

From Text to Structure: Cross-Document Narrative Analysis Using AI-Assisted Semantic Signal Extraction

[Narrovue.com](https://narrovue.com)

Abstract

We introduce **Semantic Signal Analysis (SSA)**, a computational framework for extracting, linking, and analyzing claims across heterogeneous text corpora. Unlike conventional natural language processing (NLP) approaches that operate at the document or sentence level, SSA models discourse as a **network of claims**, enabling structural analysis of narratives across documents.

Applying SSA to a corpus with substantial document-size imbalance (~10:1), we demonstrate that raw structural metrics are dominated by document volume. After normalization, however, structurally dense regions emerge within smaller documents, revealing that **narrative importance is not proportional to size**. Network analysis further shows strong intra-document cohesion and limited cross-document integration, suggesting that narrative alignment governs structural connectivity.

These results position SSA as both a methodological contribution and a practical tool for analyzing narrative systems in policy, media, and institutional contexts.

1. Introduction

Text analysis has historically focused on classification, similarity, and information retrieval. While effective, these approaches overlook a critical dimension of discourse:

Structure.

Narratives are not simply collections of sentences—they are **systems of interrelated claims**. Understanding how these claims connect, reinforce, or contradict each other is essential for analyzing policy documents, institutional messaging, and contested public narratives.

This paper introduces **Semantic Signal Analysis (SSA)**, a framework designed to:

- Extract claims as atomic analytical units
- Identify relationships between claims
- Represent discourse as a **networked structure**

We apply SSA to a cross-document corpus characterized by substantial size imbalance and ideological variation. This setting allows us to examine a key methodological challenge:

How do we compare narrative structure across unequal texts?

2. Related Work

2.1 Argument Mining

Research in argument mining has focused on identifying claims and their relationships within documents. Foundational work by Christian Stab and Iryna Gurevych demonstrated supervised extraction of argumentative components.

However, most approaches:

- Operate within single documents
- Assume hierarchical argument structures

SSA extends this paradigm by modeling **cross-document, non-hierarchical relationships**.

2.2 Semantic Representation

Embedding-based models such as Sentence-BERT and SimCSE enable scalable semantic similarity comparisons.

These methods are typically used for:

- Retrieval
- Clustering

SSA repurposes them for:

Graph construction at the claim level

2.3 Network Analysis

Work in network science—including contributions by Mark Newman and Albert-László Barabási—has shown that real-world systems exhibit skewed degree distributions and structural inequality.

SSA applies these principles to discourse by:

- Treating claims as nodes
- Measuring structural centrality

2.4 Computational Social Science

Research in computational social science and misinformation research has focused on diffusion and virality.

SSA differs by focusing on:

Internal narrative architecture rather than spread dynamics

3. Methodology

3.1 Unit of Analysis

- Texts are decomposed into **atomic claims**
- Each claim is assigned a unique identifier

3.2 Relationship Extraction

Pairs of claims are evaluated using semantic similarity and rhetorical criteria.

Relationships are defined when similarity exceeds a threshold or when an LLM identifies a meaningful connection (e.g., support, elaboration, contradiction).

3.3 Network Construction

- Nodes: Claims
- Edges: Semantic or rhetorical relationships
- Graph: Undirected network

3.4 Normalization

To correct for document-size imbalance:

Normalized Degree = Degree / Total Claims in Source Document

This allows comparison of **relative structural importance** across documents.

3.5 Cross-Document Connectivity

For each node, we compute:

- Number of distinct source documents among neighbors

This measures **narrative integration across texts**.

4. Data

The corpus consists of two primary documents:

Document	Claims
Document A	~15,055
Document B	~1,418

This ~10:1 imbalance provides a test case for evaluating normalization.

5. Results

5.1 Network Structure

The resulting graph contains:

- A giant connected component (~2,700 nodes)
- Numerous small clusters

This reflects a **core-periphery narrative structure**.

5.2 Degree Inequality

The degree distribution is highly skewed:

- A small number of claims dominate connectivity
- Most claims have low degree

This aligns with known properties of complex networks.

5.3 Network Visualization

The network exhibits:

- A dense central core
- Sparse peripheral regions
- Limited bridging between clusters

5.4 Normalization Effects

Before normalization:

- Centrality is dominated by Document A

After normalization:

- High-density claims emerge in Document B

Key Finding: Structural influence is not proportional to document size.

5.5 Cross-Document Connectivity

- Most edges occur within documents
- Cross-document links are sparse

This suggests:

Narratives remain structurally bounded

6. Claim-Level Analysis

Examples of high-centrality claims:

C15168

“The DOJ should not engage in criminal prosecution ...”

C15163

“The President should direct the DOJ ...”

C15165

“It is fundamentally unfair for the DOJ to prosecute ...”

C9958

“The DOJ should work with all federal agencies ...”

C15164

“DOJ should consult and consider all official ...”

Interpretation

These claims:

- Function as **policy directives**
- Anchor surrounding claims
- Reinforce institutional authority

7. Discussion

7.1 Structural Inequality

Narrative systems exhibit strong inequality:

- Few claims organize discourse
- Many claims remain peripheral

7.2 Importance of Normalization

Without normalization:

- Larger documents dominate analysis

With normalization:

- True structural density becomes visible

7.3 Narrative Alignment

Observed pattern:

- Strong intra-document cohesion
- Weak cross-document integration

We propose:

Narrative alignment governs structural connectivity across documents

8. Limitations

- Dependence on AI-assisted extraction
- Sensitivity to similarity thresholds
- Limited control over ideological alignment

9. Conclusion

Semantic Signal Analysis enables:

- Claim-level structural modeling
- Detection of narrative concentration
- Measurement of cross-document divergence

This work demonstrates that narratives:

- Are structurally unequal
- Are internally coherent
- Vary in integration based on alignment

10. References

- Christian Stab & Iryna Gurevych (2017)
- Nils Reimers & Iryna Gurevych (2019)
- Tianyu Gao et al. (2021)
- Mark Newman (2010)
- Albert-László Barabási (2009)
- Michael Nielsen (2012)