appendix D). These address common parameter-sensitivity concerns, but readers applying SemNA should expect to revisit these choices for their own corpus and question.

Data and Code Availability

To support adoption and replication, we release SemNA as an open-source Python package (`semna`, version 0.1.0, MIT license), provided for review through the restricted-access OSF archive (see Appendix E) and to be released publicly on PyPI and GitHub on publication. It gives organizational scholars a tested implementation for constructing explicit and latent semantic networks and applying the full set of SemNA measures, and it bundles three public-domain corpora (presidential inaugural addresses, FOMC statements, and SEC 8-K filings) and worked vignettes so they can adapt the method to their own questions.

A separate replication archive reproduces every value reported in this demonstration. Because the underlying earnings-call corpus is licensed from Financial Modeling Prep and cannot be publicly redistributed, the archive ships the trained model and derived network files as canonical artifacts, together with scripts that recompute the reported quantities, and makes these artifacts available to reviewers through a restricted-access archive. Full details, including the bundled corpora, the reproduced quantities, and the corpus retrieval protocol, appear in Appendix E.

GENERAL DISCUSSION

Meaning in organizational life resides not in individual concepts but in how they are connected to one another. This topological layer of meaning underlies a wide range of organizational constructs but has remained difficult to measure directly. This paper develops SemNA into a systematic framework for doing so, through two related advancements. The first concerns network construction, where we formalize the distinction between explicit usage and latent meaning networks, capturing meaning as both a resource mobilized through communication and a cognitive structure that shapes interpretation. The second concerns network interpretation, where we translate a comprehensive vocabulary of established network measures into the semantic domain. Together, these contributions give organizational scholars a systematic basis for analyzing the topological organization of meaning in text. In what follows, we first highlight how SemNA might advance theoretical work across multiple organizational conversations, then outline the method's key limitations and conclude with future directions.

Advancing Organizational Scholarship

We turn first to how the SemNA framework might enrich organizational research across a range of scholarly conversations. What unites these conversations is that many of their constructs turn not on individual concepts but on how meaning is topologically structured. Table 3 surveys these conversations across two domains — internal organizational activities and market and institutional activities — pairing each with research questions suited to an explicit usage or latent meaning semantic network. The sections below develop one illustrative example from each quadrant, each closing with an example hypothesis the framework makes possible, with the table serving as a broader map of where SemNA's leverage extends.

[Insert Table 3 here]

Internal Organizational Activities

Scholars studying internal organizational activities — leadership, sensemaking, teams, and culture — examine how meaning is constructed and shared within organizations. Across these literatures, meaning operates both as a resource actors mobilize through communication and as a cognitive structure that frames action. By distinguishing these networks and revealing the topological organization of each, SemNA lets researchers capture both.

Scholars can use explicit usage semantic networks to study how actors strategically mobilize language to influence others. Recent work on CEO and leader communication has used computational text analysis to measure rhetorical patterns, often through dictionary-based coding of vision-evocative imagery, regulatory focus, or temporal framing (Carton et al., 2014; Crilly, 2017; Gamache et al., 2015). These approaches identify which categories of language a leader uses but typically do not characterize how those linguistic elements are structurally connected to one another. SemNA enables this structural specification. By representing a leader's communication as an explicit usage network, researchers can apply brokerage to identify the concepts through which a leader joins otherwise-separate vocabularies. A vision organized around a high-brokerage term, such as "sustainability" or "inclusion," ties a financial-performance vocabulary to a social-responsibility one that would otherwise remain disconnected. This makes it possible to hypothesize that the brokerage of a leader's core concepts predicts how effectively the vision coordinates action across otherwise-separate functions, independent of how much vision-evocative language the leader uses, because a bridging concept gives members of different domains a shared reference that links their vocabularies.

Latent meaning semantic networks offer a different analytical lens, generating insight into the implicit structures of organizational meaning. A growing line of work on organizational culture has used computational text analysis to measure cultural fit and content from internal communications and public reviews (Goldberg et al., 2016; Corritore et al., 2020; Srivastava et al., 2018). These approaches typically operationalize culture through linguistic style matching across actors or through topic distributions across documents. SemNA enables a structural complement. By representing an organization's discourse as a latent meaning network, researchers can apply core-periphery structure to characterize how its meaning system is organized. A pronounced core-periphery structure indicates a culture built around a densely interconnected core of values, with peripheral concepts attached to that core but not to one another. The breadth of that core, and how tightly the periphery binds to it, becomes a structural signature of how unified or pluralistic the culture is. This makes it possible to hypothesize that two organizations with identical cultural content can function differently depending on how that content is structured, with a narrow, tightly bound core yielding stronger consensus but slower adaptation to novel demands than a broad, loosely bound one.

Taken together, SemNA contributes to these conversations by shifting the analytical focus from the inventory of words used to the topological organization of meaning. By creating structural representations of both communication and cognition, our framework allows scholars to empirically capture complex meaning systems and employ network measures to explain how those systems facilitate coordination, sustain culture, or shape interpretation inside organizations.

Market and Institutional Activities

Scholars studying market and institutional activities — categories, cultural entrepreneurship, and institutions — examine how markets are constructed, legitimacy is acquired, and fields evolve. Across these literatures, meaning operates as a strategic tool actors deploy to shape environments and as a sedimented reality that defines the boundaries of legitimate action. By distinguishing explicit usage from latent networks, SemNA lets researchers model this interplay of agency and structure.

Scholars can apply explicit usage networks to study how actors strategically mobilize language to reshape market boundaries. Research on cultural entrepreneurship and "optimal distinctiveness" has increasingly turned to computational text analysis to measure narrative distinctiveness in entrepreneurial pitches, crowdfunding campaigns, and platform listings (Harmon et al., 2023; Lounsbury & Glynn, 2001; Taeuscher et al., 2021). These approaches typically capture distinctiveness through topic-model distances or embedding-based similarity to category baselines (e.g., Vossen & Ihl, 2020). SemNA enables a structural specification of how this distinctiveness is achieved. By representing an entrepreneur's narrative as an explicit usage network, researchers can apply brokerage to identify the specific bridging concepts that link a dense cluster of familiar terms to a novel domain. An entrepreneurial pitch in a new technology category, for example, could be analyzed to identify the analogies and metaphors that broker between the novel technology and established concepts. This makes it possible to hypothesize that two ventures can be equally distinct from their category by scalar distance yet draw different audience support, with the one that ties its novel terms into a dense cluster of familiar terms gaining more than the one whose novelty stays structurally isolated, because bridging concepts make the unfamiliar legible in terms audiences already understand.

Latent meaning semantic networks offer a different analytical lens, revealing the underlying structures of a field. Research on market categories has increasingly used computational text analysis to characterize what makes categories cohere, who belongs to them, and how they evolve (Kovács & Hannan, 2015; Le Mens et al., 2023; Pontikes, 2022; see also Hannan, 2022 for a programmatic statement). These approaches typically operationalize category structure through embedding-based distances or topic-model groupings. SemNA enables a structural complement. By representing a field's discourse as a latent meaning network, researchers can apply modularity and shortest path length to characterize the topological organization of categories themselves. Modularity captures whether a market segments into distinct, internally coherent categorical cores, while shortest path length captures the structural distance an organization must traverse when spanning two categories. This makes it possible to hypothesize that the penalty an organization incurs for spanning two categories scales with their shortest path length in the field's latent network, so that spanning structurally distant categories is penalized more than adjacent ones even at equal feature overlap (Hsu, 2006; Hsu et al., 2009).

Taken together, SemNA provides a structural analysis of market and institutional activities that distinguishes between the strategic enactment of fields and their underlying architecture. This framework advances these conversations by moving beyond the measurement of field "states" (e.g., is this category legitimate?) to the measurement of field "topology" (e.g., how porous are its boundaries?). By creating structural representations of both strategic action and institutional constraint, SemNA allows scholars to capture the dynamic interplay of agency and structure, opening new avenues for research on how fields evolve, fracture, and settle.

Limitations of SemNA

No representation of text is perfect, and each approach — whether dictionaries, topic models, word embeddings, or SemNA — captures some aspects of meaning while inevitably flattening others (Aceves & Evans, 2024; Hannigan et al., 2019). The insights SemNA generates depend on how the network is constructed, how semantic information is simplified, and how well the method scales with corpus size. We outline limitations related to each.

A first limitation concerns the method's sensitivity to analytic design choices. Semantic networks do not naturally emerge from text but instead are built through a series of researcher decisions, such as how to define co-occurrences or semantic similarity, which thresholds to use when determining ties, how to delimit the corpus, and which embedding model or context window to adopt. Small changes in these parameters can sometimes meaningfully alter network structure. The choice of embedding method is particularly consequential, since Word2Vec, GloVe, fastText, and contextual models such as BERT each capture distributional regularities differently, and findings from a latent meaning network may not transport across embedding methods without re-validation. As a result, substantive findings may reflect modeling assumptions as much as underlying meaning. To address this, researchers should explicitly justify their design decisions and assess the robustness of results across alternative specifications (as illustrated in the Practical Application).

A second limitation is the loss of semantic information inherent in translating rich, context-dependent language into a simplified network of nodes and ties. SemNA necessarily compresses this complexity by reducing multi-dimensional meaning to a lower-dimensional network structure. SemNA should therefore be understood as an analytic abstraction rather than a literal reflection of linguistic nuance. Individual ties are best interpreted cautiously and in relation to broader network patterns, and researchers should consider supplementing SemNA with other methods to triangulate interpretation.

A third limitation involves scalability and density. As corpora grow larger, semantic networks tend to become dense because many words co-occur or exhibit high similarity in large text collections. Dense networks can become visually cluttered, computationally cumbersome to analyze, and theoretically uninformative as everything becomes connected. To mitigate these challenges, researchers should consider using stricter tie thresholds, pruning weak or trivial connections, or sampling and stratifying large corpora before constructing networks. Focusing on cluster-level or node-level patterns — rather than attempting to interpret the entire global structure — can also help preserve meaningful signal.

Future Extensions

Several extensions offer promising avenues for future work. One extension involves examining how meaning systems change over time and how they compare across contexts. Future work might model how the dynamics of meaning unfold through repeated communication, observing both when enactments become sedimented and reshape the underlying meaning structure (Green et al., 2009; Kennedy, 2008) and when existing latent structures constrain the meanings actors can legitimately draw upon (Glaser et al., 2016; McPherson & Sauder, 2013). This temporal orientation also surfaces moments when explicit discourse and latent meaning move in different directions, revealing sites of tension, drift, or semantic innovation. A complementary direction is to compare meaning systems across organizations, fields, or time periods, where network-level measures (Table 2) such as density, modularity, and centralization become especially informative.

A second extension centers on linking semantic networks with cognitive and social networks. Meaning is not only semantically embedded in language but also shaped by the social relationships through which these associations circulate. Integrating semantic structures with social structures would enable scholars to explore how interpretive resources are distributed across groups, how social position influences access to and use of particular meanings, and how collective sensemaking processes emerge from patterns of interaction.

A final extension involves combining semantic network analysis with other NLP methods. While latent meaning networks already draw on embedding models, researchers might incorporate additional tools such as topic modeling, dependency parsing, sentiment and emotion analysis, named entity recognition, and contextual embedding models. When mapped into a semantic network, these layers would reveal not only how ideas are connected but also how they are framed, evaluated, and distributed across actors.

Conclusion

Meaning is relational, and SemNA gives organizational scholars a systematic way to measure it by capturing not only which concepts populate a text but how they are organized into a larger structure. By distinguishing between explicit usage and latent meaning networks, and translating established network measures into the semantic domain, our framework allows scholars to capture meaning as both a communicative resource that actors actively mobilize and a cognitive structure that frames their action. SemNA moves the analysis of organizational text past frequency counts, topic loadings, and high-dimensional vectors toward the topological organization of meaning — a method that is both mathematically rigorous and theoretically interpretable. As textual data becomes increasingly central to organizational inquiry, we hope this framework inspires scholars to look past the mere presence of words to investigate the architecture of meaning that shapes organizational life.

Acknowledgements: The authors used an AI assistant (Claude, by Anthropic — Opus-family models, via the Claude Code interface) for writing and editing assistance (revising text, improving clarity, and flagging inconsistencies) and as a software tool to write the replication-archive code and run analysis scripts. All conceptualization, research design, and original analyses were carried out by the authors without AI. All AI-assisted outputs were verified by the authors, who are solely responsible for the content.

REFERENCES

Aceves, P., & Evans, J. A. 2024. Mobilizing conceptual spaces: How word embedding models can inform measurement and theory within Organization Science. ***Organization Science***, 35: 788–814.

Amabile, T. M. 1983. The social psychology of creativity: A componential conceptualization. ***Journal of Personality and Social Psychology***, 45: 357–376.

Bass, B. M. 1990. ***Bass & Stogdill's handbook of leadership: Theory, research, and managerial applications*** (3rd ed.). New York: Free Press.

Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. 2016. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. ***Advances in Neural Information Processing Systems***, 29: 4349–4357.

Bonacich, P. 1987. Power and centrality: A family of measures. ***American Journal of Sociology***, 92: 1170–1182.

Borgatti, S. P., & Everett, M. G. 2000. Models of core/periphery structures. ***Social Networks***, 21: 375–395.

Borgatti, S. P., & Halgin, D. S. 2011. On network theory. ***Organization Science***, 22: 1168–1181.

Borsboom, D., & Cramer, A. O. J. 2013. Network analysis: An integrative approach to the structure of psychopathology. ***Annual Review of Clinical Psychology***, 9: 91–121.

Brescoll, V. L., & Uhlmann, E. L. 2008. Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace. ***Psychological Science***, 19: 268–275.

Bundy, J., Pfarrer, M. D., Short, C. E., & Coombs, W. T. 2017. Crises and crisis management: Integration, interpretation, and research development. ***Journal of Management***, 43: 1661–1692.

Burt, R. 1992. ***Structural holes***. Cambridge, MA: Harvard University Press.

Cao, R., Koning, R., & Nanda, R. 2024. Sampling bias in entrepreneurial experiments. ***Management Science***, 70: 7283–7307.

Carley, K. 1993. Coding choices for textual analysis: A comparison of content analysis and map analysis. ***Sociological Methodology***, 23: 75–126.

Carley, K. 1994. Extracting culture through textual analysis. ***Poetics***, 22: 291–312.

Carley, K. M. 1997. Extracting team mental models through textual analysis. ***Journal of Organizational Behavior***, 18: 533–558.

Carley, K. M., & Kaufer, D. S. 1993. Semantic connectivity: An approach for analyzing symbols in semantic networks. ***Communication Theory***, 3: 183–213.

Carton, A. M., Murphy, C., & Clark, J. R. 2014. A (blurry) vision of the future: How leader rhetoric about ultimate goals influences performance. ***Academy of Management Journal***, 57: 1544–1570.

Christensen, A. P., & Kenett, Y. N. 2023. Semantic network analysis (SemNA): A tutorial on preprocessing, estimating, and analyzing semantic networks. ***Psychological Methods***, 28: 860–879.

Claes, K., & Godart, F. 2025. What keeps the market ticking? The role of third-party audiences and cognitive embeddedness in shaping competitive dynamics in luxury watchmaking. ***Strategic Management Journal***, 46: 667–692.

Collins, A. M., & Loftus, E. F. 1975. A spreading-activation theory of semantic processing. ***Psychological Review***, 82: 407–428.

Collins, A. M., & Quillian, M. R. 1969. Retrieval time from semantic memory. ***Journal of Verbal Learning and Verbal Behavior***, 8: 240–247.

Cornelissen, J. P., & Werner, M. D. 2014. Putting framing in perspective: A review of framing and frame analysis across the management and organizational literature. ***Academy of Management Annals***, 8: 181–235.
Corritore, M., Goldberg, A., & Srivastava, S. B. 2020. Duality in diversity: How intrapersonal and interpersonal cultural heterogeneity relate to firm performance. ***Administrative Science Quarterly***, 65: 359–394.
Crilly, D. 2017. Time and space in strategy discourse: Implications for intertemporal choice. ***Strategic Management Journal***, 38: 2370–2389.
Danowski, J. A. 1982. A network-based content analysis methodology for computer-mediated communication: An illustration with a computer bulletin board. ***Communication Yearbook***, 6: 904–925.
Danowski, J. A. 1993. Network analysis of message content. In W. D. Richards & G. A. Barnett (Eds.), ***Progress in communication sciences***: 198–221. Norwood, NJ: Ablex Publishing.
DeChurch, L. A., & Mesmer-Magnus, J. R. 2010. Measuring shared team mental models: A meta-analysis. ***Group Dynamics: Theory, Research, and Practice***, 14: 1–14.
Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. ***Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies***: 4171–4186. Minneapolis, MN: Association for Computational Linguistics.
Dewey, J. 1932. ***The later works of John Dewey, volume 7: Ethics***. Carbondale: Southern Illinois University Press.
Dubossarsky, H., De Deyne, S., & Hills, T. T. 2017. Quantifying the structure of free association networks across the life span. ***Developmental Psychology***, 53: 1560–1570.
Durand, R., Granqvist, N., & Tyllström, A. 2017. From categories to categorization: A social perspective on market categorization. In R. Durand, N. Granqvist, & A. Tyllström (Eds.), ***From categories to categorization: Studies in sociology, organizations and strategy at the crossroads*** (Research in the Sociology of Organizations, vol. 51): 3–30. Bingley, U.K.: Emerald Publishing.
Eagly, A. H., & Karau, S. J. 2002. Role congruity theory of prejudice toward female leaders. ***Psychological Review***, 109: 573–598.
Elsbach, K. D. 1994. Managing organizational legitimacy in the California cattle industry: The construction and effectiveness of verbal accounts. ***Administrative Science Quarterly***, 39: 57–88.
Fiss, P. C., & Zajac, E. J. 2006. The symbolic management of strategic change: Sensegiving via framing and decoupling. ***Academy of Management Journal***, 49: 1173–1193.
Freeman, L. C. 1977. A set of measures of centrality based on betweenness. ***Sociometry***, 40: 35–41.
Freeman, L. C. 1978. Centrality in social networks: Conceptual clarification. ***Social Networks***, 1: 215–239.
Gamache, D. L., McNamara, G., Mannor, M. J., & Johnson, R. E. 2015. Motivated to acquire? The impact of CEO regulatory focus on firm acquisitions. ***Academy of Management Journal***, 58: 1261–1282.
Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. ***Proceedings of the National Academy of Sciences***, 115: E3635–E3644.

Gavetti, G., Levinthal, D., & Ocasio, W. 2007. Neo-Carnegie: The Carnegie school's past, present, and reconstructing for the future. ***Organization Science***, 18: 523–536.
Giorgi, S., Maoret, M., & Zajac, E. J. 2019. On the relationship between firms and their legal environment: The role of cultural consonance. ***Organization Science***, 30: 803–830.
Glaser, V. L., Fast, N. J., Harmon, D., & Green, S. E., Jr. 2016. Institutional frame switching: How institutional logics shape individual action. In J. Gehman, M. Lounsbury, & R. Greenwood (Eds.), ***How institutions matter!*** (Research in the Sociology of Organizations, vol. 48A): 35–69. Bingley, U.K.: Emerald Group Publishing.
Godart, F. C., & Claes, K. 2017. Semantic networks and the market interface: Lessons from luxury watchmaking. In P. Groenewegen, J. E. Ferguson, C. Moser, J. W. Mohr, & S. P. Borgatti (Eds.), ***Structure, content and meaning of organizational networks: Extending network thinking*** (Research in the Sociology of Organizations, vol. 53): 113–141. Bingley, U.K.: Emerald Publishing.
Godart, F. C., & Galunic, C. 2019. Explaining the popularity of cultural elements: Networks, culture, and the structural embeddedness of high fashion trends. ***Organization Science***, 30: 151–168.
Goldberg, A. 2011. Mapping shared understandings using relational class analysis: The case of the cultural omnivore reexamined. ***American Journal of Sociology***, 116: 1397–1436.
Goldberg, A., Srivastava, S. B., Manian, V. G., Monroe, W., & Potts, C. 2016. Fitting in or standing out? The tradeoffs of structural and cultural embeddedness. ***American Sociological Review***, 81: 1190–1222.
Granqvist, N., & Siltaoja, M. 2020. Constructions, claims, resonance, reflexivity: Language and market categorization. ***Organization Theory***, 1: 2631787720968561.
Green, S. E., Jr., Li, Y., & Nohria, N. 2009. Suspended in self-spun webs of significance: A rhetorical model of institutionalization and institutionally embedded agency. ***Academy of Management Journal***, 52: 11–36.
Greenwood, R., & Hinings, C. R. 1996. Understanding radical organizational change: Bringing together the old and the new institutionalism. ***Academy of Management Review***, 21: 1022–1054.
Grimmer, J., Roberts, M. E., & Stewart, B. M. 2022. ***Text as data: A new framework for machine learning and the social sciences***. Princeton, NJ: Princeton University Press.
Hannan, M. T. 2022. Measuring memberships in collectives in light of developments in cognitive science and natural-language processing. ***Sociological Science***, 9: 473–492.
Hanneman, R. A., & Riddle, M. 2005. ***Introduction to social network methods***. Riverside, CA: University of California, Riverside.
Hannigan, T., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V., Wang, M., Kaplan, S., & Jennings, P. D. 2019. Topic modeling in management research: Rendering new theory from textual data. ***Academy of Management Annals***, 13: 586–632.
Harmon, D. 2019. When the fed speaks: Arguments, emotions, and the microfoundations of institutions. ***Administrative Science Quarterly***, 64: 542–575.
Harmon, D., Rhee, E., & Cho, Y. H. 2023. Building a bridge to the future: Prospective legitimation in nascent markets. ***Strategic Management Journal***, 44: 2597–2633.
Heilman, M. E., & Haynes, M. C. 2005. No credit where credit is due: Attributional rationalization of women's success in male-female teams. ***Journal of Applied Psychology***, 90: 905–916.

Höllerer, M. A., Jancsary, D., Barberio, V., & Meyer, R. E. 2020. The interlinking theorization of management concepts: Cohesion and semantic equivalence in management knowledge. ***Organization Studies***, 41: 1284–1310.
Hsu, G. 2006. Jacks of all trades and masters of none: Audiences' reactions to spanning genres in feature film production. ***Administrative Science Quarterly***, 51: 420–450.
Hsu, G., Hannan, M. T., & Koçak, Ö. 2009. Multiple category memberships in markets: An integrative theory and two empirical tests. ***American Sociological Review***, 74: 150–169.
Jancsary, D., Meyer, R. E., Höllerer, M. A., & Barberio, V. 2017. Toward a structural model of organizational-level institutional pluralism and logic interconnectedness. ***Organization Science***, 28: 1150–1167.
Jones, C., Maoret, M., Massa, F. G., & Svejenova, S. 2012. Rebels with a cause: Formation, contestation, and expansion of the de novo category "modern architecture," 1870–1975. ***Organization Science***, 23: 1523–1545.
Kaplan, S., & Tripsas, M. 2008. Thinking about technology: Applying a cognitive lens to technical change. ***Research Policy***, 37: 790–805.
Kenett, Y. N., & Faust, M. 2019. A semantic network cartography of the creative mind. ***Trends in Cognitive Sciences***, 23: 271–274.
Kenett, Y. N., Anaki, D., & Faust, M. 2014. Investigating the structure of semantic networks in low and high creative persons. ***Frontiers in Human Neuroscience***, 8: 407.
Kenett, Y. N., Levi, E., Anaki, D., & Faust, M. 2017. The semantic distance task: Quantifying semantic distance with semantic network path length. ***Journal of Experimental Psychology: Learning, Memory, and Cognition***, 43: 1470–1489.
Kennedy, M. T. 2008. Getting counted: Markets, media, and reality. ***American Sociological Review***, 73: 270–295.
Kim, N., Kim, J. H., Wolters, M. K., MacPherson, S. E., & Park, J. C. 2019. Automatic scoring of semantic fluency. ***Frontiers in Psychology***, 10: 1020.
Kovács, B., & Hannan, M. T. 2015. Conceptual spaces and the consequences of category spanning. ***Sociological Science***, 2: 252–286.
Kozlowski, A. C., Taddy, M., & Evans, J. A. 2019. The geometry of culture: Analyzing the meanings of class through word embeddings. ***American Sociological Review***, 84: 905–949.
Kwon, S. W., Rondi, E., Levin, D. Z., De Massis, A., & Brass, D. J. 2020. Network brokerage: An integrative review and future research agenda. ***Journal of Management***, 46: 1092–1120.
Lawrence, T. B., & Suddaby, R. 2006. Institutions and institutional work. In S. R. Clegg, C. Hardy, T. B. Lawrence, & W. R. Nord (Eds.), ***The Sage handbook of organization studies*** (2nd ed.): 215–254. London: Sage.
Le Mens, G., Kovács, B., Hannan, M. T., & Pros, G. 2023. Using machine learning to uncover the semantics of concepts: How well do typicality measures extracted from a BERT text classifier match human judgments of genre typicality? ***Sociological Science***, 10: 82–117.
Lerner, A. J., Ogrocki, P. K., & Thomas, P. J. 2009. Network graph analysis of category fluency testing. ***Cognitive and Behavioral Neurology***, 22: 45–52.
Lewis, K. 2003. Measuring transactive memory systems in the field: Scale development and validation. ***Journal of Applied Psychology***, 88: 587–604.
Lix, K., Goldberg, A., Srivastava, S. B., & Valentine, M. A. 2022. Aligning differences: Discursive diversity and team performance. ***Management Science***, 68: 8430–8448.

Lounsbury, M., & Glynn, M. A. 2001. Cultural entrepreneurship: Stories, legitimacy, and the acquisition of resources. ***Strategic Management Journal***, 22: 545–564.
Lounsbury, M., & Glynn, M. A. 2019. ***Cultural entrepreneurship: A new agenda for the study of entrepreneurial processes and possibilities*** (Elements in Organization Theory). Cambridge, U.K.: Cambridge University Press.
Lounsbury, M., & Rao, H. 2004. Sources of durability and change in market classifications: A study of the reconstitution of product categories in the American mutual fund industry, 1944–1985. ***Social Forces***, 82: 969–999.
Lyness, K. S., & Heilman, M. E. 2006. When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. ***Journal of Applied Psychology***, 91: 777–785.
Maitlis, S., & Christianson, M. 2014. Sensemaking in organizations: Taking stock and moving forward. ***Academy of Management Annals***, 8: 57–125.
Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. 2000. The influence of shared mental models on team process and performance. ***Journal of Applied Psychology***, 85: 273–283.
McNally, R. J., Robinaugh, D. J., Wu, G. W. Y., Wang, L., Deserno, M. K., & Borsboom, D. 2015. Mental disorders as causal systems: A network approach to posttraumatic stress disorder. ***Clinical Psychological Science***, 3: 836–849.
McPherson, C. M., & Sauder, M. 2013. Logics in action: Managing institutional complexity in a drug court. ***Administrative Science Quarterly***, 58: 165–196.
Mikolov, T., Chen, K., Corrado, G., & Dean, J. 2013. Efficient estimation of word representations in vector space. ***Proceedings of the International Conference on Learning Representations (ICLR), Workshop Track***: 1–12.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. 2013. Distributed representations of words and phrases and their compositionality. ***Advances in Neural Information Processing Systems***, 26: 3111–3119.
Mikolov, T., Yih, W., & Zweig, G. 2013. Linguistic regularities in continuous space word representations. ***Proceedings of NAACL-HLT***: 746–751.
Newman, M. E. J. 2006. Modularity and community structure in networks. ***Proceedings of the National Academy of Sciences***, 103: 8577–8582.
Ocasio, W. 1997. Towards an attention-based view of the firm. ***Strategic Management Journal***, 18: 187–206.
Okhuysen, G. A., & Bechky, B. A. 2009. Coordination in organizations: An integrative perspective. ***Academy of Management Annals***, 3: 463–502.
Peirce, C. S. 1934. ***Collected papers of Charles Sanders Peirce***. Cambridge, MA: Harvard University Press.
Pennington, J., Socher, R., & Manning, C. D. 2014. GloVe: Global vectors for word representation. ***Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)***: 1532–1543. Doha, Qatar: Association for Computational Linguistics.
Pontikes, E. G. 2022. Category innovation in the software industry: 1990–2002. ***Strategic Management Journal***, 43: 1697–1727.
Quillian, M. R. 1966. ***Semantic memory***. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.
Quillian, M. R. 1967. Word concepts: A theory and simulation of some basic semantic capabilities. ***Behavioral Science***, 12: 410–430.

Quine, W. V. O. 1960. ***Word and object***. Cambridge, MA: MIT Press.
Quintane, E., & Carnabuci, G. 2016. How do brokers broker? Tertius gaudens, tertius iungens, and the temporality of structural holes. ***Organization Science***, 27: 1343–1360.
Rastelli, C., Greco, A., & Finocchiaro, C. 2020. Revealing the role of divergent thinking and fluid intelligence in children's semantic memory organization. ***Journal of Intelligence***, 8: 43.
Rhee, E. Y., & Fiss, P. C. 2014. Framing controversial actions: Regulatory focus, source credibility, and stock market reaction to poison pill adoption. ***Academy of Management Journal***, 57: 1734–1758.
Rhee, L., & Leonardi, P. M. 2018. Which pathway to good ideas? An attention-based view of innovation in social networks. ***Strategic Management Journal***, 39: 1188–1215.
Rousseau, D. M., & Tijoriwala, S. A. 1999. What's a good reason to change? Motivated reasoning and social accounts in promoting organizational change. ***Journal of Applied Psychology***, 84: 514–528.
Santana, J. J., & Kim, S. 2026. From values to codes: A computational text analysis of the codification of occupational ethics. ***Organization Studies***, 47: 127–156.
Saussure, F. de. 1916. ***Course in general linguistics***. Paris: Payot.
Short, J. C., McKenny, A. F., & Reid, S. W. 2018. More than words? Computer-aided text analysis in organizational behavior and psychology research. ***Annual Review of Organizational Psychology and Organizational Behavior***, 5: 415–435.
Siew, C. S. Q., Wulff, D. U., Beckage, N. M., & Kenett, Y. N. 2019. Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. ***Complexity***, 2019: 2108423.
Song, J., & Harmon, D. 2026. Ambiguous entrepreneurial communication and mobilizing audience support for novel ideas. Working paper.
Srivastava, S. B., & Goldberg, A. 2017. Language as a window into culture. ***California Management Review***, 60: 56–69.
Srivastava, S. B., Goldberg, A., Manian, V. G., & Potts, C. 2018. Enculturation trajectories: Language, cultural adaptation, and individual outcomes in organizations. ***Management Science***, 64: 1348–1364.
Stam, D., Lord, R. G., van Knippenberg, D., & Wisse, B. 2014. An image of who we might become: Vision communication, possible selves, and vision pursuit. ***Organization Science***, 25: 1172–1194.
Suddaby, R., & Greenwood, R. 2005. Rhetorical strategies of legitimacy. ***Administrative Science Quarterly***, 50: 35–67.
Taeuscher, K., Bouncken, R. B., & Pesch, R. 2021. Gaining legitimacy by being different: Optimal distinctiveness in crowdfunding platforms. ***Academy of Management Journal***, 64: 149–179.
Thornton, P. H., Ocasio, W., & Lounsbury, M. 2012. ***The institutional logics perspective: A new approach to culture, structure, and process***. Oxford, U.K.: Oxford University Press.
Uzzi, B., & Spiro, J. 2005. Collaboration and creativity: The small world problem. ***American Journal of Sociology***, 111: 447–504.
Valentine, M. 2018. When equity seems unfair: The role of justice enforceability in temporary team coordination. ***Academy of Management Journal***, 61: 2081–2105.
Veremyev, A., Semenov, A., Pasiliao, E. L., & Boginski, V. 2019. Graph-based exploration and clustering analysis of semantic spaces. ***Applied Network Science***, 4: 104.

Vicinanza, P., Goldberg, A., & Srivastava, S. B. 2023. A deep-learning model of prescient ideas demonstrates that they emerge from the periphery. ***PNAS Nexus***, 2: pgac275.

Vossen, A., & Ihl, C. 2020. More than words! How narrative anchoring and enrichment help to balance differentiation and conformity of entrepreneurial products. ***Journal of Business Venturing***, 35: 106050.

Walsh, J. P. 1995. Managerial and organizational cognition: Notes from a trip down memory lane. ***Organization Science***, 6: 280–321.

Wasserman, S., & Faust, K. 1994. ***Social network analysis: Methods and applications***. Cambridge, U.K.: Cambridge University Press.

Watts, D. J., & Strogatz, S. H. 1998. Collective dynamics of 'small-world' networks. ***Nature***, 393: 440–442.

Weick, K. E. 1993. The collapse of sensemaking in organizations: The Mann Gulch disaster. ***Administrative Science Quarterly***, 38: 628–652.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. 2005. Organizing and the process of sensemaking. ***Organization Science***, 16: 409–421.

Wittgenstein, L. 1953. ***Philosophical investigations***. Oxford, U.K.: Basil Blackwell.

Wulff, D. U., Hills, T. T., & Mata, R. 2022. Structural differences in the semantic networks of younger and older adults. ***Scientific Reports***, 12: 21459.

Zhang, E. Y., Aven, B. L., & Kleinbaum, A. M. 2024. License to broker: How mobility eliminates gender gaps in network advantage. ***Administrative Science Quarterly***, 69: 227–270.

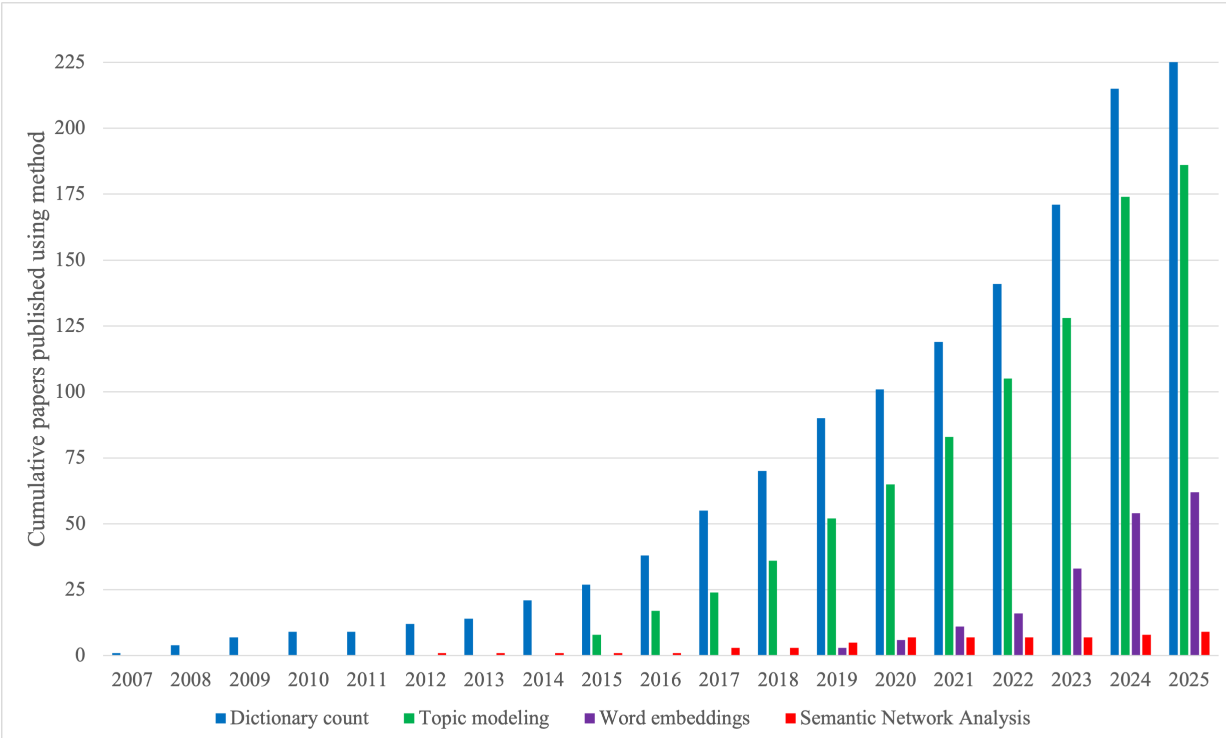
Ziegert, J. C., & Hanges, P. J. 2005. Employment discrimination: The role of implicit attitudes, motivation, and a climate for racial bias. ***Journal of Applied Psychology***, 90: 553–562.

Zilber, T. B. 2002. Institutionalization as an interplay between actions, meanings, and actors: The case of a rape crisis center in Israel. ***Academy of Management Journal***, 45: 234–254.

Zott, C., & Huy, Q. N. 2007. How entrepreneurs use symbolic management to acquire resources. ***Administrative Science Quarterly***, 52: 70–105.

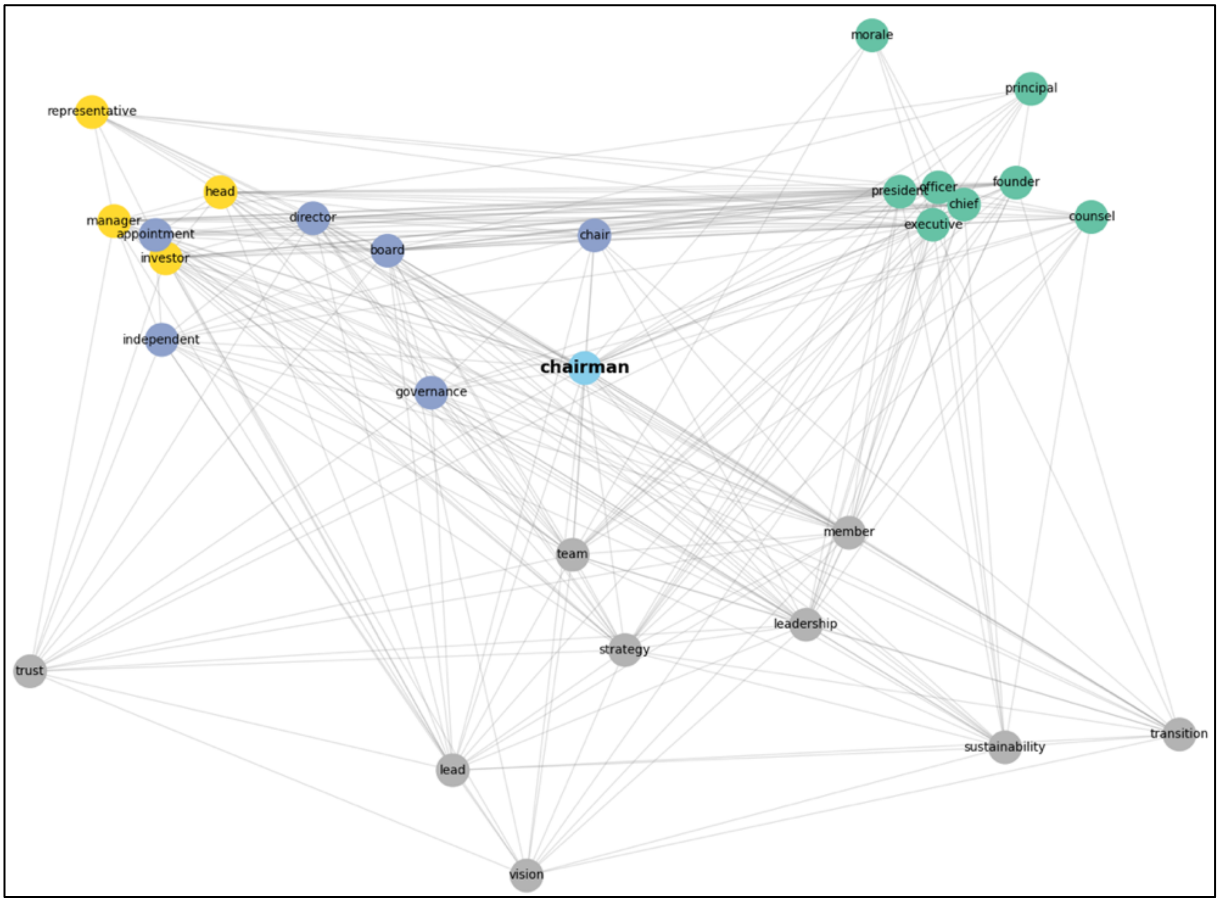
Zunino, D., Suarez, F. F., & Grodal, S. 2019. Familiarity, creativity, and the adoption of category labels in technology industries. ***Organization Science***, 30: 169–190.

Figure 1. Papers published in organization journals using NLP methods



Note: Papers published in *Academy of Management Journal*, *Administrative Science Quarterly*, *Journal of Applied Psychology*, *Journal of Management*, *Organizational Behavior and Human Decision Processes*, *Organization Science*, *Organization Studies*, *Management Science*, and *Strategic Management Journal*.

Figure 2. Ego networks of "chairman" and "chairwoman" in the explicit usage network



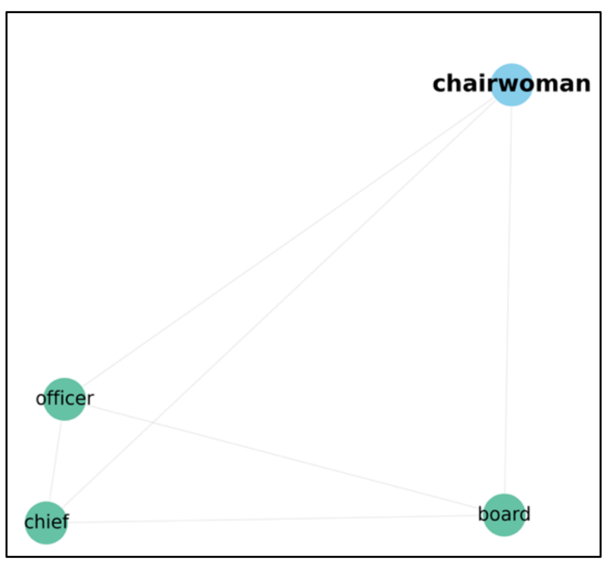


Figure 3. Ego network of “chairman” in the latent meaning network

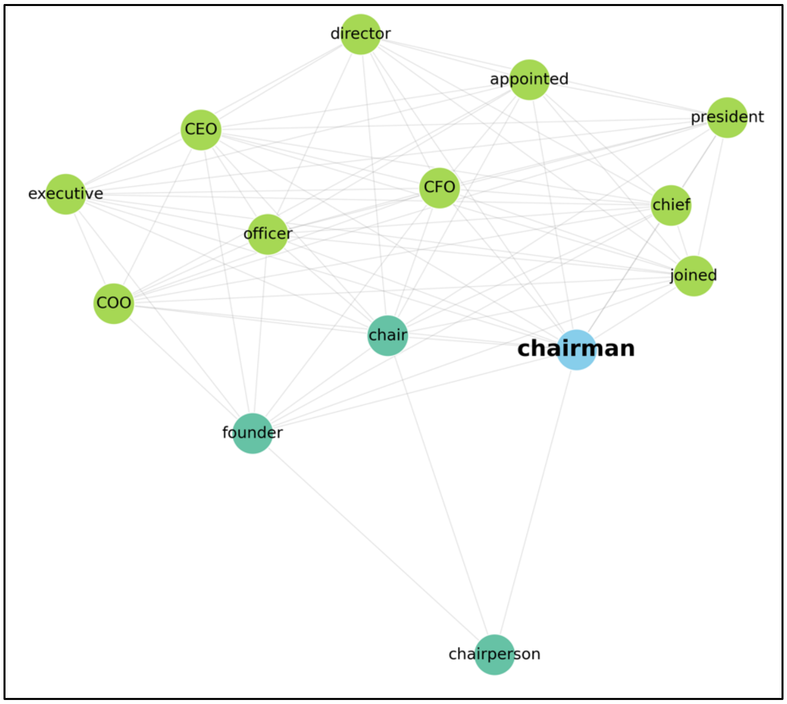
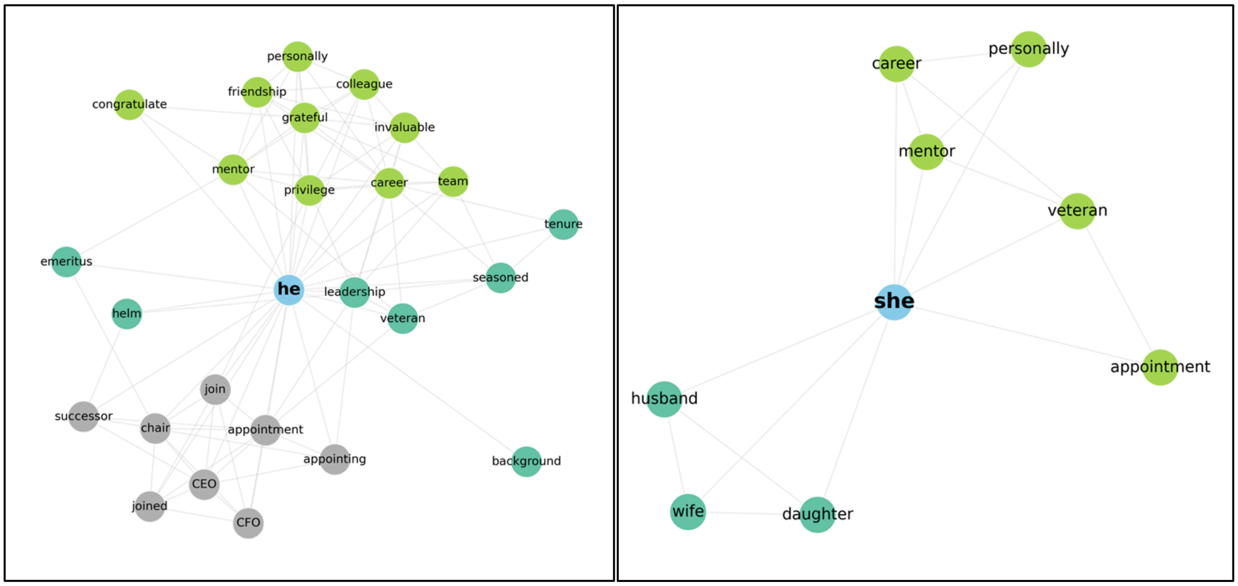


Figure 4. Ego networks of "he" and "she" in the latent meaning network



Note: "he" and "she" sit close in the embedding, so each falls in the other's neighborhood; because that tie reflects only their shared status as pronouns, it is excluded from both ego networks.

Table 1. Two Types of Semantic Networks

	Explicit Usage Semantic Network	**Latent Meaning Semantic Network**
What it captures	Concepts that authors directly link in discourse	Concepts used in similar contexts, even when not directly linked
Input corpus	Targeted and theoretically aligned with research question, need not be large	Large enough to capture background linguistic regularities
Nodes	Researcher-selected words or phrases	Vectors learned from the corpus (e.g., Word2Vec, GloVe, BERT)
Ties	Co-occurrence within a defined contextual window	Cosine similarity, continuous measure or above a chosen threshold
When useful	Studying how meaning is publicly communicated, contested, and negotiated	Studying taken-for-granted associations and cognitive schemas that surface discourse cannot reveal

Table 2. Measuring Semantic Networks

Level	Measure	Definition	Explicit Usage Interpretation	Latent Meaning Interpretation
Dyad	Existence of tie	Whether two nodes are connected	Two concepts appear together in text, indicating linguistic coupling	Two concepts have similar contextual usage, indicating shared underlying meaning
Node	**Nodal Prominence**			
	Degree centrality	Number or total weight of ties incident on a node	Concepts broadly used with others in discourse	Concepts broadly related to others in shared meaning space
	Eigenvector centrality	Influence of a node based on ties to other highly connected nodes	Concepts that anchor the core vocabulary by connecting to other widely-used terms	Concepts that anchor most cohesive regions of meaning by connecting to other broadly-related terms
	Structural Position			
	Shortest path length	Minimum number of steps connecting two nodes	Whether two concepts are directly coupled in discourse or linked only through specific intermediaries	How many associative steps separate two concepts in the underlying semantic structure
	Betweenness centrality	How often a concept lies on the shortest paths between other pairs of concepts	Concepts serving as discursive bridges between otherwise weakly linked topical domains	Concepts serving as cognitive intermediaries between otherwise distant regions of meaning
	Core-periphery	Whether a concept occupies the densely interconnected core or the structurally dependent periphery	Concepts structurally central to discourse, appearing across a wide range of contexts	Concepts in the densely interconnected core of the latent semantic space, around which peripheral concepts cluster
	Neighborhood Structure			
	Ego density	Proportion of realized ties among a node's immediate neighbors	Concepts whose immediate linguistic context is tightly interrelated	Concepts whose latent neighbors are themselves mutually similar, indicating a tightly bounded conceptual domain
	Brokerage	Degree to which a concept's ties are non-redundant, spanning structural holes among its neighbors	Concepts that integrate separate discourses or audiences	Concepts whose latent neighbors fall into multiple disconnected sub-regions, indicating the concept carries associations from semantically distinct domains

Level	Measure	Definition	Explicit Usage Interpretation	Latent Meaning Interpretation
Network	**Aggregate Connectivity**			
	Density	Proportion of all possible ties present in the network	How interlinked the language of a corpus is	How coherent the underlying meaning system is
	Centralization	Whether a few hub concepts dominate or connectivity is dispersed across many nodes	Whether discourse is organized around a small number of dominant framing concepts or is more distributed	Whether the cognitive architecture is anchored by a few central concepts or reflects a more evenly distributed structure
	Network Segmentation			
	Modularity	Extent to which nodes form internally cohesive clusters with few external ties	Discursive communities or topical clusters in language use	Distinct cognitive domains or meaning subsystems within the semantic space, with high modularity indicating clearly separated regions of meaning
	Network Reach			
	Average path length	Average number of steps between any two concepts in the network	Whether any two concepts can be linked through few intermediary co-occurrences	Whether any two concepts are associatively proximate or separated by many steps
	Small-worldliness	Whether the network simultaneously exhibits strong local clustering and short global path lengths	Discourse exhibiting both coherent topical clusters and short global pathways	Meaning system exhibiting both tight local coherence and short global associative distances

Table 3. Using SemNA to Advance Organizational Scholarship

Level of Analysis	Explicit Usage Semantic Network	Latent Meaning Semantic Network
Internal Organizational Activities		
Leader Communication	How do leaders motivate employees or influence external stakeholders? (Carton et al., 2014; Elsbach, 1994; Rhee & Fiss, 2014, Stam et al., 2014)	What are the implicit biases or stereotypes that leaders convey in their communications? (Heilman & Haynes, 2005; Lyness & Heilman 2006; Ziegert & Hanges, 2005)
Sensemaking & Crises	How do actors give sense to or construct a coherent narrative during a crisis? (Bundy et al., 2017; Fiss & Zajac 2006; Maitlis & Christianson, 2014)	What is the underlying structure of reality during a crisis, and does it resist change in the face of external shocks? (Harmon, 2019; Weick, 1993; Weick et al., 2005)
Teams & Groups	How do teams actively coordinate and develop a shared vocabulary to manage daily tasks and conflict? (Okhuysen & Bechky, 2009; Valentine, 2018)	To what extent do teams share a deep, implicit structure of tasks or knowledge? (Lewis, 2003; DeChurch & Mesmer-Magnus, 2010; Mathieu et al., 2000)
Organizational Attention & Cognition	How is organizational attention explicitly directed toward specific issues during decision-making? (Ocasio, 1997; Rhee & Leonardi, 2018)	What are the sedimented cognitive structures or dominant logics that filter information and constrain search? (Gavetti et al., 2007; Kaplan & Tripsas, 2008; Walsh, 1995)
Organizational Culture	How does an organization's culture change over time? (Greenwood & Hinings, 1996; Rousseau & Tijoriwala, 1999; Zilber, 2002)	What are the deep cultural codes that persist beneath the surface of an organization's culture? (Corritore et al., 2020; Goldberg, 2011; Srivastava & Goldberg, 2017)
Market and Institutional Activities		
Categories	How do new categories emerge or how do actors use categories strategically? (Durand et al., 2017; Granqvist & Siltaoja, 2020; Kennedy, 2008)	What are the features of a given market category, and how do they overlap with other market categories? (Hsu, 2006; Hsu et al., 2009; Lounsbury & Rao, 2004; Zunino et al., 2019)
Cultural Entrepreneurship	How do entrepreneurs construct or shape new markets? (Harmon et al., 2023; Lounsbury & Glynn, 2001; Zott & Huy, 2007)	What are the latent cognitive landscapes entrepreneurs can draw upon as a stock of knowledge to gain legitimacy? (Lounsbury & Glynn, 2019, Song & Harmon, 2026)
Institutions	How do actors use rhetorical strategies to create, disrupt, or maintain institutions? (Lawrence & Suddaby, 2006; Suddaby & Greenwood, 2005)	What are the deep, sedimented belief systems and cultural codes that structure a field? (Jones et al., 2012; Glaser et al., 2016; Jancsary et al., 2017; Thornton et al., 2012)

APPENDIX A

Earnings call data construction and text cleaning

Our sample includes all publicly available earnings call transcripts from U.S. firms from Q1 to Q4 2024, obtained through the Financial Modeling Prep (FMP) API. The corpus contains transcripts for the 3,195 companies with at least one transcript available during the period, providing substantial variation in linguistic style, topical content, and firm context. Each earnings call includes both a prepared presentation, where executives deliver scripted remarks, and a Q&A session, where analysts' questions often prompt unscripted responses. For preprocessing we used NLTK. We lowercased the text and removed URLs, user mentions, non-ASCII characters, numbers, and other special characters, segmented each transcript into sentences at sentence-final punctuation, and retained sentences with more than two tokens. We removed English stop words but retained the gendered pronouns (he, him, his, himself, she, her, hers, herself) as distinct tokens and lemmatized the remaining tokens with the WordNet lemmatizer. This resulted in a total of 7,698,626 sentences accounting for 74,983,189 words.

APPENDIX B

Node and tie selection criteria for explicit usage network

Nodes. The explicit usage network's nodes are drawn from a conceptually grounded lexicon of gender-, role-, and leadership-related terms, organized into six theoretically grounded categories: gender identity terms, family roles, professional titles, leadership traits, leader competence, and leader actions. Within each category, specific terms are selected by drawing on prior organizational and linguistic research, including agentic and communal trait taxonomies from Brescoll and Uhlmann (2008) and Eagly and Karau (2002), gendered occupational title lists from prior computational linguistics work (Bolukbasi et al., 2016; Garg et al., 2018), and competence and action vocabularies drawn from leadership and strategy research (Bass, 1990; Carton et al., 2014). The full word list across the six categories is provided in Table B1 below.

Table B1. Lexicon for the explicit usage network, by category and subcategory.

Category	Subcategory	Terms
Gender identity	Gendered pronouns	he, his, him, himself, she, her, hers, herself
	Gendered nouns	man, men, woman, women, boy, girl
Family roles	Female	mother, mom, daughter, sister, wife, grandmother, granddaughter, aunt, niece, matriarch
	Neutral or male	father, dad, son, brother, husband, grandfather, grandson, uncle, nephew, patriarch
Professional titles	Neutral	chairperson, chief, chair, head, lead, leader, director, boss, counsel, operator, attorney, board, member, colleague, mentor, successor, founder, co-founder, cofounder, president, officer, executive, investor, manager, principal, partner, analyst, consultant, advisor, appointee, nominee, spokesperson, representative, delegate, recruiter, pioneer, veteran, expert
	Gendered male	chairman, chairmen, businessman, businessmen, salesman, salesmen, foreman, foremen, spokesman, spokesmen
	Gendered female	chairwoman, chairwomen, businesswoman, businesswomen, saleswoman, saleswomen, forewoman, forewomen, spokeswoman, spokeswomen
Leadership traits	Agentic, male-coded	aggressive, assertive, ambitious, competitive, decisive, dominant, forceful, independent, confident, bold, driven, tough, disciplined, tenacious, analytical, logical, strategic, visionary, pioneering, outspoken, spearhead,

		champion, outperform, execute, deliver, win, lead, dominate, accelerate, disrupt, strong, skilled, effective, efficient, experienced, expert, proven, qualified
	Communal, female-coded	collaborative, supportive, empathetic, understanding, nurturing, helpful, cooperative, communal, capable, compassionate, interpersonal, inclusive, transparent, trustworthy, honest, responsible, committed, dedicated, caring, warm, listening, mentoring, developing, cultivating, engaging, connecting, uniting, partnering
Leader competence	Strategic competence	strategy, vision, roadmap, direction, priority, initiative, objective, goal, milestone, execution, planning, transformation, pivot, repositioning, expansion, diversification, differentiation, focus, alignment, capability, advantage, positioning, portfolio, allocation, investment, mission, growth, performance, profitability, efficiency, spearhead, drive, lead, execute
	Operational competence	execution, operational, efficiency, scale, process, discipline, rigor, optimization, streamline, restructure, turnaround, integration, implementation, delivery, accountability, oversight, governance, control, compliance, infrastructure
	Financial performance	efficiency, productivity, performance, results, outlook, forecast, guidance, beat, exceed, outperform, underperform
	Market and competitive position	market, share, competition, competitive, positioning, differentiation, leadership, dominance, advantage, moat, brand, reputation, customer, demand, retention, loyalty, pricing, value, deliver, build, scale, transform, accelerate
	Innovation and growth	innovation, disruption, transformation, technology, digitization, automation, AI, platform, ecosystem, product, pipeline, launch, commercialization, scaling, adoption, penetration, expansion, acquisition, partnership, collaboration
	Risk and resilience	risk, resilience, stability, durability, sustainability, continuity, recovery, mitigation, uncertainty, volatility, headwind, tailwind, challenge, crisis, pressure, navigate, adapt, respond
	People and talent	talent, team, culture, hiring, retention, development, succession, diversity, inclusion, engagement, morale, capability, bench, organization, structure
Leader actions	Leadership and decision-making	decision, judgment, leadership, direction, stewardship, oversight, accountability, credibility, integrity, confidence, conviction, commitment, resolve, clarity, communication, transparency, trust, consistency, experience, expertise, track record, tenure, succession, appointment, transition

Ties. Ties in the explicit usage network are formed based on sentence-level co-mentions. Two terms are considered connected if they appear together within the same sentence. We determine the co-occurrence threshold using the t-score, a statistical measure that assesses whether the level of co-occurrence between two terms is above the one that would be expected by chance given their individual frequencies. Examining the distribution of t-scores across all possible word pairs, we found that the top five percent corresponded to a value of approximately 2.5, which indicates a substantively meaningful association rather than random co-usage. We therefore adopted t-score ≥ 2.5 as our criterion for establishing ties.

Robustness. To ensure the validity of this threshold, we evaluated several global properties of networks constructed at adjacent cutoffs. Specifically, we compared network density, the number of connected components, and modularity at values near 2.5. These structural features remained largely stable across nearby thresholds, indicating that the network does not undergo discontinuous shifts at 2.5. We also found that the conceptual patterns and examples used in the Practical Application section remained consistent under alternative thresholds at the top one percent ($t = 6.10$) and top ten percent ($t = 1.89$) of the t-score distribution. Together, these checks provide confidence that the insights drawn from the explicit usage network are not sensitive to the threshold used.

APPENDIX C

Node and tie selection criteria for latent meaning network

Nodes. The latent meaning network's nodes are word vectors learned through Word2Vec embedding over the full vocabulary of the preprocessed 2024 earnings call corpus. We use Word2Vec rather than contextual embedding models such as BERT for two reasons. First, Word2Vec produces a single static vector per word, aligning directly with our framework's treatment of each word as a node in the semantic network; contextual embeddings vary by sentence and require additional aggregation choices to form a network (e.g., averaging across token instances, see Vicinanza, Goldberg, & Srivastava, 2023). Second, Word2Vec has been the dominant embedding model in adjacent organizational scholarship that constructs semantic similarity at the word level (e.g., Kozlowski et al., 2019; Lix et al., 2022), supporting comparability with established applications in this literature. We set the vector dimensionality to 100 to provide a sufficiently rich semantic space while minimizing the error of the word-context matrix (Mikolov, Sutskever, et al., 2013) and exclude words appearing fewer than 20 times to ensure stable vector estimates. This procedure produces a high-dimensional vector representation for each retained word, with semantically similar terms placed closer together in the embedding space based on the distributional contexts in which they appear across earnings calls.

Pronoun consolidation. The latent model is trained on the same preprocessed corpus as the explicit network, with one modification specific to the embedding. Where the explicit network retains the gendered pronoun forms as distinct tokens (Appendix A), the latent model consolidates them before training, collapsing him, his, and himself into "he" and her, hers, and herself into "she." This pools the distributional evidence across all surface forms, so that each gendered referent is represented by a single vector reflecting every context in which it appears rather than splitting that evidence across inflections.

Ties. Ties in the latent meaning network are defined based on cosine similarity between word vectors. Cosine similarity captures the degree to which two words appear in similar contexts and is an established measure of latent semantic association (Lix et al., 2022; Kozlowski et al., 2019; Mikolov, Yih, et al., 2013; Mikolov, Chen, et al., 2013; Vossen & Ihl, 2020). We adopt a threshold of 0.55 to designate the presence of a tie, a level that reflects a substantively meaningful degree of semantic overlap based on inspection of in-sample word pairs at and around this cutoff.

Robustness. To assess the robustness of this threshold, we examined global network properties — such as density, number of components, and modularity — across adjacent similarity cutoffs. These properties remained stable across nearby thresholds, indicating that the network does not undergo structural discontinuities at 0.55. We also confirmed that the he/she asymmetry reported in the gender-bias application is stable across these cutoffs. Table C1 reports the ego density and brokerage of "he" and "she" at cosine thresholds of 0.50, 0.55, and 0.60. At every cutoff, "she" occupies a denser and less-brokered neighborhood than "he," matching the directions reported in the main text. Together, these checks provide confidence that the findings are not sensitive to the cosine similarity threshold used.

Table C1. Ego density and brokerage of "he" and "she" in the latent meaning network across cosine-similarity thresholds

Threshold	**"he" ego density**	**"she" ego density**	**"he" brokerage**	**"she" brokerage**
0.50	0.174	0.271	0.970	0.951
0.55	0.243	0.321	0.936	0.838
0.60	0.200	0.333	0.762	0.689

APPENDIX D

Frequency-matched bootstrap robustness for the chairman/chairwoman comparison

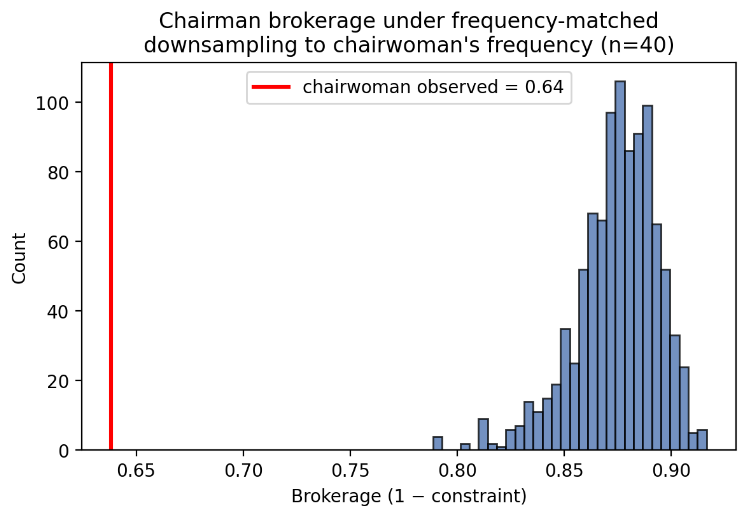
A potential concern with the chairman/chairwoman comparison in the explicit usage network is that the observed structural contrast — chairman as a multi-domain broker, chairwoman as enclosed within a narrow governance subspace — could reflect differences in frequency rather than genuine differences in topology. "Chairman" appears far more often than "chairwoman" in the 2024 corpus, and a more frequent term naturally has more opportunities to co-occur with diverse other terms, potentially expanding its observed neighborhood.

Our tie-selection procedure already addresses part of this concern by using t-scores rather than raw co-occurrence counts. T-scores adjust for baseline word frequencies, treating two terms as connected only when their level of co-occurrence exceeds what would be expected by chance given their individual frequencies. However, frequency also affects the number of opportunities a term has to co-occur with others — more sentences containing the term means more chances to find statistically meaningful associations across diverse contexts.

To isolate structural differences from sampling opportunity, we implement a sentence-level bootstrap. Specifically, we repeatedly downsample the set of sentences containing "chairman" to match "chairwoman's" total sentence count and reconstruct chairman's ego network from each downsample using the same t-score thresholding described in the main text. This yields a distribution of what chairman's ego network would look like if it occurred as rarely as chairwoman.

Figure D1 reports the resulting distributions. Even when chairman is downsampled to chairwoman's frequency, his brokerage remains well above hers in every iteration (see Figure D1). The bootstrap distribution (median 0.88) lies entirely above chairwoman's observed 0.64. This indicates that the asymmetry between the two terms reflects genuine topological differences in how they are used across leadership discourse rather than an artifact of unequal sampling opportunity.

Figure D1. Bootstrap distribution of chairman's brokerage under frequency-matched downsampling to chairwoman's frequency (n = 40).



Note: The set of sentences containing "chairman" is repeatedly downsampled (1,000 iterations) to match chairwoman's total sentence count, and chairman's brokerage (1 − constraint) is recomputed from each downsample. The red line marks chairwoman's observed brokerage (0.64). Chairman's brokerage remains well above hers in every iteration, indicating that the asymmetry reflects genuine topological differences rather than unequal sampling opportunity.

The same concern applies to nodal prominence. Eigenvector centrality rewards a term for connecting to other well-connected terms, and because a more frequent term has more opportunities to form such ties, chairman's higher eigenvector centrality (0.014, versus 0.001 for chairwoman) could in principle reflect frequency rather than a genuinely more central position. We test this with the same frequency-matched downsampling. In each iteration we draw chairman's sentences down to chairwoman's count, restrict chairman's ties to the alters that survive in that subsample, and recompute his eigenvector centrality on the full network. Across 1,000 iterations, chairman's matched eigenvector centrality has a median of 0.0058 (95%

interval [0.0024, 0.0094]), and in none of the iterations does it fall to or below chairwoman's observed value of 0.001 (P = 0.000). Frequency accounts for part of the raw gap — the matched median sits well below the observed 0.014 — but it does not explain it away. Even with as few sentences as chairwoman, chairman remains several times more central. The prominence asymmetry is therefore a structural feature of how the two titles are positioned in the discourse rather than an artifact of unequal sampling.

APPENDIX E

Data and code availability

Note for review team: For peer review, the `semna` package, the replication scripts, and the data artifacts needed to run them are all available through a restricted-access, view-only OSF archive (https://osf.io/zq2x4/?view_only=c79866358d7f4492b7997ffcc1bce6be). The public PyPI release, the GitHub repositories, and a Zenodo DOI described below follow on publication.

The `semna` Python package. We implement SemNA as an open-source Python package, `semna`, that constructs explicit usage and latent meaning semantic networks, computes the full set of topological measures introduced in the paper (Table 2), and exports networks for downstream analysis or visualization. The package also ships three public-domain corpora with worked vignettes so the framework can be applied without proprietary data. The package (version 0.1.0, MIT license) is provided through the restricted-access OSF archive for review. On publication it will be released on PyPI (installable via `pip install semna`) and on GitHub at [repository URL added on publication], with a Zenodo DOI for the version used here.

Replication archive. A separate replication archive at [repository URL added on publication] (Zenodo DOI to be minted on acceptance) reproduces every value reported in the empirical illustration. The archive ships the canonical artifacts — the trained word-embedding model and the derived explicit usage and latent meaning network files (in GML format) — together with three scripts that recompute the reported quantities: one that computes the explicit usage network measures from the shipped network file; one that reconstructs the latent meaning network from the trained model at the reported cosine threshold and computes its measures (the consolidated pronoun measures, the reported cosine similarities, and the threshold-robustness values in Table C1), with the shipped latent network file provided for inspection; and one for the frequency-matched bootstrap underlying Appendix D. A `README` maps each reported value to the script that produces it.

Reproducibility. We distribute the trained model and derived networks as the canonical reference because the embedding step is not bit-for-bit deterministic. Word2Vec training with multiple worker threads depends on the order in which gradient updates are applied, so even with a fixed random seed, retraining on the same corpus produces a model whose cosine similarities differ from the shipped model by a small amount (on the order of thousandths to hundredths per word pair); ties near the latent network's cosine threshold can flip in or out on a retrain, and downstream measures inherit this drift at the margin. The explicit usage network, by contrast, is fully deterministic given the same corpus and library versions: its ties are integer co-occurrence counts filtered by an exact t-score threshold (see Appendix A).

Earnings call corpus. The empirical analyses use earnings call transcripts from all U.S. public firms in calendar year 2024, obtained through the Financial Modeling Prep (FMP) API. The 3,195 firms in the corpus are those with available 2024 transcripts, as identified by FMP's earnings-call transcript-list endpoint (not the general stock universe, which is far larger); transcripts were then retrieved for each firm and quarter through FMP's earnings-call transcript endpoint (v3). The raw corpus is approximately 862 MB. After preprocessing (see Appendix A), it yields 7,698,626 sentences accounting for 74,983,189 words.

Data access. The underlying transcripts are licensed from FMP and cannot be publicly redistributed. The derived artifacts needed to run the replication scripts — the trained model and the two network files — are deposited in a restricted-access archive for peer review (anonymized view-only link: https://osf.io/zq2x4/?view_only=c79866358d7f4492b7997ffcc1bce6be) and will be made publicly available, to the extent FMP's terms permit, on publication. Researchers who wish to reconstruct the corpus can obtain the transcripts directly through an active FMP "Ultimate"-tier subscription (https://site.financialmodelingprep.com/pricing-plans) and apply the preprocessing in Appendix A; because the embedding step is not deterministic (see Reproducibility), a corpus rebuilt this way reproduces the explicit usage network exactly but yields latent-network values that differ slightly from those reported, which is why the trained model is shipped as the canonical reference. The Appendix D bootstrap additionally draws on the processed corpus, which is not redistributed; its precomputed output is included in the archive and the script is provided for audit.

Figures. The reported figures visualize data the replication scripts reproduce (the focal terms' ego-network node and edge sets, and the bootstrap distributions). The published renderings are produced separately for presentation; force-directed layouts are not bit-reproducible across software versions.