BEYOND TOPICS: FROM WEAK SIGNALS TO PERSISTENT ISSUES IN EVOLVING PUBLIC DISCOURSE

Anonymous authors
Paper under double-blind review

ABSTRACT

Important collective problems may begin as sparse and heterogeneous observations long before they become obvious topics. Existing discourse-analysis methods can identify recurring semantic patterns and track their prevalence over time, but they do not explicitly determine when distributed evidence is sufficient to form a collective issue, or whether subsequent observations reflect the same evolving problem or a genuinely distinct one. We formulate this challenge as **Continual Emerging Issue Discovery (CEID)**, an open-world task that models how collective issues emerge from fragmented evidence, persist as their manifestations change, and differentiate over time. We instantiate CEID with an issue-centric continual GraphRAG that separates Discourse, Evidence, and Issue representations, retains weak signals under insufficient support, and preserves issue identities across temporal updates. We construct a controlled temporal benchmark grounded in real-world public-service feedback and evaluate issue formation, discourse grounding, temporal identity preservation, and emergence latency. Our method improves issue discovery and discourse grounding, substantially strengthens temporal identity preservation, and detects emerging issues earlier than topic-, LLM-, retrieval-, and graph-based baselines. These results suggest that evolving public feedback is better modeled as an evidence-accumulation process over persistent collective issues than as a sequence of independently discovered topics.

1 INTRODUCTION

Important collective problems often begin as sparse and heterogeneous observations long before they become obvious topics. In public-service feedback, one person may report, "I have been waiting for months since my referral," another that "nobody has contacted me," and another that an appointment has repeatedly been postponed. Individually, these statements may appear to be unrelated complaints; collectively, they may reveal an emerging problem such as delayed access to care. No single expression necessarily names the problem. Rather, the issue becomes visible only as evidence accumulates across people and time.

This creates a challenge that goes beyond discovering recurring themes. Early observations may be too weak to justify declaring a new issue, yet discarding them risks recognizing the problem only after it has become widespread (Parekh et al., 2024). Once an issue has formed, later observations raise a second question: do they reflect new manifestations of the same underlying problem, or evidence of a genuinely distinct one? For example, a report that "my mother cannot use the online booking system" may represent another barrier within a broader access problem, or the emergence of a more specific issue concerning digital exclusion. The challenge is therefore not only to identify *what people are discussing*, but to determine *when distributed evidence is sufficient to form a collective issue and whether that issue retains its identity as its manifestations change*.

Existing discourse-analysis methods are not designed around this distinction. Topic models, including dynamic and online variants, are effective at identifying recurring semantic patterns and tracking changes in their prevalence (Blei & Lafferty, 2006; Zhang & Lauw, 2022; Wu et al., 2024). However, changing topic distributions are not equivalent to forming and maintaining the identity of an evolving real-world problem. A topic may capture that appointments, waiting, and access frequently occur together, yet it does not explicitly determine whether sparse observations have accumulated into a new systemic issue, whether subsequent evidence should update that issue, or whether a re-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

lated but distinct issue has emerged. Likewise, analysing temporal snapshots independently can recover changing themes while losing continuity between the underlying problems represented at different times. The limitation is therefore not merely temporal; it is representational (Karakkaparambil James et al., 2024). Topics capture recurring semantic regularities, whereas the target in CEID is an evidence-supported collective issue with persistent identity.

We formulate this problem as **Continual Emerging Issue Discovery (CEID)**. CEID considers an open-world stream in which the set of underlying issues is neither fixed nor known in advance. An Emerging Issue is not merely a cluster of semantically similar documents, but a collective problem supported by heterogeneous evidence whose identity may persist as its manifestations and temporal state change. CEID therefore couples two scientific questions. **Issue Formation** asks when fragmented and individually incomplete observations provide sufficient collective evidence to constitute an issue. **Issue Evolution** asks how an issue can preserve its identity as new evidence arrives while allowing genuinely distinct but related concerns to differentiate. Together, these questions imply three representational requirements: a system should defer commitment when collective evidence is insufficient, preserve issue identity as manifestations and states change, and create a distinct but related issue when the accumulated evidence warrants differentiation. Weak evidence therefore need not be forced into an existing issue or discarded as noise; it may remain unresolved until further observations provide sufficient support.

To realize these requirements, we introduce an issue-centric continual GraphRAG (Edge et al., 2024) that maintains a persistent, evidence-grounded representation of the stream. The framework separates raw Discourse, problem-oriented Evidence, and collective Issue representations. Evidence with insufficient support is retained as a candidate rather than prematurely committed to an Issue. Evidence supporting an existing Issue updates its state and support while preserving its identity; evidence indicating a sufficiently distinct but related problem instead creates a new Issue linked to its predecessor through an evolution relation. Historical evidence and Issues remain retrievable, enabling the model to reason over an evolving graph rather than rediscovering each temporal wave independently. We evaluate CEID on a temporally structured benchmark grounded in real-world citizen feedback. The benchmark tests whether systems can form evidence-grounded collective issues, preserve their identities across waves, and detect emerging issues before they become dominant patterns. Our work makes three main contributions:

- **Continual Emerging Issue Discovery.** We formulate CEID as an open-world problem that shifts the unit of analysis from recurring semantic topics to evidence-supported collective issues that form, persist, and differentiate over time.
- **Persistent issue-centric GraphRAG.** We introduce a persistent issue-centric GraphRAG that accumulates heterogeneous evidence, defers commitment to weak signals under insufficient support, preserves issue identity across temporal updates, and explicitly represents issue differentiation through evolution relations.
- **Temporal benchmark and evaluation.** We establish a temporally structured benchmark and evaluation framework for measuring issue formation, grounding, identity preservation, state tracking, and emergence latency.

2 Related Work

Topic and temporal discourse modeling. Topic modeling has long been used to uncover latent semantic structure in large text collections, while dynamic and online extensions model how topic content or prevalence changes over time Blei et al. (2001); Blei & Lafferty (2006); Hoffman et al. (2010). More recent embedding-based approaches further improve semantic coherence and flexibility in topic discovery Grootendorst (2022). These methods are effective for identifying recurring themes and tracking which themes become more or less prominent. CEID addresses a different representational problem: changes in topic prevalence do not determine when heterogeneous observations collectively provide sufficient evidence for a new Issue, nor whether subsequent observations represent new manifestations of the same Issue. We therefore shift the analytical target from temporal semantic regularities to evidence-supported collective issues with persistent identity.

Emerging event discovery, novelty detection, and online clustering. A related line of work studies emerging-event detection, event analysis, novelty detection, burst detection, and online or stream

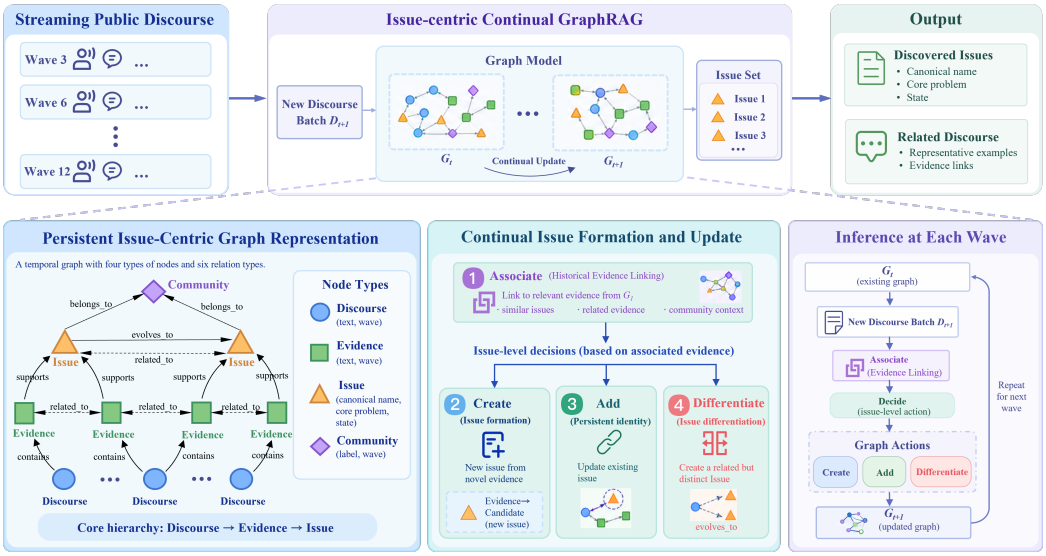

Figure 1: Overview of CEID and the Issue-Centric Continual GraphRAG Framework

clustering Allan et al. (1998); Kleinberg (2002); Kasiviswanathan et al. (2011); Aggarwal et al. (2003); AlSumait et al. (2008). These approaches identify novel content, rapidly growing signals, or clusters that appear and change as new observations arrive; some further align or track clusters across temporal windows Blei & Lafferty (2006). CEID shares this continual and open-world setting, but novelty alone does not imply that a collective Issue has formed. A novel observation may remain an unsupported weak signal, while substantial linguistic change may still reflect the same underlying Issue. CEID must therefore reason not only about novelty, but also about evidence accumulation, identity preservation, and issue differentiation.

GraphRAG and persistent memory. Retrieval-augmented generation extends language models with external evidence, while GraphRAG organizes retrieved information through entities, relations, and higher-order graph structure to support structured retrieval and reasoning Lewis et al. (2020); Edge et al. (2024). Related work on long-term memory, continual knowledge updating, and dynamic graph maintenance further enables models to reuse historical information across interactions or evolving data streams Packer et al. (2023); Zhong et al. (2024); Gutiérrez et al. (2025); Xu et al. (2025). These capabilities provide useful foundations for CEID, but persistent memory alone does not specify how issue identity should be formed and maintained. Our framework organizes persistent graph memory around three task-specific requirements: retaining weak evidence before sufficient support is available, preserving an existing issue while its state and supporting evidence change, and creating an explicitly related but distinct issue when genuine differentiation occurs.

These research directions provide complementary capabilities in semantic organization, novelty detection, temporal tracking, and persistent retrieval, but none directly models how a previously unnamed collective issue emerges from distributed evidence, acquires a stable identity, and subsequently evolves over time. CEID takes this process itself as the central inference problem. Accordingly, our issue-centric continual GraphRAG is organized around three requirements: *issue formation*, *persistent identity*, and *issue differentiation*, rather than treating temporal discourse as a sequence of independently discovered topics or clusters.

3 Proposed Method

To address this gap, we formulate CEID as an open-world continual inference problem and instantiate it with a persistent issue-centric graph separating Discourse, Evidence, and Issue representations. We then introduce mechanisms for historical association, deferred issue formation, identity-preserving update, and issue differentiation.

3.1 CEID TASK FORMULATION

We study Continual Emerging Issue Discovery (CEID) from a stream of public discourse. Let the incoming discourse be divided into temporal waves $\mathcal{D} = \{D_1, \ldots, D_T\}$, where D_t denotes the discourse observed at wave t. Each discourse instance is represented as $d_i = (id_i, w_i, x_i)$, where id_i is the discourse identifier, w_i its observation wave, and x_i the raw textual expression.

At wave t, the system has access only to the accumulated discourse and must infer the current set of collective issues:

$$D_{\leq t} = \bigcup_{\tau=1}^{t} D_\tau \quad \longrightarrow \quad \mathcal{I}_t = \{I_1, \ldots, I_{K_t}\}. \tag{1}$$

Neither K_t nor the identities of the underlying issues are assumed to be fixed or known in advance.

Each issue is represented as $I = (id, c, n, s, E)$, where c denotes its canonical name, n its core problem, s its current temporal state, and E its supporting evidence. We use $s \in \{\textit{emerging}, \textit{growing}, \textit{persistent}, \textit{declining}\}$.

In addition to established issues, the system maintains a candidate pool $\mathcal{P}_t$ for coherent but insufficiently supported evidence. For evaluation and temporal tracking, we additionally define a broader lifecycle status:

$$\ell \in \{\textit{candidate}, \textit{emerging}, \textit{growing}, \textit{persistent}, \textit{declining}\} \tag{2}$$

where *candidate* denotes evidence that remains insufficiently supported for full issue formation. CEID thus requires both **issue formation**, determining when accumulated evidence supports a new issue, and **issue evolution**, determining whether subsequent evidence updates an existing issue while preserving its identity or indicates a related but distinct issue.

3.2 PERSISTENT ISSUE-CENTRIC GRAPH REPRESENTATION

To support evidence accumulation and persistent identity across waves, we represent the evolving discourse stream as a persistent issue-centric graph that separates Discourse, Evidence, Issue, and Community representations. At wave t,

$$G_t = (V_t, R_t), \tag{3}$$

where the node set consists of four types:

$$V_t = V_t^D \cup V_t^E \cup V_t^I \cup V_t^C, \tag{4}$$

corresponding to **Discourse, Evidence, Issue, and Community** nodes.

A Discourse node represents an observed public expression, and an Evidence node represents a problem-oriented interpretation extracted from one discourse instance. An Issue node represents a collective problem consolidated from multiple supporting evidence units and stores its canonical name, core problem, and temporal state. Its supporting evidence is represented explicitly through incoming `supports` relations. Community nodes provide auxiliary structural context for retrieval and do not directly determine issue assignments.

The graph contains six relation types:

$$\begin{aligned}
&\text{Discourse} \rightarrow \text{Evidence}: && \texttt{contains},\\
&\text{Evidence} \leftrightarrow \text{Evidence}: && \texttt{related_to},\\
&\text{Evidence} \rightarrow \text{Issue}: && \texttt{supports},\\
&\text{Issue} \leftrightarrow \text{Issue}: && \texttt{related_to},\\
&\text{Issue} \rightarrow \text{Issue}: && \texttt{evolves_to},\\
&\text{Issue} \rightarrow \text{Community}: && \texttt{belongs_to}.
\end{aligned}$$

These relations serve distinct roles. Evidence-level relations connect heterogeneous observations expressing sufficiently related underlying problems, while `supports` relations ground each Issue in its source evidence. Persistent Issue nodes provide stable identities across temporal waves, and the directed `evolves_to` relation records explicit issue differentiation while preserving lineage between related problems. Community structure provides auxiliary context for retrieval.

3.3 CONTINUAL ISSUE FORMATION AND UPDATE

As a new discourse batch D_{t+1} arrives, the persistent graph is incrementally updated:

$$G_{t+1} = \mathrm{Update}(G_t, D_{t+1}). \tag{5}$$

The update cycle first converts new discourse into problem-oriented evidence, associates new evidence with relevant historical evidence and context, and then determines whether the new evidence should form a new Issue, update an existing Issue, differentiate into a related but distinct Issue, or remain unresolved in the candidate pool.

Evidence extraction. For each newly observed discourse instance d, an LLM extracts one or more problem-oriented evidence units, $d \to \{e_1, \ldots, e_m\}$, and each evidence unit is represented as a vector $z_e = f(e)$. These evidence units constitute the basic signals used for subsequent retrieval and issue-level reasoning.

Associate: historical evidence linking. For each new evidence unit e, we retrieve potentially related historical evidence using approximate nearest-neighbor search implemented with FAISS (Johnson et al., 2019):

$$\mathcal{N}(e) = \mathrm{ANN}(z_e, Z_t^E, k), \tag{6}$$

where Z_t^E denotes the vector representations of evidence already stored in G_t. The retrieved candidates are then verified by an LLM, and verified pairs are linked through `related_to` relations. These verified associations are subsequently expanded from local evidence to the persistent issue structure:

$$E_{\mathrm{new}} \to E_{\mathrm{old}} \to I_{\mathrm{old}} \to I_{\mathrm{neighbor}}. \tag{7}$$

This association process provides both evidence-level correspondences and issue-level historical context for the subsequent formation and update decisions.

Create: deferred issue formation. Issue formation is triggered by accumulated evidence rather than by individual discourse instances. Let E_c denote a coherent set of related evidence. The system either consolidates E_c into a new Issue when collective support is sufficient, or retains it as an unresolved candidate otherwise:

$$E_c \to \begin{cases} I_{\mathrm{new}}, & \text{if sufficient collective support exists,} \\ P_j \in \mathcal{P}_t, & \text{otherwise.} \end{cases} \tag{8}$$

As additional evidence arrives, the candidate is re-evaluated and may lead to a new Issue, support an existing Issue, or remain unresolved. This deferred-commitment mechanism preserves early weak signals without prematurely promoting them to Issues.

Add: identity-preserving update. When the associated historical context indicates that new evidence supports an existing Issue, the evidence is attached to that Issue rather than creating a new identity. The Issue therefore follows a persistent trajectory

$$I_t \to I_{t+1}, \tag{9}$$

while retaining the same Issue identifier and core problem.

We explicitly separate Issue identity from temporal state. The identity is represented by the relatively stable collective problem captured by its canonical name and core problem, whereas the state reflects its current temporal condition. Thus, new evidence may induce a state transition $s_t(I) \to s_{t+1}(I)$, for example from *emerging* to *growing* or from *persistent* to *declining*, without implying identity replacement. In this case, the change is within the same Issue:

$$\text{State Change} : I \to I. \tag{10}$$

Differentiate: related but distinct issue formation. New evidence may instead indicate a collective problem that is related to, but sufficiently distinct from, an existing Issue. Rather than continuously broadening the original Issue, the framework creates a new Issue node and explicitly records its lineage:

$$I_a \xrightarrow{\texttt{evolves_to}} I_b. \tag{11}$$

The original Issue remains in the graph together with its historical evidence and trajectory, while the differentiated Issue acquires its own identity, supporting evidence, and temporal state. This mechanism allows the graph to represent structural changes such as issue splitting without forcing semantically distinct evidence into an increasingly broad or obsolete Issue. We therefore distinguish issue differentiation, $I_a \to I_b$, from identity-preserving state change, $I \to I$.

Historical retention and community context. Issues remain part of the persistent graph even when their current state becomes *declining*:

$$state(I) = \textit{declining} \not\Rightarrow I \notin G_t. \tag{12}$$

Their historical evidence and relations therefore remain retrievable, allowing future observations to be interpreted in the context of previously active problems.

After Issue-level relations have been updated, the higher-level Community structure is refreshed using graph community detection over the Issue–Evidence connectivity graph. Communities provide auxiliary structural context for subsequent retrieval but do not directly determine whether new evidence is assigned to an Issue.

3.4 Inference at Each Wave

At each temporal wave t, the current discourse batch is processed against the persistent graph from the previous wave. The overall inference process is

$$D_t \to E_t \to \text{Associate}(E_t, G_{t-1}) \to \text{Decide} \to G_t \to \mathcal{I}_t. \tag{13}$$

Given the associated historical context, the decision stage determines whether newly observed evidence should

$$\text{Decide} \in \{\text{Create}, \text{Add}, \text{Differentiate}\}. \tag{14}$$

Here, CREATE forms a new Issue from sufficiently supported evidence, ADD updates an existing Issue while preserving its identity, and DIFFERENTIATE creates a related but distinct Issue linked to its predecessor. When the available evidence does not yet support any of these decisions, it remains unresolved in the candidate pool until additional evidence becomes available.

After all discourse instances in the current wave have been processed, the system outputs the active Issue set together with each Issue's canonical identity, temporal state, and supporting evidence:

$$O_t = \{(id_i, c_i, n_i, s_i, E_i)\}_{i=1}^{K_t}. \tag{15}$$

Because the graph persists across waves, O_t is not an independent reconstruction of the current discourse. Instead, it represents the current state of an evolving issue structure in which historical evidence, unresolved candidates, established Issues, and explicit evolution relations remain available for subsequent reasoning.

4 Experiments

We evaluate our approach on Continual Emerging Issue Discovery over a sequence of temporal waves. Our experiments are designed to answer two scientific questions: **SQ1: Issue Formation**, which asks whether fragmented and heterogeneous discourse can be consolidated into coherent collective issues; and **SQ2: Issue Evolution**, which asks whether discovered issues can preserve their identities as new evidence arrives while allowing related but distinct issues to differentiate. We compare against representative topic-modeling, LLM-based, retrieval-based, and graph-based baselines under a unified issue-level output format.

4.1 Dataset

Source Corpus. The CEID benchmark is grounded in a private corpus of NHS citizen feedback previously used in our earlier work (Shi et al., 2025). The source corpus is used to ground the semantic content, expression patterns, and temporal characteristics of the benchmark. It is not publicly released under the governing data-sharing agreement.

Scientific Question	Metric	What it measures	Direction
SQ1: Issue Formation	Issue F1	Accuracy of discovering collective issues	↑
	Discourse F1	Whether discovered issues are supported by the correct discourse evidence	↑
	NMI/AMI/ARI	Agreement between predicted and ground-truth discourse groupings	↑
SQ2: Issue Evolution	Temporal Identity Preservation (TIP)	Whether the same issue identity is maintained across waves	↑
	State Macro-F1	Accuracy of lifecycle-status tracking over time	↑
	Emergence Detection Delay (EDD)	How quickly newly emerging issues are detected	↓

Table 1: Evaluation metrics for the two CEID scientific questions. SQ1 evaluates issue formation and discourse grounding, while SQ2 evaluates temporal identity preservation, lifecycle-state tracking, and emergence latency.

Controlled Benchmark Construction. For evaluation, we construct a controlled synthetic benchmark grounded in the source corpus. The benchmark is designed to reflect the empirical characteristics of the original feedback stream while providing explicit ground truth for issue formation and evolution. The construction process consists of three stages. First, we extract problem-oriented issue mentions from the source feedback and consolidate them into a high-recall candidate issue bank. Second, candidate issues are normalized, merged, split, and filtered to construct a canonical Ground-truth Issue Bank, in which each issue is associated with its core problem, issue type, supporting evidence, and temporal trajectory. Third, conditioned on the resulting issue bank and a latent temporal world state, we generate temporally structured discourse that simulates how the corresponding issues may be expressed across successive waves.

All quantitative experiments reported in this paper are conducted on the resulting controlled benchmark rather than directly on the private source corpus.

4.2 Evaluation Protocol and Metrics

We evaluate all methods under the same issue-level output format. Since different methods may produce different numbers of issues and use different surface descriptions, we first perform **semantic matching** between predicted and ground-truth issues. Candidate pairs are identified using vector similarity and verified by an LLM, followed by one-to-one Hungarian matching (Kuhn, 1955). The resulting alignment is then used for both issue formation and temporal evaluation.

SQ1: Issue Formation. Issue F1 is our primary metric and is computed from the one-to-one semantic alignment. Discourse Precision, Recall, and F1 evaluate whether the predicted issues are grounded in the appropriate discourse evidence. NMI, AMI, and ARI provide complementary measures of discourse-level grouping quality.

SQ2: Issue Evolution. TIP evaluates whether an issue is represented by a consistent predicted identity across temporal waves. State Macro-F1 measures the correctness of the predicted lifecycle state ℓ (*candidate*, *emerging*, *growing*, *persistent*, or *declining*). EDD measures the temporal gap between the ground-truth emergence of an issue and its first successful detection. Issue differentiation is further examined qualitatively through the evolution cases in Section 4.5.

4.3 Baselines and Implementation

We compare against five representative baselines spanning topic modeling, direct LLM inference, retrieval-augmented generation, and graph-based retrieval: LDA, BERTopic, Zero-Shot LLM, Naive RAG, and Static GraphRAG. For all LLM-based methods, we use the same underlying LLM and the same structured output schema. All retrieval-based methods use the same L2-normalized hashing-based text representation wherever applicable. All experiments are implemented in Python. For semantic retrieval, we use approximate nearest-neighbor search implemented with FAISS (Johnson et al., 2019), and graph communities are detected using the Leiden community detection algorithm (Traag et al., 2019). Retrieval depths and other model configurations are fixed across experiments, with full values reported in Appendix A. Each experiment is repeated with the same configuration across temporal waves.

Method	Issue F1	Discourse F1	NMI	AMI	ARI	TIP	State F1	EDD
LDA	0.26	0.26	0.55	0.15	0.03	0.15	0.24	0.71
BERTopic	0.27	0.35	0.50	0.16	-0.01	0.16	0.27	0.91
Zero-Shot LLM	0.27	0.18	0.25	0.08	-0.03	0.09	0.29	0.83
Naive RAG	0.32	0.37	0.40	0.21	0.01	0.21	0.26	0.79
Static GraphRAG	0.19	0.08	0.32	0.12	0.02	0.10	0.24	1
Ours	**0.37**	**0.55**	**0.67**	**0.25**	**0.08**	**0.50**	0.23	**0.53**

Table 2: Main results on Continual Emerging Issue Discovery.

4.4 Main Results

Table 2 reports the main results on Continual Emerging Issue Discovery. Overall, our method achieves the best performance on the primary issue discovery metric and on most discourse-level and temporal metrics, indicating that explicitly modeling evidence-grounded and persistent issue structures is beneficial for continual public discourse analysis.

Issue formation and discourse grounding. Our method obtains the highest score of 0.37 in Issue F1, outperforming the strongest baseline, Naive RAG (0.32), by 5 percentage points. The advantage is more pronounced for Discourse F1, where our method achieves 0.55 compared with 0.37 for Naive RAG and 0.35 for BERTopic. This suggests that the proposed evidence-centric representation is particularly effective at associating heterogeneous discourse expressions with their underlying collective issues rather than relying only on surface-level semantic similarity. Our method also achieves the highest scores of 0.67, 0.25, and 0.08 in NMI, AMI, and ARI, respectively, indicating improved agreement between the discovered issue structure and the underlying discourse organization. In contrast, conventional topic models provide relatively strong NMI in the case of LDA (0.55) but substantially lower AMI and ARI, suggesting that broad thematic grouping alone does not necessarily recover the underlying issue structure. Naive RAG improves over the LLM-only baseline, while the lower performance of Static GraphRAG shows that introducing a graph structure without continual issue maintenance does not by itself guarantee better issue formation. The Discourse F1 of 0.55 is notably higher than the Issue F1 of 0.37, indicating that the model can associate discourse with the appropriate problem signals more successfully than it can consolidate all such signals into a compact canonical issue inventory. This gap is consistent with the difficulty of issue formation: the task requires not only recognizing related observations, but also deciding when heterogeneous evidence constitutes one collective issue rather than multiple neighboring concerns.

Issue Evolution: Persistent identity and early emergence. The largest improvements appear in temporal metrics. Our method achieves a score of 0.50 in TIP, substantially higher than all baselines, with the next-best method reaching only 0.21. This result supports the central role of persistent issue representations in maintaining issue identity across temporal waves. Similarly, our method obtains the lowest EDD (0.53), compared with 0.71 for LDA, 0.79 for Naive RAG, 0.83 for Zero-Shot LLM, and 1.00 for Static GraphRAG. The result indicates that the proposed continual update mechanism can recognize newly emerging issues earlier rather than waiting for them to become dominant semantic clusters. However, State F1 remains low (0.23), indicating that lifecycle-state prediction remains challenging; among established issues, distinguishing emerging, growing, persistent, and declining states is particularly difficult. The result also reveals an important limitation of the current framework: the graph can preserve an issue as discourse evolves, but state estimation requires more reliable temporal evidence and a more explicit model of state transition dynamics. We therefore treat state tracking as a remaining challenge rather than attributing the overall temporal gains to uniformly improved state prediction.

4.5 Case Study and Graph Visualization

To provide an interpretable view of how the proposed framework operates, we visualize representative subgraphs and trace three types of issue formation and evolution cases. The cases are selected to illustrate complementary properties of the framework.

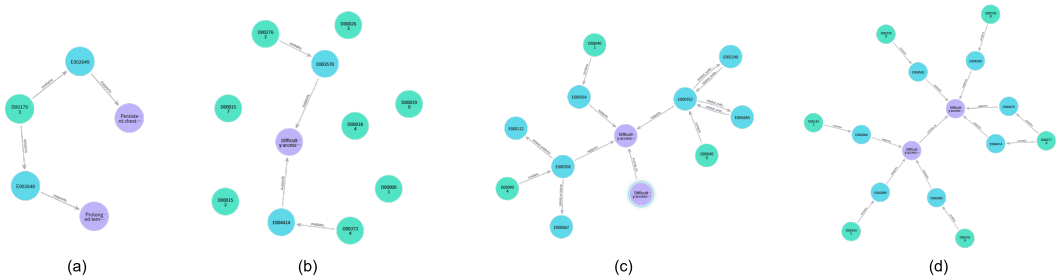

Figure 2: Representative CEID cases and graph evolution: (a–b) An instance of heterogeneous discourse is broken down into several units of evidence, each of which supports the formation of a different issue; (c) Persistent Issue identities are maintained across temporal waves as new evidence is incorporated and temporal states change; (d) Issues evolve into new issues and are differentiated as new waves of discourse arrive.

Case 1: Evidence Decomposition and Issue Formation. Figure 2 (a) shows a discourse containing multiple problem signals: "*Sometimes I think kind of, we've been stuck in this temporary flat since winter hit. The damp's getting to me, and I can't shake this heavy feeling. It's like the walls are closing in, and my chest is tight all the time.*" The discourse corresponds to two distinct issues: prolonged temporary housing and persistent chest tightness. Our framework decomposes it into multiple evidence units, allowing different problem signals to be associated with different issues. Figure 2 (b) illustrates how related evidence is aggregated into an Issue, while unsupported discourse is retained in the *candidate* pool for subsequent aggregation.

Case 2: Persistent Issue Identity. Figures 2 (c) presents two issues that recur across multiple temporal waves while changing in their surface manifestations and contextual realization. The graph preserves the corresponding Issue identities while continuously attaching newly observed evidence and updating their states. Instead of independently rediscovering semantically related topics at each wave, the persistent graph connects observations across time and exposes their evolution trajectories.

Case 3: Issue Differentiation. Figure 2(d) illustrates how the framework distinguishes identity-preserving updates from the emergence of a related but distinct Issue. At the 9m wave, the identified Issue is "People experience long waits when trying to get healthcare." At the 12m wave, newly arriving discourse instead provides evidence for a different access problem: "People experience difficulties navigating between services when trying to get healthcare." Although the two Issues are related through the broader theme of healthcare access, their core problems are distinct. The former concerns delays in receiving care, whereas the latter concerns difficulties in navigating and coordinating across services. The new evidence therefore cannot be explained simply as an additional manifestation of the waiting-time Issue. Rather than attaching it to the existing Issue or replacing the original identity, the framework creates a new Issue node and links it to the earlier Issue through the `evolves_to` relation. This preserves the historical trajectory of the original Issue while explicitly representing the emergence of a related but distinct collective problem.

5 Conclusion

We introduced CEID to model how collective issues form, persist, and differentiate in evolving public discourse. Our issue-centric continual GraphRAG accumulates heterogeneous evidence, preserves issue identity across temporal updates, and explicitly represents related but distinct issue evolution. Experiments on a temporally structured benchmark grounded in real-world citizen feedback show improved issue discovery and discourse grounding, substantially stronger temporal identity preservation, and earlier emergence detection than representative baselines. These results suggest that evolving discourse is better modeled as an evidence-accumulation process over persistent collective issues rather than as a sequence of independently discovered topics.

AI USE STATEMENT

In this work, LLMs were used as part of the construction of the CEID benchmark. The benchmark is grounded in a private corpus of real NHS citizen feedback collected during the COVID-19 pandemic. LLMs assisted with extracting and normalizing problem-oriented issue descriptions, constructing preliminary issue-level annotations, and generating temporally structured discourse conditioned on the resulting issue bank and temporal trajectories. The benchmark therefore contains synthetically generated discourse derived from the semantic content, expression patterns, and temporal structure of the underlying NHS corpus. The original NHS feedback itself was not generated by AI and is not publicly released under the governing data-sharing agreement. AI-assisted benchmark outputs were reviewed and verified by the authors before use in the experiments.

We have not used generative AI tools to design the research methodology, implement the method, formulate theoretical claims, or interpret results. Additionally, we used AI tools for manuscript preparation: drafting text and improving readability. All AI-assisted work was reviewed and approved by the authors.

We take full responsibility for the final content of this work, including all text, claims, data records, and artifacts produced with the aid of generative AI.

REFERENCES

Charu C. Aggarwal, Jiawei Han, Jianyong Wang, and Philip S. Yu. A framework for clustering evolving data streams. In *Proceedings of the 29th International Conference on Very Large Data Bases*, pp. 81–92, 2003.

James Allan, Ron Papka, and Victor Lavrenko. On-line new event detection and tracking. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 37–45, 1998.

Loulwah AlSumait, Daniel Barbará, and Carlotta Domeniconi. On-line lda: Adaptive topic models for mining text streams with applications to topic detection and tracking. In *2008 eighth IEEE international conference on data mining*, pp. 3–12. IEEE, 2008.

David Blei, Andrew Ng, and Michael Jordan. Latent dirichlet allocation. *Advances in neural information processing systems*, 14, 2001.

David M Blei and John D Lafferty. Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, pp. 113–120, 2006.

Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, Dasha Metropolitansky, Robert Osazuwa Ness, and Jonathan Larson. From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*, 2024.

Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*, 2022.

Bernal Jiménez Gutiérrez, Yiheng Shu, Weijian Qi, Sizhe Zhou, and Yu Su. From RAG to memory: Non-parametric continual learning for large language models. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 21497–21515. PMLR, 2025.

Matthew Hoffman, Francis Bach, and David Blei. Online learning for latent dirichlet allocation. *advances in neural information processing systems*, 23, 2010.

Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 7(3):535–547, 2019. doi: 10.1109/TBDATA.2019.2921572.

Charu Karakkaparambil James, Mayank Nagda, Nooshin Haji Ghassemi, Marius Kloft, and Sophie Fellenz. Evaluating dynamic topic models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, pp. 160–176. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.acl-long.11.

Shiva Prasad Kasiviswanathan, Prem Melville, Arindam Banerjee, and Vikas Sindhwani. Emerging topic detection using dictionary learning. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pp. 745–754, 2011.

Jon Kleinberg. Bursty and hierarchical structure in streams. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 91–101, 2002.

Harold W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1–2):83–97, 1955. doi: 10.1002/nav.3800020109.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33: 9459–9474, 2020.

Charles Packer, Sarah Wooders, Kevin Lin, Vivian Fang, Shishir G Patil, Ion Stoica, and Joseph E Gonzalez. Memgpt: Towards llms as operating systems. *arXiv preprint arXiv:2310.08560*, 2023.

Tanmay Parekh, Anh Mac, Jiarui Yu, Yuxuan Dong, Syed Shahriar, Bonnie Liu, Eric Yang, Kuan-Hao Huang, Wei Wang, Nanyun Peng, and Kai-Wei Chang. Event detection from social media for epidemic prediction. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 5758–5783. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.naacl-long.322.

Daqian Shi, Xiaolei Diao, Jinge Wu, Honghan Wu, Xiongfeng Tang, Felix Naughton, and Paulina Bondaronek. Graph-based llm over semi-structured population data for dynamic policy response. In *International Workshop on Efficient Medical Artificial Intelligence*, pp. 278–288. Springer, 2025.

Vincent A. Traag, Ludo Waltman, and Nees Jan van Eck. From louvain to leiden: Guaranteeing well-connected communities. *Scientific Reports*, 9(1):5233, 2019. doi: 10.1038/s41598-019-41695-z.

Xiaobao Wu, Xinshuai Dong, Liangming Pan, Thong Nguyen, and Anh Tuan Luu. Modeling dynamic topics in chain-free fashion by evolution-tracking contrastive learning and unassociated word exclusion. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 3088–3105. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.findings-acl.183.

Wujiang Xu, Zujie Liang, Kai Mei, Hang Gao, Juntao Tan, and Yongfeng Zhang. A-MEM: Agentic memory for llm agents. In *Advances in Neural Information Processing Systems*, volume 38, pp. 20004–20031, 2025. doi: 10.52202/085713-0593.

Delvin Ce Zhang and Hady Lauw. Dynamic topic models for temporal document networks. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 26281–26292. PMLR, 2022.

Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. Memorybank: Enhancing large language models with long-term memory. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 19724–19731, 2024.

A Implementation Details

We use deepseek-v4-flash as the common LLM for evidence extraction, evidence-relation verification, issue formation, issue differentiation, state updating, issue canonicalization, and benchmark generation. All semantic representations are produced by Hashing bag-of-unigrams+bigrams vector and L2-normalized before similarity computation. We use top-12 retrieval for Evidence and top-5 retrieval for Issues. Retrieved candidates are subsequently verified by the LLM rather than being linked solely by embedding similarity. Community structure is constructed using the Leiden algorithm. All LLM decisions are implemented with fixed structured JSON prompts covering evidence extraction, relation verification, issue Create/Add/Differentiate decisions, and state estimation, using

Wave Transition	Retained Evidence	Updated Issues	Candidate Promotions	Issue Evolutions
3m → 6m	1387	293	109	98
6m → 9m	2775	495	60	187
9m → 12m	4352	590	75	287

Table 3: A qualitative inspection experiment of the evolution trajectory, where "Retained Evidence" refers to the number of pieces of evidence retained, "Updated Issues" refers to the number of new issues, "Candidate Promotions" refers to the increase in the number of candidates, and "Issue Evolutions" refers to the number of issue evolutions

temperature 0.2 and a maximum output length of 196608. For continual updates, an incoming Evidence is first matched to historical Evidence and then expanded to candidate Issues and neighboring Issues. Embedding similarity is used only for retrieval and blocking, while final relation and issue decisions are made by the LLM. For evaluation, predicted and ground-truth Issues are first filtered by an embedding threshold 0.15 and then aligned using one-to-one Hungarian matching with the same semantic verification procedure across all methods. All experiments use fixed random seeds, identical model configurations where applicable, and cached LLM responses and embeddings for reproducibility.

B Cross-Wave Update Ablation Analysis

We further analyze the role of cross-wave graph updates by contrasting our continual configuration with Static GraphRAG, which reconstructs the graph independently at each checkpoint. While this comparison is not a strict component-wise ablation, it provides a counterfactual reference for understanding the effect of preserving graph state across waves. As reported in Table 2, the continual configuration substantially improves Temporal Identity Preservation (TIP), increasing it from 0.10 to 0.50, and reduces Emergence Detection Delay (EDD) from 1.00 to 0.53. These differences suggest that retaining historical evidence and issue structures across checkpoints is important for maintaining issue continuity and recognizing emerging issues earlier. Qualitative inspection of the graph trajectories in Table 3 further shows that cross-wave updates allow new discourse to be associated with previously accumulated evidence and issues, whereas independently rebuilt graphs tend to represent temporally related observations as separate structures.