

Exploring Narrative Continuity, Narrative Drift, and Narrative Gaps in Social Media

Lisa Grobelscheg

CAMPUS 02

University of Applied Sciences Vienna University of Economics and Business Vienna University of Economics and Business
Graz, Austria Vienna, Austria Vienna, Austria

lisa.grobelscheg@campus02.at

Ema Kahr

Institute for Complex Networks

Vienna, Austria
ema.kahr@wu.ac.at

Mark Strembeck

Institute for Complex Networks

Vienna, Austria
mark.strembeck@wu.ac.at

Abstract—While topic models help organize large datasets into smaller, semantically coherent units, substantial manual effort is still required to describe and interpret the resulting topics and to understand how they evolve over time. We present the *Narrative Evolution Explorer* (NEEx), a framework that combines the automatic generation of interpretable narratives and the representation of narrative evolution as a network. NEEx is an analytical framework that allows researchers to trace narrative trajectories and identify points of semantic continuity, change, and drift. We contribute a graph-based representation that operationalizes four forms of narrative discontinuity and provide an empirical demonstration based on three real-world datasets to reveal how semantic change, fragmented storylines, isolated narratives, and narrative termination can be identified. Among other things, we found that engagement-only continuity overestimated how long a narrative’s underlying semantics actually persist. Thus, our study provides empirical evidence that points to a risk of overstating how stable a narrative’s meaning actually remains over time.

Index Terms—chronological networks, narratives, retrieval-augmented generation, Twitter

I. INTRODUCTION

Sociopolitical events, such as political elections, demonstrations, or social movements, are heavily discussed on social media platforms. The discourse around such events evolves rapidly as the event unfolds. People seek meaning, search for information, and interpret the event, leading to the formation of narratives. These narratives undergo transformations. Some narratives evolve, while others disappear over time. Narrative trajectories have been studied in computational social science, crisis informatics, and political communication [1], [2], [3].

Traditionally, narratives are detected through topic modeling methods, which rely on keyword distributions. Yet, keywords alone fail to capture the semantic meaning of a discussion [4]. Moreover, labeling of the topics requires substantial manual effort. This led to methods such as Narrative Topic Labels derived with Retrieval Augmented Generation (NTRL-RAG) [5], which automatically produce interpretable narrative labels by relying on retrieval-augmented generation (RAG). While NTRL-RAG generates concise, human-understandable topic labels grounded in trusted news sources, each narrative is generated independently, and relationships between the narratives over time are not explicitly modeled.

While tracing the evolution of narratives, two challenges remain open: i) a storyline may seem incoherent or fragmented to human users due to incomplete social media data or coarse-grained topic detection, and ii) current approaches primarily focus on identifying and refining coherent storylines rather than explicitly diagnosing temporal continuity, narrative drifts, and gaps.

Recently, GraphRAG approaches have begun integrating graph structures with the retrieval component of a RAG system [6]. However, most of the focus has been placed on entity graphs and knowledge graphs. Although these approaches improve document retrieval, they do not explicitly model evolving narratives and treat graphs as static knowledge structures.

In this paper, we explicitly model continuity of narratives alongside three dimensions: engagement (are users continuing to participate in a narrative?), semantic (does the content of the narrative remain semantically stable over time?), and structural (do gaps, isolated narratives, weak transitions, and termination occur?). We propose the *Narrative Evolution Explorer* (NEEx), a framework that combines topic modeling, RAG-supported label generation, and the construction of a temporal narrative graph to allow for a graph-based analysis of transitions, isolated narratives, and narrative isolation.

In this study, we address the following research questions:

- 1) RQ1: How can engagement and semantic continuity be used to characterize the evolution of narratives in online social media?
- 2) RQ2: How can graph topologies be used to characterize different patterns of narrative evolution in social media discourse?

The remainder of the paper is structured as follows. Section II summarizes related work. The research method is outlined in Section III. We present the results in Section IV and discuss them in Section V. The paper concludes with Section VI and provides directions for future research.

II. RELATED WORK

The core functionality of topic models has long been the detection of topics in large textual corpora [7], with notable methods including Latent Dirichlet Allocation and Dynamic Topic Models [8] [9]. Extended topic modeling techniques

have been developed to capture the prevalence of topics over time and identify emerging and declining topics.

In the context of social media, event detection methods face a unique challenge in identifying emerging topics from a large number of short and sparse texts that contain hashtags, abbreviations, emojis, and other linguistic features not commonly found in standard texts [10]. In this context, several studies have analyzed the temporal dynamics of online discourse, including public opinion and sentiment on Twitter during various sociopolitical events [11], [12], [13]. While studying the propagation of rumors on social media, [14] highlighted the dynamic nature of social media narratives. However, most existing approaches treat discourse as a sequence of independent topics associated with discrete time intervals.

Another line of studies applied event extraction methods to identify structured representations of actions (who did what to whom) [15], or presented as a sequence of events that together form a storyline [16]. These approaches predominantly rely on clustering, temporal ordering, and semantic similarity. Yet these studies offer no explicit validation of narrative coherence and do not allow for the analysis of evolving or contradictory storylines. Shahaf, Dafna and Guestrin proposed an algorithm based on semantic similarity and optimization criteria such as coherence, diversity, and brevity to find an optimal path for building a coherent narrative chain [17]. Pradhan et al. built a narrative graph that captures semantic relationships between tweets and detected chain-of-events clusters to study evolving variations of the event [18]. Though similar in intent to our framework, these studies do not explicitly model evolving narrative graphs, neither detect structural inconsistencies, or iteratively refine storylines through additional evidence retrieval.

Large Language Models (LLMs) have been shown to achieve comparable quality judgments to humans while assessing LLM-generated texts [19]. Building on this capability, recent work has explored using LLMs not only for evaluation but also for iterative self-refinement [20] or self-correction during generation [21]. For example, [20] implemented a mechanism in which an LLM iteratively critiques and refines its own output without external supervision. Having a complementary direction, [21] implemented a mechanism in which LLMs reflect on feedback to their task, store their own reflective texts in an episodic memory buffer, and use these to improve performance in subsequent trials. Zhang and Long introduced an LLM-based method to identify gaps in narratives and generate bridging sentences to match the story’s emotional and logical flow [22].

Recent studies combine LLMs with graphs for information retrieval and reasoning [6]. These approaches construct graphs via entity and relation extraction or LLM-assisted structuring. For instance, [23] constructed a weighted graph over keywords and class labels to support retrieval in dynamic text classification tasks. Graph-based traversal is then used to identify and prioritize relevant contextual information for each input. These approaches focus primarily on pre-constructed graphs and do not support iterative narrative refinement or the correction of missing temporal links in evolving storylines. Closest to

our work is [24], which used RAG and graph structural entropy for unsupervised social event discovery to trace the evolution of events from social media streams. However, this approach focuses on event evolution and does not model or trace narratives.

TopicTracker [25] addresses a similar problem to NEEEx, introducing a platform for topic trajectory identification and visualization. Topic trajectory is described as “*a main stream or an evolutionary pathway of topics over time*”. The approach differs from NEEEx as evaluation is not based on user engagement over time but on tree-based structures. Furthermore, it does not consider semantic similarity.

Moreover, [26] developed an interactive system, Narrative Maps Visualization Tool (NMVT), that extracts narratives from a collection of textual documents, iteratively refines them, and visualizes them as narrative maps. These narratives are explained by AI components that interpret the connections between narratives and generate storyline labels. Unlike NMVT, which relies on semantic similarity together with topical and entity-based relationships to construct coherent connections among narratives, NEEEx relies on semantic similarity as well as temporal and user-engagement links to diagnose narrative continuity and detect narrative drift within an evolving narrative.

III. RESEARCH METHOD

NEEEx is an analytical framework that builds upon four processing stages: 1) topic modeling, 2) narrative topic labeling (NTRLAG: Narrative Topic Labels derived with retrieval-augmented generation, [5]), 3) the construction of a chronologically ordered network (ChronCN: Chronological Chain Network, [27]), and 4) an analytical layer for examining narrative evolution through engagement and semantic continuity.

In principle, any topic model that assigns topic identifiers to each document and keywords to each topic can be used in the proposed framework. The output is forwarded to NTRLAG, which then derives Narrative Topic Labels in the form of: (i) actor, (ii) action, (iii) event, and (iv) an LLM-based description for each document cluster. NTRLAG incorporates a retrieval-augmented generation (RAG) component that retrieves relevant news articles from trusted sources to refine and validate the generated descriptions. To create a chronological representation of the narratives, we use ChronCN to build a graph that visually connects topics which are sequentially engaged with by the same users. In ChronCN, every node represents a narrative-day combination (e.g., C1D1 denotes narrative 1 on day 1, where C originates from the topic model’s cluster assignment), and each edge reflects user engagement in both adjacent nodes.

Building on these representations, NEEEx introduces an additional analytical layer that allows for the examination of narrative evolution. NEEEx builds on ChronCN networks by adding a semantic representation. For each narrative-day combination, we compute the centroid of the embeddings of the corresponding tweets and use cosine similarity between

the centroids of connected nodes to quantify semantic continuity. This procedure allows NEEEx to distinguish between engagement continuity and semantic continuity and identify four patterns of narrative evolution: weak transitions, missing transitions, isolated narratives, and storyline termination.

Figure 1 presents the pipeline and the main task of each step.

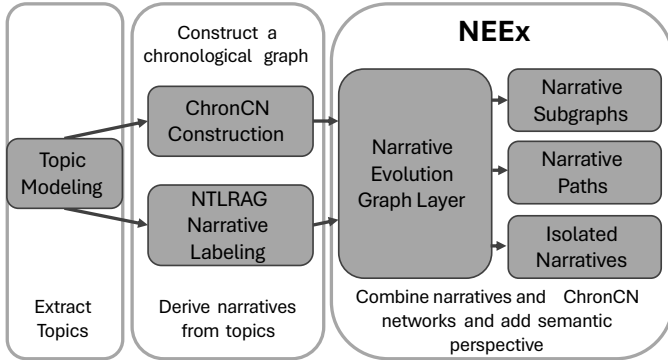


Fig. 1. A pipeline containing: Topic modeling, NTLRAG, ChronCN, and Narrative Evolution Explorer.

A. Dataset Curation

To demonstrate NEEEx, we applied it to three real-world datasets collected via Twitter’s Academic API (before its transition to X). For the purposes of this paper, we only considered unique messages to limit any amplification effects from excessive retweeting and performed basic pre-processing (removal of retweet identifiers, user mentions, special characters, URLs, hashtags, and conversion to lowercase). Table I contains basic information about the datasets used.

TABLE I
BASIC INFORMATION ABOUT DATASETS USED FOR NARRATIVE EVOLUTION EXPLORER.

Location	Extraction period	Messages (unique)	Users
Las Vegas (NV)	2-14 Oct 2017	3,436,187 (505,850)	1,394,070
Santa Fe (TX)	18-25 May 2018	967,674 (113,146)	458,644
El Paso (TX)	3-18 Aug 2019	2,307,577 (318,368)	939,940

We deliberately aim to represent similar types of events to facilitate analysis across datasets. The Las Vegas dataset covers tweets around a mass shooting at the Route 91 Harvest music festival on October 1, 2017. The Santa Fe dataset was collected after a school shooting event that happened in May 2018 in Santa Fe, Texas. The El Paso dataset refers to a shooting in a Walmart in El Paso, Texas, on August 3, 2019.

NTLRAG requires an additional corpus of news articles for each dataset in order to validate the label appropriateness (subject-verb-object), for details see [5].

B. Topic Detection

We identified topics using BertTopic [28] and obtained topic clusters, along with representative keywords and tweets for each topic. Hyperparameter tuning was carried out using the Optuna framework [29], with the resulting parameter values presented in Table II. Topic modeling identified 948 topics for the El Paso dataset, 2,335 topics for the Las Vegas dataset, and 294 for the Santa Fe dataset.

TABLE II
MINIMUM HYPERPARAMETER VALUES FOR BERTOPIC FOR EACH DATASET AS DETERMINED BY OPTUNA.

Dataset	Clustersize	Samples	Topicsize	Metric
Las Vegas	20	25	42	euclidean
Santa Fe	44	25	25	euclidean
El Paso	28	27	32	cosine

C. RAG-based Narrative Generation

NTLRAG derives a Narrative Topic Label for each topic produced by BERTopic. It uses a dual retriever setting (vector-based and BM25) to obtain relevant documents from each topic’s document collection and the corresponding news corpus. The query string consists of the BERTopic representative keywords for each topic. We restricted the retrieval space to the top $k=5$ news articles published during the interval of ± 3 days. NTLRAG extracts the Narrative Topic Labels from the extracted documents, validates them, and, if necessary, refines them again. For all generation tasks, we used Llama 3.1, available through the Ollama framework.

D. Narrative Evolution Graph Generation

Per day, each narrative is treated as a node in the Narrative Evolution Graph (NEG) (e.g., narrative 1 on day 1 is a node) and is associated with a narrative topic label and a narrative information score, which indicates the prevalence of a narrative on a certain day [27]. A directed edge from node A to node B is created if there exists at least one user who contributes to the narrative associated with node A and subsequently contributes to the narrative associated with node B on a later day.

To include cosine similarity in the graph, we first compute embeddings for the documents in each narrative-day combination. Then we use the centroid of these embeddings and add it to the edge incident to the corresponding nodes. For example, the cosine similarity between the centroids of day 1 and day 2 for a narrative might be 0.745, which is then assigned to the edge between those nodes. Note that edges are not only possible between consecutive days but also between any two nodes, as long as the graph remains directed and acyclic.

Given the directed acyclic NEG $G = (V, E)$, we define a narrative subgraph $G_n = G[V_n]$, where $V_n = \{v \in V \mid n(v) = n\}$ denotes the set of vertices assigned to narrative n . The corresponding set of edges $E_n = \{(u, v) \in E \mid u, v \in V_n\}$ contains all directed edges connecting vertices within the same narrative. NEG differentiates between engagement

TABLE III
GRAPH TOPOLOGIES, THEIR CONTINUITY ANALYSIS TYPE, AND THEIR RECOMMENDED INTERPRETATIONS FOR NARRATIVE EVOLUTION.

Name	Topology	Cont.Type	Interpretation
Weak transition	A directed edge $e = (u, v) \in E_n$ of a narrative subgraph $G_n = (V_n, E_n)$ whose cosine similarity falls below a threshold $\theta_i \in \Theta$.	semantic	The corresponding narrative exhibits potential semantic drift.
Missing transition	Narrative exists at t , is absent at $t + 1$, and reappears at $t + k$, $k > 1$	engagement	The involvement in the narrative has temporarily ended and subsequently resumes.
Isolated narrative	A vertex $v \in V_n$ of a narrative subgraph $G_n = (V_n, E_n)$ with no incident edges.	engagement	Narrative activity occurs on this day without observed user engagement on other days.
Storyline termination	Narrative exists at t , and there are no subsequent nodes for that narrative above the threshold $\theta_i \in \Theta$.	semantic	No subsequent narrative activity is observed for the remainder of the observation period.

continuity and semantic continuity. The Narrative Subgraphs, for example, are constructed based on engagement continuity, with edges representing ongoing user activity across days. Therefore, a single narrative can comprise multiple unconnected graphs (e.g., a user engages in narrative 1 on days 1, 2, and 3, and other users engage in narrative 1 on days 4 and 5). The Narrative Paths in the Narrative Subgraphs, on the other hand, reflect semantic continuity (see III-E).

E. Graph-Based Narrative Evolution Analysis

Once the initial version of the NEG has been constructed, NEEEx analyzes its narrative subgraphs to assess temporal and semantic continuity. We focus on the analysis of four distinct observations connected to the narrative subgraphs G_n : (i) weak transition, (ii) missing transition, (iii) isolated narrative, and (iv) storyline termination, as detailed in Table III.

To observe the concepts outlined in Table III, we introduce *Narrative Paths*. They represent the longest chronological sequence of narrative-day nodes within a narrative subgraph for which the cosine similarity of every consecutive transition satisfies a predefined threshold θ . To select a suitable set of thresholds for constructing the paths, we use density plots of cosine similarity values for each dataset and visually inspect their distributions. The resulting set of cosine similarity thresholds for our three datasets is $\Theta = \{0.25, 0.5, 0.75, 0.9, 0.99\}$, where $\theta \in \Theta$ represents a cosine similarity cut-off value. For example, suppose the edge C1D1→C1D2 has a cosine similarity of 0.98, but node C1D2 has no outgoing edges. At the same time, the path C1D1→C1D3→C1D5 exists with cosine similarities of 0.97 and 0.96, respectively. Assuming a threshold of 0.90, the latter path is selected because it yields the longest coherent narrative evolution and exhibits semantic similarity above the threshold, despite containing edges with slightly lower cosine similarity.

F. Comparison With Engagement-Only Narrative Tracking

In order to evaluate the contribution of the semantic component in narrative trajectory tracing via NEEEx, we compare it against a baseline that captures engagement-only narrative tracking.

To do so, we took the narrative subgraphs for each narrative across all three datasets. We then filtered all edges with a cosine similarity below a chosen $\theta \in [0.25, 0.5, 0.75, 0.9, 0.99]$ to capture weak, medium, and strong continuity. We then recorded the longest path, which reveals a daily span of how long a narrative remained semantically coherent. The created subgraphs for each narrative are then used to construct additional graphs by removing all edges (and incident nodes) below the chosen threshold. We analyze node retention across all thresholds.

IV. RESULTS

A. Narrative Evolution Graph Characteristics

Once the NEG and all corresponding Narrative Subgraphs have been constructed, we feed them into NEEEx. Table IV presents the basic properties of the created NEG without any further modifications.

TABLE IV
BASIC INFORMATION ON THE NARRATIVE EVOLUTION EXPLORERS CREATED FROM THREE DATASETS.

Dataset	Nodes	Edges
El Paso	9,025	58,379
Las Vegas	17,916	127,272
Santa Fe	1,992	10,692

B. Semantic Continuity Across Narrative Evolution Graphs

Table V presents the Narrative Evolution Graphs and their cosine similarity inspection results, such as the value associated with a percentile and the number of connected and isolated nodes. The Las Vegas dataset produced the densest graph because it is substantially larger, whereas Santa Fe has comparatively fewer nodes/edges. At equivalent percentile thresholds, the proportion of isolated vs connected nodes differs. The higher the value associated with a certain threshold, the higher the semantic continuity in the dataset. For our three cases, we found that El Paso has an overall lower semantic continuity than the other two datasets (except for the Top 90% compared to the Las Vegas dataset). One possible explanation

is the longer observation window, which provides greater temporal distance over which narrative content may change (16 days versus 8 or 12 days). The table can be read as follows: 90% of all the edges in the Narrative Evolution Graph exhibit a cosine similarity higher than 0.194 for El Paso, 0.191 for Las Vegas, and 0.230 for Santa Fe.

C. Interpreting Semantic Similarity

To assess how cosine similarity reflects semantic continuity within NEG, we inspected messages associated with edges that exhibit different similarity values. Table VI presents example messages for a certain cosine similarity value. The message pair with a cosine similarity of 0.968 (Las Vegas dataset) contains nearly identical content; i.e., such a high similarity represents a reformulation of essentially the same message. In contrast, the example from the El Paso dataset has a cosine similarity of 0.581. Even though both messages are related to shooting and share related contextual information, their content differs. The first message expresses a description of El Paso as a friendly and welcoming city, while the second discusses concerns about a school from El Paso visiting another school due to safety considerations. Our third example illustrates a substantially lower similarity of 0.207 (Santa Fe dataset). The first message concerns students anticipating graduation, whereas the second refers to a moment preceding a basketball game. Both of these narratives provide little evidence of semantic continuity.

Thus, cosine similarity should not be interpreted as a measure of whether two messages are about the same event, but rather serve as indicators in NEE_x of whether the semantic content remains stable across connected narrative-day nodes. Moreover, semantic similarity complements the temporal engagement structure of the NEG as two nodes may remain connected because users continue to engage with a narrative, while their decreasing semantic similarity can indicate that the meaning of that narrative is changing. This distinction provides the basis for identifying weak transitions and potential narrative drift.

D. Weak Transitions: Detecting Narrative Drift

Figure 2 presents the Narrative Subgraph of the C0 Narrative in the Santa Fe dataset at the cosine similarity threshold of 0.9. The Narrative Path for this configuration includes the following nodes: C0D1, C0D2, C0D3, C0D4, and C0D5. All of these edges exhibit a semantic continuity above the chosen threshold (0.9). The corresponding narrative is:

Actor: user; **Action:** offer; **Event:** prayer.

Description: Prayer without action is dead. The narrative captures not only the notion of praying for the victims, but also the idea that prayer should be accompanied by action.

The nodes on the Narrative Path reflect both ideas, e.g., C0D2 "I'm deeply angered prayers and my deepest sympathies for the victims of [name of shooter]..." or C0D3 "yes please pray then call your congressmen". In contrast, the weak transition C0D1 to C0D7 does not cover the aspect of action only: "Yesterday I took time to meditate by the ocean. I looked

down at the T-shirt I wore and prayed". This leads to the insight that narrative 0 shows engagement continuity from day 1 to day 8, but semantic continuity remains high only until day 5. By inspecting the weak transition (C0D1 to C0D7), we can either assume that the user who engaged on day 1 and day 7 might have posted off-topic, i.e., the post was falsely attributed to the Santa Fe school shooting, or the user has changed perspective on the topic.

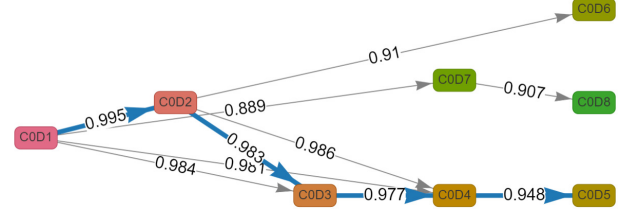


Fig. 2. A Narrative Path in the Santa Fe Dataset with 0.90 Cosine Similarity Threshold. The Path edges are highlighted in blue; values on the edges are cosine similarity values of the corresponding edge.

E. Missing Transitions: Detecting Fragmented Storylines

In the Las Vegas dataset, analysis revealed a fragmented storyline for the narrative:

Actor: people, concert organizers, country;

Action: attend, enjoy, escape;

Event: festivals, concerts, music;

Description: Music festivals are meant to be fun and enjoyable experiences for people.

Figures 3 and 4 present the Narrative Subgraphs for the narrative above, with a Narrative Path at threshold 0.9 and nodes: C6D1, C6D2, and C6D6. This topology suggests a strong semantic and engagement continuity of the narrative, but a missing transition ranging from C6D2 to C6D6. Figure 4 allows for further investigation of this missing transition and presents the second Subgraph belonging to this narrative, where a new storyline has emerged on day 3 (C6D3) and is active until day 5. Investigation into messages belonging to each subgraph indicates a change of perspective, where the first storyline primarily talks about a hypothetical visit to a concert that will not take place because of the corresponding shooting incident, whereas the second storyline (second subgraph) describes how users want to go or have already gone to concerts and events despite what has happened.

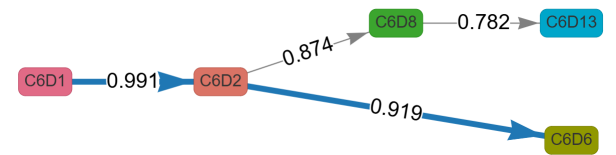


Fig. 3. The first Narrative Subgraph of Narrative C6 in the Las Vegas Dataset, with a Narrative Path at 0.9 threshold.

TABLE V
SUMMARY STATISTICS INCLUDING CONNECTED AND ISOLATED NODES OF NARRATIVE EVOLUTION GRAPHS AT DIFFERENT COSINE SIMILARITY THRESHOLDS.

Threshold	El Paso				Las Vegas				Santa Fe			
	Value	Edges	con. Nodes	iso. Nodes	Value	Edges	con. Nodes	iso. Nodes	Value	Edges	con. Nodes	iso. Nodes
Top 5%	0.686	2919	2161	6864	0.699	6364	4278	13638	0.786	535	506	1486
Top 10%	0.580	5838	2893	6132	0.628	12728	5368	12548	0.676	1070	679	1313
Top 25%	0.457	14595	4222	4803	0.527	31818	7289	10627	0.520	2673	989	1003
Top 50%	0.354	29190	5387	3638	0.415	63636	9359	8557	0.408	5346	1253	739
Top 75%	0.266	43784	6325	2700	0.293	95454	11474	6442	0.310	8019	1438	554
Top 90%	0.194	52541	6904	2121	0.191	114544	12810	5106	0.230	9622	1526	466

TABLE VI
EXAMPLE MESSAGES WITH DIFFERENT COSINE SIMILARITY VALUES.

Case	Edge	Msg.1	Msg.2	Cosine
Las Vegas	C0D7 to C0D11	Las Vegas Shooting, Proposed Regulations Alternate title:	Las Vegas shooting, proposed regulations	0.968
El Paso	C3D1 to C0D14	This is my home. This is my city. A city filled with loving, kind, friendly people. Everywhere you go complete strangers	So a school from Plano TX doesn't want a school from El Paso to go play them because of safety concerns?	0.581
Santa Fe	C1D1 to C6D8	those students were ready for the school year to be over for graduation to happen for celebrations to occur now	wow what a moment before the rockets game	0.207



Fig. 4. The second Narrative Subgraph of Narrative C6 in the Las Vegas Dataset, with a Narrative Path at 0.9 threshold.

F. Isolated Narratives: Detecting Temporally Disconnected Activity

In addition to weak and missing transitions, the NEG can reveal narratives that are disconnected from other occurrences of the same narrative. We distinguish between two forms of narrative disappearance: *isolated narratives* and *storyline termination*. An isolated narrative is a narrative with no incident edges to other nodes belonging to the same narrative. Hence, isolated narratives indicate that the user engagement provides no evidence of continuity. Thus, an isolated narrative may represent a short-lived narrative or a limitation in the observed data (e.g., missing data). Most isolated narratives occur in the last third of the data extraction period, due to ceasing activity over time. In the El Paso dataset, the following narrative:

Actor: White supremacist males, **Action:** radicalized, threaten, terrorize, **Event:** a terrorist threat in America, white supremacist terrorism, domestic terrorism,
Description: White supremacist males have been the biggest domestic threat to America for generations

consists of three Narrative Subgraphs and eight isolated nodes. The first isolated narrative node is on day 4, followed by day 6 and day 8, which is comparatively early, considered an extraction period of 16 days. Figure 5 presents the three Narrative Subgraphs for the narrative C12, with a short

Narrative Path of length three. We conclude that C12 is a fragmented narrative with low continuity in semantics and engagement. Since activity in the narrative rises and falls during the extraction period, one possibility is high sensitivity to external events or even bot activity that could prompt messages in response to their posts.



Fig. 5. The three Narrative Subgraphs of Narrative C12 in the El Paso Dataset, with a Narrative Path at 0.5 threshold.

G. Storyline Termination: Detecting Narrative Disappearance

Storyline termination indicates that there is no subsequent activity associated with a narrative for the remainder of the observation period. Figure 3 presents the Narrative Path of the C6 narrative in the Las Vegas dataset. The semantically coherent storyline ends on day 6, whereas engagement with the narrative continues until day 13. A comparison of messages in both storylines reveals the semantic difference; for example, a message from C6D13: “that’s so right, my girlfriend has sooo much class and inspiration for every venue...” versus “politics sickens me when I’ve been awake for 30 hours making sure 2 of my friends attending concert...” from C6D2.

H. Analytical Value of NEEx

We evaluate the analytical value of NEEx by comparing engagement and semantic continuity using longest-path comparison. Figure 6 presents the mean proportion of the longest path that was kept after applying semantic similarity thresholds. We found that path length drops by around 30% at the 0.9 cosine similarity threshold for all three datasets before collapsing to under 10% at the 0.99 threshold in all three datasets. This points to engagement continuity and semantic continuity diverging mainly in the upper range of the threshold scale. We next inspect more closely where this divergence occurs and whether the node retention exhibits a uniform effect across each narrative.

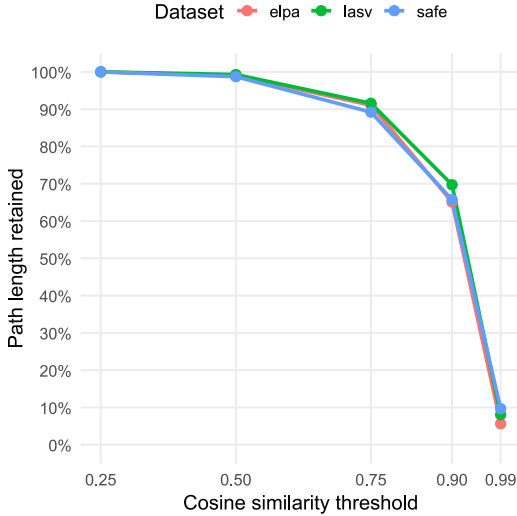


Fig. 6. Mean proportion of longest paths in Narrative Subgraphs for thresholds 0.25,0.5,0.75,0.9 and 0.99 compared to the 0 threshold (engagement-only trajectory).

To investigate the effects of adding the semantic perspective, we created subgraphs for each narrative and each threshold and analyzed node retention compared to an engagement-only graph construction. Figure 7 illustrates a rate of node removal as thresholds increase across datasets. At thresholds up to 0.75, a substantial number of narratives lose nodes before the strict end of the threshold range. The effect becomes dominant at 0.90, where the median drops to approximately 55% (El Paso), 67% (Las Vegas), and 63% (Santa Fe), while the interquartile range spans the full scale. At 0.99, we observed a node retention collapse to nearly 0% in all three datasets.

V. DISCUSSION

Our results demonstrate that narrative evolution on Twitter can be examined through the continuity of user engagement and content semantics. Existing approaches that focus on topic and narrative tracking represent temporal evolution as a sequence of topics or events (see, e.g., [11], [12], [13], [14]). These approaches provide information about whether a topic persists, but do not address whether the meaning associated

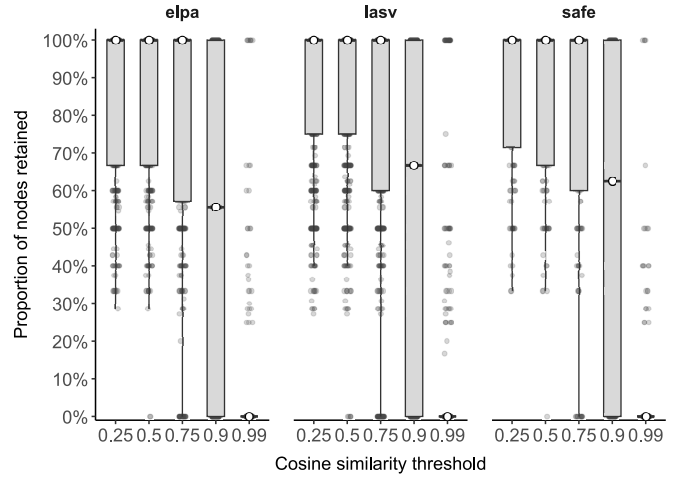


Fig. 7. Boxplots presenting node retention of Narrative Subgraphs for thresholds 0.25,0.5,0.75,0.9 and 0.99 compared to 0 threshold (engagement-only trajectory). The white dot represents the median.

with that topic remains stable. The analysis of semantic similarity shows why this distinction is important. Narrative-day nodes may remain connected through user engagement, even when their textual content differs substantially. Thus, persistence of a narrative in the engagement network should not be interpreted as persistence of the narrative’s underlying meaning. In NEEx, we showed that weak transitions capture this situation: users continue to engage with the same narrative, but its semantic meaning changes.

Our comparison in Section IV-H shows that engagement-only continuity overestimated how long a narrative’s underlying semantics actually persist. We showed that the semantic retention of the same narrative drops to a median of 55-70% at a cosine similarity threshold of 0.9, and collapses further at 0.99. These results suggest that judging a narrative solely by engagement does not reveal semantic changes in the underlying discourse. This is especially relevant for studies that rely on engagement or keyword continuity as a signal of narrative persistence. Thus, our study provides empirical evidence that points to a risk of overstating how stable a narrative’s meaning actually remains over time.

A. Limitations

NEEx relies on the quality of the topic assignments and narrative labels produced by BERTopic and NTLRAG. Any errors resulting from these two components would propagate into the NEG. Moreover, cosine similarity directly depends on the embedding model. While detecting isolated nodes and terminating narratives, we cannot claim the disappearance of narratives on social media in general, given potential changes in participating users, incomplete data collection, or data collection temporal scope.

VI. CONCLUSION

In this study, we presented Narrative Evolution Explorer (NEEx), an analytical framework for analyzing temporal and

semantic evolution of narratives in social media messages. The framework is built upon topic modeling, RAG-based narrative labeling, a chronological user engagement network, and semantic similarity analysis to construct a graph-based representation of the evolution of narratives. NEEEx represents narrative-day combinations as nodes and user engagement relationships as temporal edges. With NEEEx, we identified examples demonstrating a narrative that remains active while its semantic content changes. We considered four topologies, namely weak transitions, missing transitions, isolated narratives, and storyline termination to provide distinct structural patterns that aid in interpreting semantic change, fragmentation, discontinuity, and disappearance of a narrative. Our framework was tested on Twitter datasets from three shooting events to demonstrate how it can be used to reveal different forms of narrative continuity across temporal social media discourse.

We found that narrative persistence should not be interpreted only as the continued presence or prevalence of a topic. A narrative may persist in terms of user engagement while simultaneously undergoing semantic changes, or it may disappear temporarily and subsequently re-emerge. With NEEEx, we make these distinctions explicit and provide a framework for studies about social media narrative evolution.

Our future work will extend the analysis by validating semantic similarity against human assessment. Moreover, we plan to incorporate inter-narrative relationships to identify transitions between narratives rather than focusing on individual narratives. Our future work will also explore cases of high semantic similarity with contradictions and opposing positions.

NEEEx is available to explore on shiny.io (<https://lisagrobrel.shinyapps.io/NEEEx/>).

REFERENCES

- [1] M. Seeger and T. L. Sellnow, *Narratives of crisis: Telling stories of ruin and renewal*. Stanford University Press, 2016.
- [2] A. Pattison, W. Cipolli III, J. Marichal, and C. Cherniakov, "Fracking Twitter: Utilizing machine learning and natural language processing tools for identifying coalition and causal narratives," *Politics & Policy*, vol. 51, no. 5, pp. 755–774, 2023.
- [3] A. Aravamudan, X. Zhang, J. Song, S. M. Fiore, and G. C. Anagnostopoulos, *Influence Dynamics Among Narratives: A Case Study of the Venezuelan Presidential Crisis*. Springer International Publishing, 2021, p. 204–213.
- [4] A. Eklund, M. Forsman, and F. Drewes, "Comparing human-perceived cluster characteristics through the lens of CIPHE: Measuring coherence beyond keywords," *J. Data Min. Digit. Humanit.*, vol. NLP4DH, no. 15044, Mar. 2025.
- [5] L. Grobelscheg, E. Kahr, and M. Strembeck, "NTLRAG: Narrative topic labels derived with retrieval augmented generation," 2026. [Online]. Available: <https://arxiv.org/abs/2602.17216>
- [6] B. Peng, Y. Zhu, Y. Liu, X. Bo, H. Shi, C. Hong, Y. Zhang, and S. Tang, "Graph retrieval-augmented generation: A survey," *ACM Trans. Inf. Syst.*, vol. 44, no. 2, Dec. 2025.
- [7] J. Allan, *Introduction to topic detection and tracking*. USA: Kluwer Academic Publishers, 2002, p. 1–16.
- [8] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, no. null, p. 993–1022, Mar. 2003.
- [9] D. Hall, D. Jurafsky, and C. D. Manning, "Studying the history of ideas using topic models," in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. Honolulu, Hawaii: ACL, Oct. 2008, pp. 363–371.
- [10] A. Steinskog, J. Therkelsen, and B. Gambäck, "Twitter topic modeling by tweet aggregation," in *Proceedings of the 21st Nordic Conference on Computational Linguistics*, J. Tiedemann and N. Tahmasebi, Eds. Gothenburg, Sweden: ACL, May 2017, pp. 77–86.
- [11] L. Grobelscheg, E. Kusen, and M. Strembeck, *Automated Narratives – On the Influence of Bots in Narratives during the 2020 Vienna Terror Attack*, ser. COMPLEXIS. SciTePress, Apr. 2022, pp. 15–25.
- [12] J. Kim and Y. Kim, "Using structural topic modeling to explore the climate change discourse about the Paris Agreement on social media," *Telematics and Informatics Reports*, vol. 15, p. 100148, 2024.
- [13] A. Mishler, E. S. Crabb, S. Paletz, B. Hefright, and E. Golonka, "Using structural topic modeling to detect events and cluster Twitter users in the ukrainian crisis," in *HCI International 2015 - Extended Abstracts*, C. Stephanidis, Ed. Cham: Springer International Publishing, 2015, pp. 639–644.
- [14] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media: A survey," *ACM Comput. Surv.*, vol. 51, no. 2, Feb. 2018.
- [15] N. Chambers and D. Jurafsky, "Unsupervised learning of narrative event chains," in *Proceedings of ACL-08: HLT*, J. D. Moore, S. Teufel, J. Allan, and S. Furui, Eds. Columbus, Ohio: ACL, Jun. 2008, pp. 789–797.
- [16] E. Filatova and V. Hatzivassiloglou, "Event-based extractive summarization," in *Text Summarization Branches Out*. Barcelona, Spain: ACL, Jul. 2004, pp. 104–111.
- [17] D. Shahaf and C. Guestrin, "Connecting the dots between news articles," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '10. New York, NY, USA: ACM, 2010, p. 623–632.
- [18] A. K. Pradhan, H. Mohanty, and R. P. Lal, "Edtbet: Event detection and tracking in Twitter using graph clustering and pre-trained language model," *Procedia Computer Science*, vol. 233, pp. 481–491, 2024, 5th International Conference on Innovative Data Communication Technologies and Application (ICIDCA 2024).
- [19] J. Wang, Y. Liang, F. Meng, Z. Sun, H. Shi, Z. Li, J. Xu, J. Qu, and J. Zhou, "Is ChatGPT a good NLG evaluator? a preliminary study," in *Proceedings of the 4th New Frontiers in Summarization Workshop*, Y. Dong, W. Xiao, L. Wang, F. Liu, and G. Carenini, Eds. Singapore: ACL, Dec. 2023, pp. 1–11.
- [20] A. Madaan, N. Tandon, P. Gupta, S. Hallinan, L. Gao, S. Wiegrefe, U. Alon, N. Dziri, S. Prabhume, Y. Yang, S. Gupta, B. P. Majumder, K. Hermann, S. Welleck, A. Yazdanbakhsh, and P. Clark, "Self-refine: iterative refinement with self-feedback," in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, ser. NIPS '23. Red Hook, NY, USA: Curran Associates Inc., 2023.
- [21] N. Shinn, F. Cassano, A. Gopinath, K. Narasimhan, and S. Yao, "Reflexion: language agents with verbal reinforcement learning," in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, ser. NIPS '23. Red Hook, NY, USA: Curran Associates Inc., 2023.
- [22] J. Zhang and Y. Long, "MLD-EA: Check and complete narrative coherence by introducing emotions and actions," in *Proceedings of the 31st International Conference on Computational Linguistics*. Abu Dhabi, UAE: ACL, Jan. 2025, pp. 1892–1907.
- [23] Y. Wang, H. Li, F. Teng, and L. Chen, "Gorag: Graph-based online retrieval augmented generation for dynamic few-shot social media text classification," in *Proceedings of the ACM Web Conference 2026*, ser. WWW '26. New York, NY, USA: ACM, 2026, p. 9148–9159.
- [24] Q. Liu, H. Peng, Z. Li, X. Meng, Z. Yang, J. Li, L. Sun, and P. S. Yu, "Effective and unsupervised social event detection and evolution via RAG and structural entropy," in *Proceedings of the ACM Web Conference 2026*, ser. WWW '26. New York, NY, USA: ACM, 2026, p. 4529–4540.
- [25] Y.-B. Kang and T. Sellis, "Topictracker: A platform for topic trajectory identification and visualisation," 2021. [Online]. Available: <https://arxiv.org/abs/2103.01432>
- [26] B. Keith Norambuena, "Narrative maps visualization tool (NMVT): An interactive narrative analytics system based on the narrative maps framework," *SoftwareX*, vol. 32, p. 102377, 2025.
- [27] L. Grobelscheg, E. Kahr, and M. Strembeck, "ChronCN: Visualizing topic model output with chronological chain networks," in *2025 12th International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2025, pp. 151–158.

- [28] M. Grootendorst, “Bertopic: Neural topic modeling with a class-based tf-idf procedure,” 2022. [Online]. Available: <https://arxiv.org/abs/2203.05794>
- [29] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.