SoftwareX 32 (2025) 102377

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original software publication

Narrative Maps Visualization Tool (NMVT): An interactive narrative analytics system based on the narrative maps framework

Brian Keith Norambuena

Universidad Católica del Norte, Av. Angamos 0610, Antofagasta, Chile

ARTICLE INFO

Keywords:
Narrative extraction
Visual analytics
Sensemaking
Explainable AI

ABSTRACT

The **Narrative Maps Visualization Tool** (NMVT) is an interactive visual analytics system designed to help analysts understand complex narratives from collections of text documents. NMVT leverages graph-based representations to extract and visualize coherent storylines, showing how events connect over time. The system integrates advanced features including document clustering, coherence-based optimization, storyline extraction, and explainable AI components that provide interpretable insights into narrative connections. NMVT supports both directed analysis (connecting specific events) and exploratory analysis (discovering emerging storylines). By enabling analysts to make sense of large document collections, NMVT addresses critical challenges in intelligence analysis, computational journalism, and misinformation research, allowing users to effectively *connect the dots* between seemingly unrelated events. The system has been successfully demonstrated on news data by extracting coherent narrative structures that capture both main storylines and alternative perspectives. Case studies show that NMVT's semantic interaction capabilities enable analysts to refine narratives based on domain expertise, while the explainable AI components increase trust in the system's outputs.

Code metadata

Current code version	v2.1.0
Permanent link to code/repository used for this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-25-00326
Permanent link to Reproducible Capsule	–
Legal Code License	MIT License
Code versioning system used	Git (GitHub)
Software code languages, tools, and services used	Python
Compilation requirements, operating environments & dependencies	Python 3.12, Dash, Flask for more details please see repository.
If available Link to developer documentation/manual	Readme: https://github.com/briankeithn/narrative-maps/blob/main/README.md Tutorial: https://github.com/briankeithn/narrative-maps/blob/main/tutorial/TUTORIAL.md
Support email for questions	brian.keith@ucn.cl

1. Motivation and significance

The information landscape has become increasingly complex in the last decades. In this context, analysts face significant challenges managing and making sense of massive document collections. The exponential growth of the World Wide Web and improved accessibility of information has exacerbated the problem of information overload in our daily lives [1]. This becomes particularly problematic when trying to process and understand the constant barrage of news events [2]. Furthermore, the rise of social media has created an environment where alternative narratives and misinformation can flourish [3], making it increasingly difficult to identify factual connections and underlying patterns.

Narratives – systems of stories interrelated with coherent themes [4] – provide a natural way to capture relationships between sequences of events, as well as the goals, motivations, and plans of actors [5]. They are fundamental to our understanding of the world [6] and

E-mail address: brian.keith@ucn.cl.

https://doi.org/10.1016/j.softx.2025.102377
Received 17 May 2025; Received in revised form 6 August 2025; Accepted 17 September 2025
Available online 25 September 2025

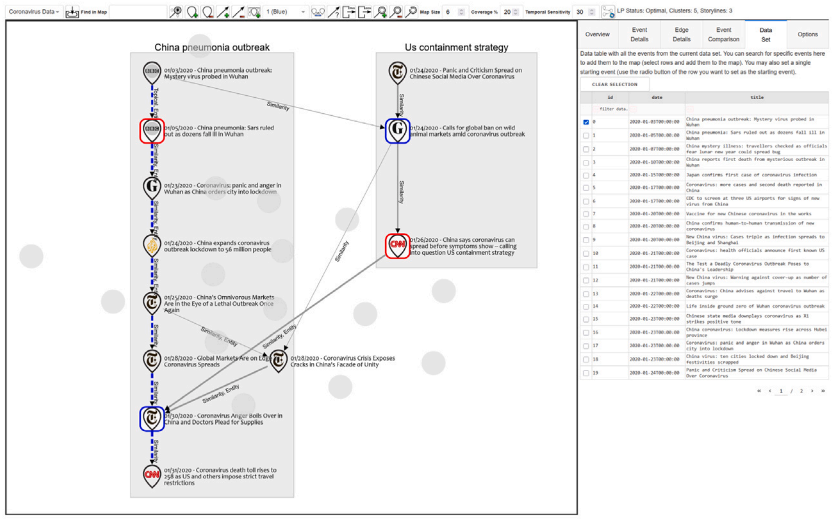


Fig. 1. Interface of the NMVT system with an instanced Narrative Map extracted from the COVID-19 data set.

play a key role in collaborative sensemaking in society [7]. However, modern online information systems rarely incorporate narratives in their representation of events happening over time, instead relying on chronological listings or simplistic clustering approaches.

The **Narrative Maps Visualization Tool** (NMVT) addresses these challenges by providing a computational framework to represent narratives as graph structures and extract them from large collections of text documents. NMVT helps analysts in several areas:

- **Narrative sensemaking**: NMVT enables analysts to understand the big picture of complex narratives by visualizing relationships between events and storylines, helping them to *connect the dots* between seemingly unrelated pieces of information [8].
- **Information exploration**: The system provides both directed analysis capabilities (tracing connections between specific events) and exploratory functions (discovering emerging storylines and outcomes) [9].
- **Misinformation modeling**: By representing competing narratives and their connections, NMVT can help identify how misinformation spreads and competes with factual reporting [9,10].
- **Intelligence analysis**: The system supports structured analytical techniques needed for intelligence work by providing a framework for organizing information and identifying causal relationships [11,12].

NMVT leverages the narrative maps framework [9], which has several advantages over existing timeline-based approaches [13,14] and tree-based representations [15,16] by implementing a directed acyclic graph (DAG) structure that can model both divergent and convergent storylines [9]. This enables a more flexible and expressive representation of complex narratives that evolve over time.

The software uses an optimization approach that maximizes coherence – how much sense a storyline makes – subject to structural and topic coverage constraints [9]. This ensures that the extracted narrative maps capture the most relevant connections while providing comprehensive coverage of the important topics. Additionally, NMVT integrates explainable AI components that help users understand why certain events are connected and how the overall narrative structure was determined [17].

In experimental settings, users interact with NMVT through a web-based interface that displays the narrative map alongside additional information panels. Users can manipulate the map through semantic interactions [12], such as adding or removing events, highlighting connections, and clustering related events. These interactions influence the underlying extraction model, allowing for an iterative refinement process that aligns with the incremental formalism observed in sensemaking tasks [18].

NMVT builds upon prior work in visual analytics for sensemaking [19], narrative visualization [20], and computational approaches to narrative extraction [10]. NMVT integrates these areas into a single system that helps in narrative understanding and provides a foundation for future research in computational journalism, intelligence analysis, and misinformation studies.

2. Software description

The Narrative Maps Visualization Tool is an interactive visual analytics system designed to help analysts explore, understand, and refine narrative structures extracted from text documents. Built upon the theoretical foundations of narrative maps [9], the tool enables users to visualize connections between events, identify storylines, and analyze the overall narrative structure while supporting human-AI collaboration through semantic interactions [12] and explainable AI components [17]. NMVT provides an intuitive interface for narrative map visualization and interaction. Fig. 1 shows the main interface of NMVT with its key components labeled.

2.1. Core architecture and functionality

NMVT follows a modular architecture implemented primarily in Python, integrating multiple components that work together to support the narrative sensemaking process. The system architecture, illustrated in Fig. 2, demonstrates how these components interact through a bidirectional pipeline paradigm where user interactions inform the underlying models, which in turn generate updated visualizations.

The *Data Processing Module* handles various input formats including CSV files with required columns comprising document identifiers, titles, URLs, dates, publication sources, full text content, and embeddings. The system uses sentence transformers, specifically the all-MiniLM-L6-v2 model, for document embedding generation and supports custom data sets through a straightforward import process. This module uses pandas for efficient data manipulation and preprocessing, ensuring that documents are properly formatted and temporally ordered for narrative extraction.

The *Narrative Extraction Engine* implements the core algorithm that transforms document collections into coherent narrative structures.

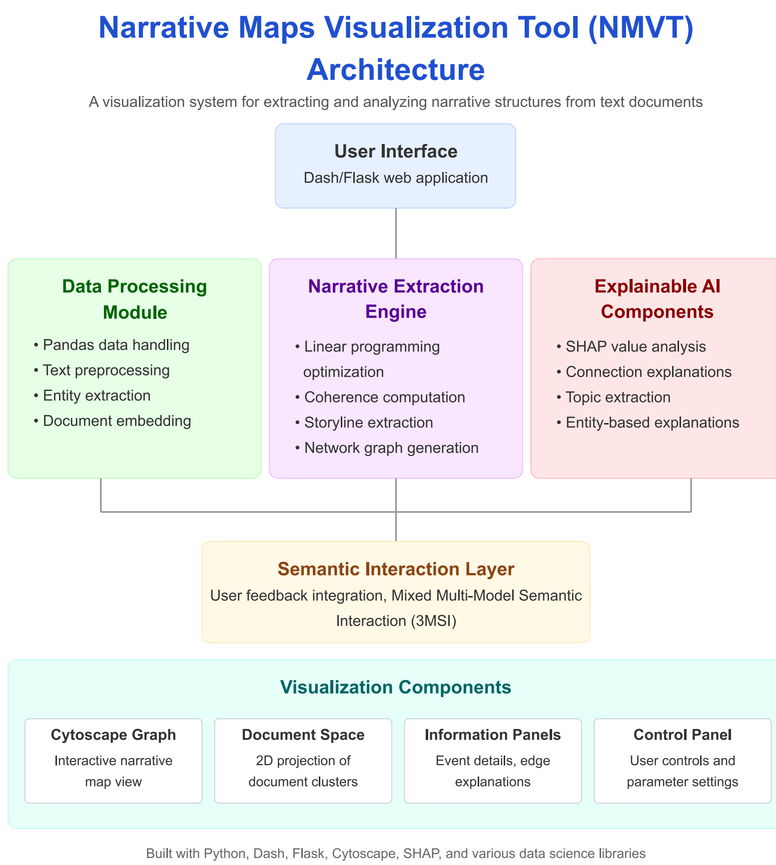


Fig. 2. System architecture diagram showing the relationships between the five main components of the NMVT system.

This algorithm generates narrative maps from text data using a multi-step process [9], transforming document collections into coherent narrative structures. In particular, this process involves document clustering using UMAP for dimensionality reduction combined with HDBSCAN for identifying topical groups. Then, it computes coherence scores based on semantic similarity, topical relationships, and entity co-occurrence patterns. The optimization process employs linear programming through PuLP to construct optimal narrative structures that maximize overall coherence while satisfying structural constraints. The engine automatically detects main storylines that represent the most coherent narrative path, identifies side stories that provide alternative perspectives, and isolates singleton events that stand alone in the narrative landscape.

The *Interactive Visualization Interface* provides the primary means for users to explore and understand narrative maps. Built with Dash and Dash Cytoscape, this component renders narrative structures as directed acyclic graphs where events appear as nodes and their relationships as edges. The interface presents temporal flow from top to bottom, allowing users to trace how narratives evolve over time. It supports intuitive interactions including zooming, panning, and element selection, while providing details-on-demand through information panels that reveal event content, edge explanations, and data set overviews. Users can export their findings in PNG format for visual documentation or JSON format for further computational analysis.

The *Semantic Interaction Layer* implements the Mixed Multi-Model Semantic Interaction (3MSI) pipeline [12], enabling analysts to refine narrative structures based on their domain expertise. This layer supports event manipulation through adding or removing nodes, connection modification by creating or deleting edges, and cluster assignment to group related events. Each interaction triggers real-time map regeneration that reflects the user's analytical insights, creating an iterative refinement process that aligns with natural sensemaking workflows.

The *Explainable AI Components* provide transparency into the system's decision-making processes by generating explanations at multiple levels [17]. These components generate connection explanations that reveal why events are linked, whether through similarity, topical relationships, or shared entities. They employ SHAP (SHapley Additive exPlanations) values to analyze keyword contributions, helping users understand which terms drive narrative connections. The system also provides event comparison functionality that explains coherence calculations between any pair of events, and generates storyline names based on content analysis to provide meaningful labels for narrative threads.

2.2. Design for non-technical users

The NMVT system incorporates complex algorithms for narrative extraction and analysis. Thus, efforts have been made to ensure accessibility for non-technical users through intuitive interfaces and abstractions validated through pilot testing with users. Rather than exposing users to the underlying mathematical optimization constraints, the system presents the algorithm's parameters in terms of their practical effects. The **map size** parameter K is described as the "expected main storyline length", helping users understand they are controlling how many events should appear in the main story of the narrative. It should be noted that statistically, the underlying linear program should produce main stories with this average length [2]. Furthermore, the **coverage** parameter (mincover) is presented as a percentage value corresponding to the "proportion of topics covered in the map", avoiding technical discussions of the actual topical cluster coverage constraints. The **temporal sensitivity** parameter, when needed, is expressed in days rather than mathematical decay functions, making it intuitive for users working with time-based data. The tutorial provides recommended starting values for different data set types, eliminating guesswork and allowing users to achieve good results without understanding the underlying algorithms.

In general, the narrative maps framework is based on visual metaphors and indicators that seek to transform algorithmic outputs into easily interpretable visual elements [9]. Representative events appear with red borders and are described as “events that best summarize their storyline”, avoiding technical explanations of centroid calculations in embedding space. Structurally important events show blue borders and are presented as “connection hubs in the narrative map”, making their role immediately clear without discussing node degree calculations. Gray background nodes are explained as “related events that didn’t make the final cut”, helping users understand these represent similar but less coherent alternatives without explaining the underlying optimization procedure that selects the nodes. Connection labels use plain language – Similarity, Topical, or Entity-based – rather than exposing similarity scores or algorithmic classifications.

Moreover, the system employs progressive disclosure to hide complexity while maintaining power for advanced users. Basic event details show only essential information like headlines and dates, with full text and metadata available on demand. The “Explain Edge” button reveals detailed connection analysis only when users seek deeper understanding, presenting SHAP value visualizations as simple bar charts labeled “keyword importance” rather than feature attribution scores.

The interface follows a guided workflow that matches natural analytical progression. Users begin by loading data through a simple dropdown and button interface, then set basic parameters using the intuitive controls described above. After generating an initial map, they can explore the visualization through familiar interactions like clicking for details or dragging to rearrange. The refinement phase uses semantic interactions that mirror how analysts naturally work with information [12]—grouping related items, removing irrelevant content, and drawing connections between events. Finally, users can export their refined narratives in formats suitable for their needs. This workflow requires no technical knowledge while supporting sophisticated analysis.

2.3. Computational performance and scalability

NMVT has been tested in two primary environments to understand its performance characteristics. The main development and testing occurred on a local machine equipped with an Intel i7 10th generation processor, 32 GB of RAM, and an NVIDIA RTX 2060 GPU with 6 GB VRAM. These specifications represent the recommended configuration for optimal performance. Additionally, the system was temporarily deployed on PythonAnywhere for user testing and demonstrations, using a more modest configuration with 5000 CPU seconds per day, 2 web workers, and 9 GB of disk space. We note the cloud deployment of NMVT has been discontinued at the time of publication. While the cloud deployment exhibited reduced performance compared to the local environment, it remained functional for demonstration purposes, illustrating the system’s adaptability to different computational resources. It is important to note that both deployments serve research and demonstration purposes rather than production use.

Under these testing conditions, NMVT successfully handled data sets containing up to 500 documents while maintaining reasonable interactivity. Performance varies with data set size: smaller data sets of fewer than 100 documents process in just a few seconds, providing near-instantaneous results. The first execution with any new data set requires approximately 3 min for medium-sized collections, as the system computes similarity matrices and extracts entity information. These computations are cached for subsequent use, reducing future map generation times to under 30 s for most operations. Complex semantic interactions that significantly alter the narrative structure may trigger more extensive recomputation, temporarily extending processing times.

The system can theoretically process up to 1000 documents, though performance degrades substantially beyond the 500-document threshold, making interactive exploration increasingly challenging. This limitation stems from the computational complexity of the underlying algorithms, particularly the pairwise similarity calculations and linear programming optimization, combined with the research-oriented nature of the implementation that prioritizes analytical flexibility over computational efficiency.

As a research prototype, NMVT’s performance may vary based on data set properties and interaction complexity. The system performs optimally with focused document collections of a few hundred items, aligning with its design goal of supporting detailed narrative analysis rather than large-scale document processing. Users planning to work with NMVT should consider these performance characteristics when preparing their analyses.

3. Illustrative examples

This section provides a detailed walkthrough of NMVT’s capabilities using the COVID-19 news dataset [9], demonstrating how analysts can extract, visualize, and refine narrative structures. We follow a complete analytical workflow from initial map generation through iterative refinement, explaining each interface element and interaction in detail. For readers seeking hands-on experience, the repository includes a detailed tutorial with step-by-step instructions for reproducing these examples.

3.1. Interface and basic functionality

Before generating a narrative map, it is essential to understand how NMVT visualizes narrative structures and what each visual element represents. Fig. 1 shows the main interface with a generated narrative map. The visualization employs several distinct visual encodings to convey different aspects of the narrative structure.

Event nodes appear as circular elements representing individual news articles or documents. These nodes are organized into **storylines**, which are visually enclosed in gray rectangular boxes. Each storyline represents a coherent sequence of related events that the algorithm has identified as forming a meaningful narrative thread. The **main storyline** stands out through its blue-colored connections, representing the most coherent path through the entire narrative according to the optimization algorithm. Some events exist as **singleton storylines**—isolated nodes without gray boxes that represent events important enough to include but not strongly connected to other narrative threads.

Another important element in the interface is the presence of **background events**, shown as gray nodes scattered around the main visualization. These represent documents from the dataset that are semantically similar to events in the extracted narrative but were not included in the final map. The optimization algorithm determined that including these events would reduce the overall coherence of the narrative structure, or they may have been filtered out during post-processing due to weak connections. However, NMVT retains these as gray background nodes for several reasons: they provide context about what information exists in the dataset beyond the main narrative, they allow analysts to understand what the algorithm considered but ultimately excluded, and they remain interactive—analysts can click on them to view details or even add them to the map through semantic interactions if their domain expertise suggests they should be included. By default, for each event in the narrative map, NMVT displays the two most similar excluded events as background nodes, though this number can be adjusted in the system options.

The system highlights **important events** through colored borders. Events with red borders are **representative events**, computed as the centroid of their storyline in the embedding space, summarizing the content of that narrative thread. Events with blue borders are **structurally important**, identified by their high degree of connectivity within the network, serving as narrative hubs that bridge multiple story elements. **Edge labels** appear on connections between events, indicating whether the relationship is based on semantic similarity, topical clustering, or shared entities, providing transparency about why the algorithm connected these particular events. In particular, the displayed labels indicate the relationship type–*Similarity*, *Topical*, or *Entity*-based–providing immediate insight into why events are linked.

The interface includes a right panel with multiple tabs providing additional information about events, edges, and the entire data set. The top menu offers options for manipulating the map, including tools for adding or removing events, creating connections, and adjusting extraction parameters. This design supports the narrative sensemaking process by combining visualization with detailed contextual information.

3.2. Creating and analyzing a narrative map

To demonstrate NMVT's analytical process, we walk through generating a narrative map from the COVID-19 dataset covering January 2020, which corresponds to the period when the virus emerged and began spreading globally. The generation process begins with parameter configuration. We set $K = 6$ to indicate we expect approximately six events in the main storyline, providing a concise but complete narrative. The coverage parameter $\mathsf{mincover} = 20\%$ ensures the narrative touches on at least 20% of the topical clusters identified in the dataset, balancing focus with breadth. For this demonstration, we also set structural constraints by marking the first document (early reports of a mystery virus) as the required starting point and the last document (international responses) as the required ending point, guiding the narrative to span the full temporal range of our dataset. The map generation process is initiated by clicking the Generate Map button, which triggers the extraction algorithm.

The resulting map, shown in Fig. 3, reveals multiple interconnected storylines that capture different facets of the early pandemic. The main storyline (with blue connections) traces a clear progression: from initial reports of a "mystery pneumonia outbreak" in Wuhan, through confirmation of human-to-human transmission, to the implementation of lockdowns and travel restrictions, and early mentions of vaccine development efforts. This main narrative captures what retrospectively proved to be an important sequence of events that shaped the global pandemic response. The side storylines provide essential context that enriches our understanding. One storyline focuses on economic impacts, connecting events about oil price declines, market volatility, and airlines suspending flights to China. Another captures social and political dimensions, including public anger over the handling of the outbreak, criticism on Chinese social media, and the strain on medical supplies. A third thread examines the Chinese government's response and its effect on public trust. These alternative narratives demonstrate how NMVT captures not just the primary sequence of events but also parallel developments that provide a more complete picture of this complex global event.

Notably, several singleton storylines appear on the map, which represent isolated nodes that the algorithm deemed important enough to include despite lacking strong narrative connections. For instance, an article about "omnivorous markets" appears as context about potential disease origins but does not flow naturally into other narrative threads. The gray background nodes surrounding the main visualization represent the broader information landscape, which are articles about similar topics that did not meet the coherence threshold for inclusion but remain available for exploration. An analyst might click on these to understand what related information exists in the dataset or to potentially incorporate overlooked events through semantic interaction.

The map visualization illustrates how the virus transformed from a "mystery virus" to a global threat, capturing both health and socioeconomic dimensions of the early pandemic narrative.

3.3. Semantic interaction for map refinement

While the initial map provides valuable insights, analysts often need to refine the narrative based on their specific analytical goals or domain expertise. NMVT's semantic interaction capabilities enable iterative refinement of the narrative structure. We demonstrate two contrasting approaches to focusing the narrative on economic impacts of the early pandemic. Fig. 4 illustrates these approaches that transform the initial map into a version that better focuses on specific narrative aspects.

Approach 1: Clustering and Cleaning. In the first approach (Fig. 4a), the analyst begins by removing events that add noise without advancing the narrative. Using shift-click to select multiple nodes, they remove three events: articles about omnivorous markets (tangential to the main narrative), a personal quarantine diary (too specific for a macro-level analysis), and other peripheral content. The analyst then identifies that three events about economic impacts – oil price declines, Saudi Arabia's market interventions, and airline suspensions – are scattered across different storylines despite their thematic similarity. Using the "Add to Cluster" function, they assign all three events to the same topical cluster (cluster 1, indicated by light blue). This clustering operation has two effects: it modifications the underlying embedding space using semi-supervised dimensionality reduction to ensure these events are positioned near each other, and it adds weak connectivity constraints to the optimization problem to encourage their connection in the narrative structure.

After regenerating the map with these interactions, the system reorganizes the narrative. In particular, the main storyline now focuses on "Coronavirus anger", capturing social responses to the pandemic including lockdowns, supply shortages, and public frustration. The economic events have coalesced into a coherent side storyline that traces how virus fears led to oil price declines, prompted Saudi intervention, and resulted in airline suspensions—a clear causal chain that was fragmented in the original map.

Approach 2: Direct Connection Manipulation. The second approach (Fig. 4b) demonstrates more surgical modifications to the narrative structure. After removing the same irrelevant events, the analyst directly creates connections between related economic events using the "Add Edge" function. They connect the oil prices event to the Saudi Arabia intervention, making explicit the causal relationship that the algorithm initially missed. The analyst then searches for additional economic content using the "Find in Map" function with the query "market*" (using wildcards to match "market", "markets", etc.). From the highlighted results, they add an event about "markets on edge" and connect it to the existing economic narrative. They also browse the dataset table to find and add an article about Asian markets closing with losses.

This approach includes a subtle technique for handling irrelevant content. Rather than directly removing an event about "life inside ground zero", the analyst removes its edges using shift-click selection. This constraint forces the algorithm to either find new, more meaningful connections for this event or naturally exclude it if no coherent connections exist. The resulting map shows stronger economic storylines while maintaining the algorithm's ability to discover unexpected connections.

3.4. Other features

NMVT's explainable AI components provide transparency into how the system determines narrative connections. When analysts click on any edge and use the **Explain Edge** functionality, the system reveals the computational reasoning behind that connection. Fig. 5 shows an example explanation for a connection between two events. The visualization uses SHAP values to show which keywords contributed most strongly to the connection. In this example, shared terms related to the outbreak, geographic locations, and temporal markers drive the similarity score. The explanation distinguishes between different

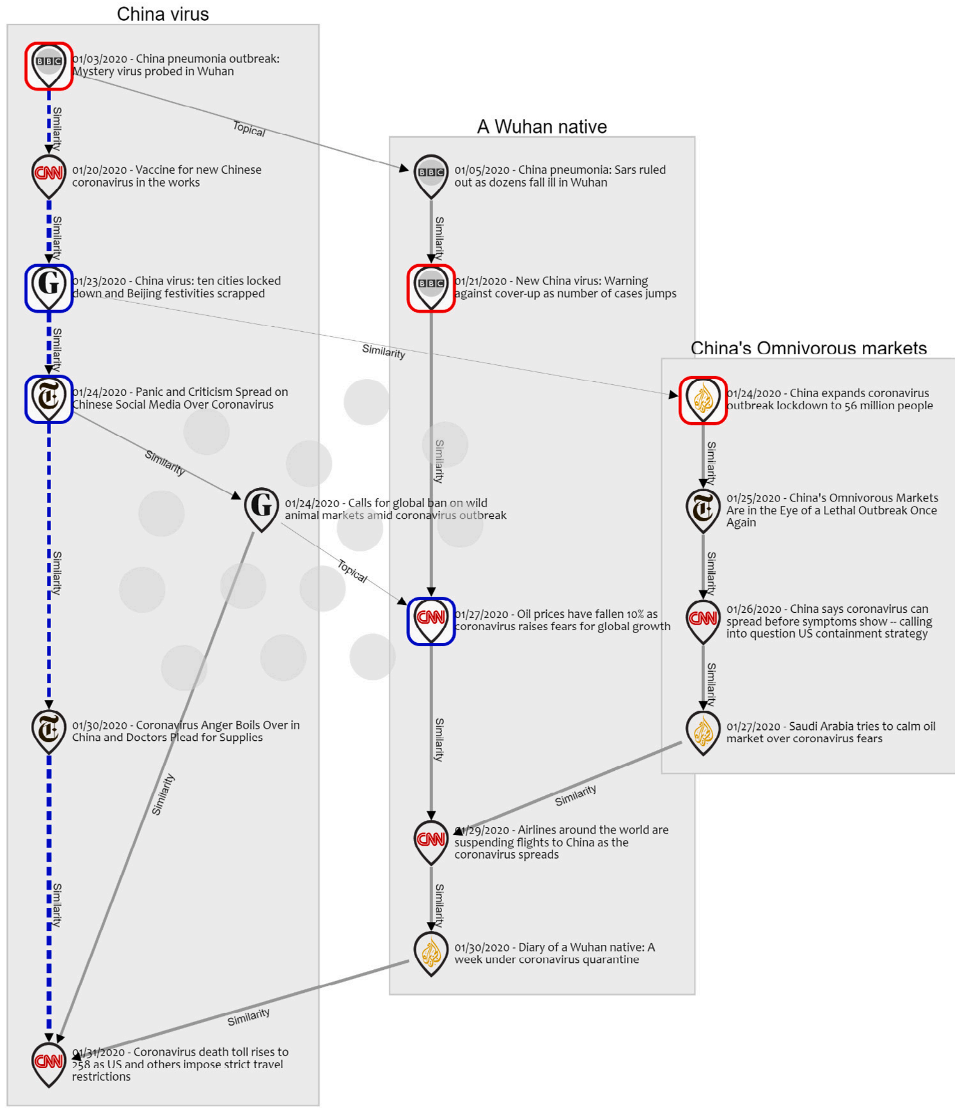
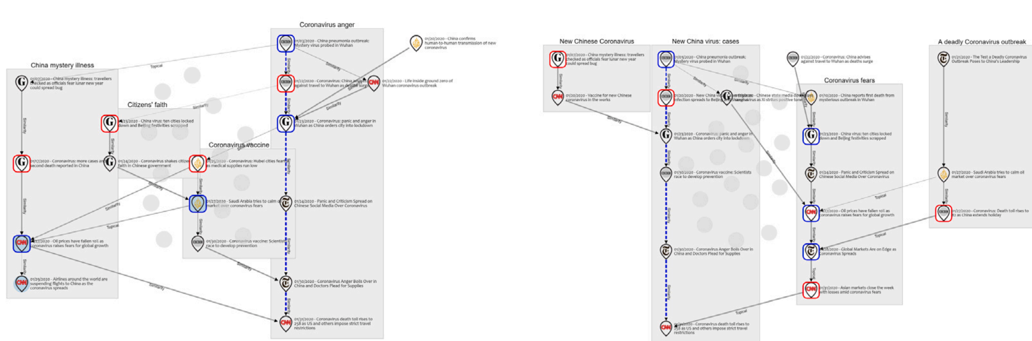


Fig. 3. Initial COVID-19 narrative map from January 2020. The main storyline (blue) shows progression from mystery virus to global response. Side storylines capture economic impacts (oil prices, markets), social effects (anger, lockdowns), and political dimensions. Gray background nodes show related articles not included in the extracted narrative, while singleton storylines provide additional context without strong narrative connections.



(a) Map refined using Remove Event and Add to Cluster interactions

(b) Map refined using Add Event, Add/Remove Edge interactions

Fig. 4. Semantic interaction examples showing two approaches to refining a narrative map. (a) Using Remove Event to eliminate irrelevant content and Add to Cluster to group related economic events. (b) Using Add Edge to connect related events and Remove Edge to eliminate unwanted connections, highlighting economic impacts.

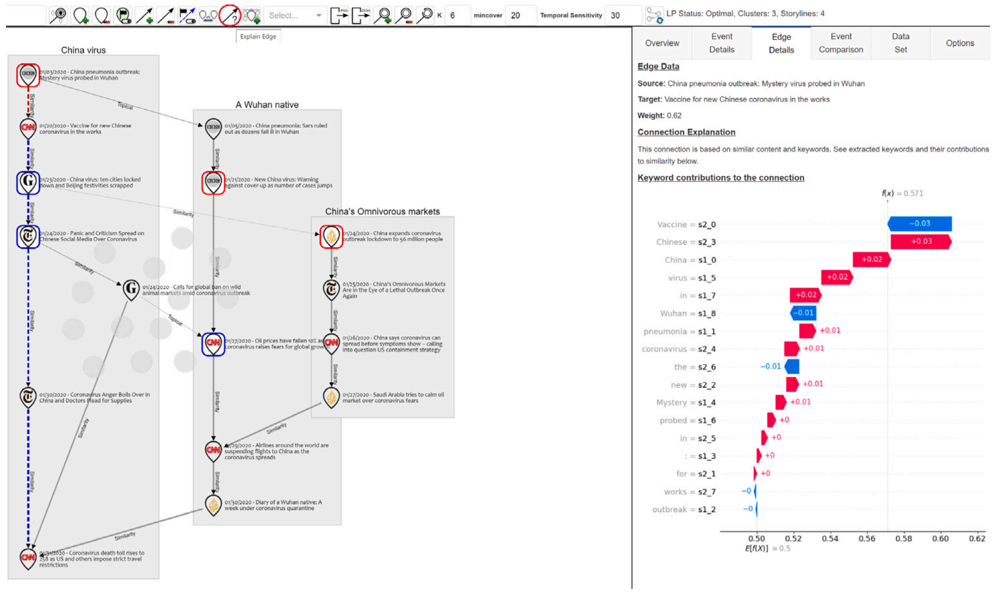


Fig. 5. The Edge Details tab showing an explanation of why two events are connected, including a visualization of keyword contributions to the connection.

connection types: similarity-based connections rely on overall semantic similarity, topical connections emerge from shared cluster membership in the embedding space, and entity-based connections identify shared named entities like people, organizations, or locations.

The system provides similar explanatory capabilities for comparing any two events, even if they are not connected. The *Event Comparison* capabilities enable users to select any two events and analyze their similarities and differences through the Compare Events function. This helps explain why certain events are connected while others remain separate, providing insight into the underlying coherence calculations that drive narrative formation. In terms of functionality, the event comparison tool works in the same way as edge explanations.

The *Overview Tab* complements these detailed explanations with a macro-level view, showing the 2D projection space with topical clusters, extracted keywords for each cluster using TF-IDF scoring, and a frequency-sorted entity list. This bird's eye view helps analysts understand the information landscape before diving into specific narratives. This bird's-eye view of the data set helps analysts identify dominant themes and important entities before diving into the detailed narrative structure.

Additional features include *Map Exporting* with options to download the map in PNG format for inclusion in reports or JSON format for further computational analysis, and the ability to toggle various visualization elements such as main storyline highlighting and important event markers. These advanced capabilities illustrate how NMVT supports the narrative sensemaking process by enabling analysts to extract, visualize, interact with, and refine narrative maps, ultimately leading to deeper understanding of complex narratives in news data.

These examples demonstrate how NMVT supports the full narrative analysis workflow, starting from initial exploration through iterative refinement to final presentation. The combination of algorithmic extraction, visual representation, semantic interaction, and explainable AI creates a practical environment for making sense of complex narratives in document collections. Readers interested in hands-on exploration can find detailed step-by-step instructions in the tutorial included with the software repository.

4. Impact

The Narrative Maps Visualization Tool has contributed to the field of narrative analytics by providing researchers and analysts with a framework for understanding complex information narratives. The impact of NMVT spans several domains and has supported research directions in narrative visualization and analysis.

The initial target audience for NMVT consisted of three primary groups: intelligence analysts for connecting dots between events in intelligence reports, computational journalists for understanding evolving information landscapes, and misinformation researchers for modeling how misinformation spreads through narrative structures [9]. While user studies have predominantly been conducted in the news domain, particularly with COVID-19 data sets, and the system was evaluated by experts from the intelligence community [12]. However, the framework has been applied to other domains such as social media analysis for studying online social movements [21], and historical document analysis for examining photographic collections [22]. Although most concrete validations remain within the news and text analysis domains, this tool is designed to be agnostic to the user's domain, as described by the initial publication of the narrative maps model [9].

As illustrated in the preceding paragraph, NMVT has opened new research avenues in multimodal narrative extraction with historical photographs. In particular, building upon the foundation established by NMVT, German et al. [22] extended the narrative mapping approach to visual media, developing a semi-supervised framework for extracting narratives from historical photographic collections. This demonstrates how NMVT's conceptual framework can transcend textual data to address the challenges of visual narrative construction. Similarly, German et al. [23] introduced Narrative Trails, which improves coherent storyline extraction through maximum capacity path optimization, directly building on NMVT's graph-based representation model.

The framework has also influenced narrative visualization research more broadly. Fan et al. [24] utilized the narrative maps framework and developed their own variation to explore narrative structure and visualization techniques in the news media, demonstrating how the approach of NMVT can be adapted and extended by other research teams. This adoption by the visualization community indicates the framework's utility beyond its original implementation.

The tool has contributed to intelligence analysis and computational journalism practices by providing an approach to narrative sensemaking tasks. As demonstrated by Keith et al. [9], NMVT enables analysts to identify important events, examine possible causal relationships, and recognize divergent storylines within complex news narratives. This capability can be valuable in scenarios where understanding evolving narratives is important. The semantic interaction capabilities implemented in NMVT [12] have enhanced the tool by allowing analysts to directly manipulate the narrative structure based on their domain expertise, creating an interactive process between the human analyst and the computational model.

NMVT has introduced a structured approach to narrative sensemaking that combines human expertise with computational methods. The design guidelines established through empirical studies with analysts [25] have provided insight into how narrative structures can be visualized and interacted with. The incorporation of explainable AI components [17] has addressed transparency concerns in the system's outputs, which is important for human-AI collaborative systems.

The framework has been applied beyond its initial target audience of intelligence analysts to include journalists, digital humanities researchers, and misinformation researchers. The NMVT framework has been adapted for specialized applications, including the analysis of social media narratives. Keith et al. [21] applied the framework to characterize narratives of social movements in online communities, specifically analyzing the 2021 Cuban protests on Reddit, demonstrating its utility to understand contemporary social phenomena. The framework has also been applied to the examination of historical photo collections [22].

NMVT represents an effort to bridge computational narrative extraction and human cognitive processes. By providing a visual, interactive, and explainable interface for narrative analysis, NMVT contributes to the development of collaborative human-AI systems to address analytical challenges [12].

5. Conclusions

The Narrative Maps Visualization Tool addresses critical challenges in narrative sensemaking by providing a unified framework for extracting, visualizing, and interacting with narrative structures from text document collections. By implementing a directed acyclic graph representation that can model both divergent and convergent storylines, NMVT offers advantages over traditional timeline and tree-based approaches. The integration of explainable AI components improve trust and transparency, while semantic interaction capabilities enable analysts to refine narrative structures based on their domain expertise. Through its application in intelligence analysis, computational journalism, and misinformation research, NMVT has demonstrated its effectiveness in helping analysts understand complex narratives and *connect the dots* between seemingly unrelated events.

Future development of NMVT will focus on improving scalability for larger data sets, expanding multimodal capabilities to incorporate visual and audio data, and refining the explainable AI components to provide even more intuitive explanations. The open source nature of the tool invites community contributions that can extend its functionality to new domains and use cases. As information overload continues to challenge effective sensemaking across disciplines, NMVT provides a valuable foundation for human-AI collaborative systems that leverage the complementary strengths of computational methods and human expertise to navigate increasingly complex information landscapes.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been supported by ANID (National Research and Development Agency of Chile) FONDECYT de Iniciación en Investigación 2025 Grant 11250039 Project. The author is also supported by Project 202311010033-VRIDT-UCN. The author also thanks his Ph.D. advisors Chris North and Tanu Mitra for their feedback and support during the early development of this system.

References

[1] Ho J, Tang R. Towards an optimal resolution to information overload: An infomediary approach. In: Proc. of the 2001 int. ACM SIGGROUP conference on supporting group work. GROUP '01, New York, NY, USA: ACM; 2001, p. 91–6.

[2] Shahaf D, Guestrin C. Connecting the dots between news articles. In: Proc. of the 16th ACM SIGKDD int. conf. on knowledge discovery and data mining. KDD '10, New York, NY, USA: ACM; 2010, p. 623–32.

[3] Nied AC, Stewart L, Spiro E, Starbird K. Alternative narratives of crisis events: Communities and social botnets engaged on social media. In: Companion of the 2017 ACM conference on computer supported cooperative work and social computing. CSCW '17 companion, New York, NY, USA: ACM; 2017, p. 263–6.

[4] Halverson J, Corman S, Goodall HL. Master narratives of Islamist extremism. 175 5th Ave., New York, NY, USA: Springer; 2011.

[5] Finlayson MA, Corman SR. The military interest in narrative. Sprache Und Datenverarb 2013;37(1–2):173–91.

[6] Abbott HP. The Cambridge introduction to narrative. One Liberty Plaza, New York, NY, USA: Cambridge University Press; 2008.

[7] Wilson T, Zhou K, Starbird K. Assembling strategic narratives: Information operations as collaborative work within an online community. Proc the ACM HCI 2018;2(CSCW):1–26.

[8] Hossain MS, Andrews C, Ramakrishnan N, North C. Helping intelligence analysts make connections. In: Workshops at the twenty-fifth AAAI conference on artificial intelligence. 2011, p. 1–11.

[9] Keith Norambuena B, Mitra T. Narrative maps: An algorithmic approach to represent and extract information narratives. In: Proc. ACM hum.-comput. interact., vol. 4, New York, NY, USA: ACM; 2020, 33 pages.

[10] Keith Norambuena BF, Mitra T, North C. A survey on event-based news narrative extraction. ACM Comput Surv 2023;55(14s). http://dx.doi.org/10.1145/3584741.

[11] Endert A, Hossain MS, Ramakrishnan N, North C, Fiaux P, Andrews C. The human is the loop: new directions for visual analytics. J Intell Inf Syst 2014;43(3):411–35.

[12] Keith Norambuena BF, Mitra T, North C. Mixed multi-model semantic interaction for graph-based narrative visualizations. In: Proceedings of the 28th international conference on intelligent user interfaces. IUI '23, New York, NY, USA: Association for Computing Machinery; 2023, p. 866–88. http://dx.doi.org/10.1145/3581641.3584076.

[13] Wang L, Cardie C, Marchetti G. Socially-informed timeline generation for complex events. In: Proc. of the 2015 conference of the North American chapter of the ACL: human language technologies. Denver, Colorado: ACL; 2015, p. 1055–65.

[14] Shahaf D, Guestrin C. Connecting two (or less) dots: Discovering structure in news articles. ACM Trans Knowl Discov from Data (TKDD) 2012;5(4):1–31.

[15] Ansah J, Liu L, Kang W, Kwashie S, Li J, Li J. A graph is worth a thousand words: Telling event stories using timeline summarization graphs. In: The world wide web conference. WWW '19, New York, NY, USA: ACM; 2019, p. 2565–71.

[16] Liu B, Niu D, Lai K, Kong L, Xu Y. Growing story forest online from massive breaking news. In: Proc. of the 2017 ACM on conference on information and knowledge management. CIKM '17, New York, NY, USA: ACM; 2017, p. 777–85.

[17] Keith B, German F, Krokos E, Joseph S, North C. Explainable AI components for narrative map extraction. In: Text2Story@eCIR 2025. 2025, p. 3–14.

[18] Shipman FM, Marshall CC. Formality considered harmful: Experiences, emerging themes, and directions on the use of formal representations in interactive systems. Comput Support Coop Work (CSCW) 1999;8(4):333–52.

[19] Cook KA, Thomas JJ. Illuminating the path: The research and development agenda for visual analytics. Tech. rep., Pacific Northwest National Lab.(PNNL), Richland, WA; 2005.

[20] Segel E, Heer J. Narrative visualization: Telling stories with data. IEEE Trans Vis Comput Graphics 2010;16(6):1139–48.

[21] Keith Norambuena BF, Mitra T, North C. Characterizing social movement narratives in online communities: The 2021 Cuban protests on reddit. In: Computation + journalism conference (c+j 2022). New York, NY, USA; 2022, p. 9–11, arXiv:2208.04465.

[22] German F, Keith B, Matus M, Urrutia D, Meneses C. Semi-supervised image-based narrative extraction: A case study with historical photographic records. In: European conference on information retrieval. Springer; 2025, p. 248–62.

[23] German F, Keith B, North C. Narrative trails: A method for coherent storyline extraction via maximum capacity path optimization. In: Text2Story@eCIR 2025. 2025, p. 15–27.

[24] Fan S, Ying Y, Goodwin S, Angus S, Purchase H, Dwyer T. Understanding the narrative: An exploration of narrative structure and visualisation techniques in news media. In: IEEE VIS 2023 posters. Melbourne, Australia: Monash University; 2023, URL https://ieeevis.b-cdn.net/vis_2023/posters/v-vis-posters-1093.pdf. Poster #1093.

[25] Keith Norambuena BF, Mitra T, North C. Design guidelines for narrative maps in sensemaking tasks. Inf Vis 21(3):220–45.