A Knowledge-Based Framework for Urban Event Detection via Temporal Knowledge Graphs and Large Language Models

Aleksandr Antonov
ITMO University
St. Petersburg, Russia
asantonov@itmo.ru

Liubov Tupikina
ITMO University
St. Petersburg, Russia,
Bell labs, Paris, France
lvtupikina@itmo.ru

Abstract

Urban agencies need event analytics that are accurate and explainable. We present Knowledge-based Event Detection (KED), which models events as temporal knowledge graphs to preserve participants, actions, places, and times. The specificity of KED is that it integrates LLM-assisted extraction of entities and relations, temporal knowledge graph construction with geocoded toponyms and proximity edges. In addition it incorporates semantic/contextual links allowing parametric community detection and adjustable event granularity. On Event2012 benchmarks, KED is competitive on long-tailed corpora and outperforms strong text baselines in balanced, lower-data regimes (a setting common for city-specific monitoring) while delivering auditable, graph-structured outputs that facilitate downstream situational awareness and decision support. By aligning geo-temporal structure with semantic context, KED advances explainable, transferable event detection for smart-city applications such as incident monitoring, mobility disruptions, and environmental alerts.

1 Introduction

The task of Event detection (ED) from semi-structured and unstructured text sets of messages is well-known and typical in the management of modern cities, where citizens' appeals are one of the sources of data on the state of the urban environment [1]. However, the intensity of such requests and the specifics of the topics require the use of special algorithms for their processing to reduce the burden on city services [2, 3]. One of the typical tasks is event extraction, which involves the detection of hidden, most often negative, events in the urban environment (accidents, emergencies, public order violations), which lead to multiple intensive or regular appeals from citizens [1]. The ability to identify such situations determines the effectiveness of eliminating the causes of an urban problem and determines the effectiveness of city management.

1.1 Problem Statement

Existing approaches to event detection do not allow us to consider the time dilation of some types of events [4] and the association of time-separated messages with a single "hidden" event. At the same time, large language models (LLM) address this problem being effective in extracting information, but they may overlook correlations between events. On the other hand semantic clustering methods such as BERTopic and LDA, despite their strengths, tend to hide objects related to a specific event (actors, locations, objects) and can create distortions related to object binding [5–7]. For example,

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: UrbanAI: Harnessing Artificial Intelligence for Smart Cities Workshop.

when using BERTopic HDBSCAN to embed or cluster documents, priority is given to contextual similarity rather than preserving explicit entity relationships [8].

2 Methodology

2.1 Proposed approach

To address the limitations highlighted in section 1.1, especially for urban monitoring where actors, infrastructure, places, and times are central for decision-making, we propose KED-method. It detects events using temporal knowledge graphs (TKG) and performs community detection over multimodal (semantic and geographic) edges of TKGs. In our approach we use the combination of methods, which allow us to extract the additional textual characteristics. We refer to our approach as Knowledge-based Event Detection (KED). Figure 1 illustrates the overall pipeline of the method, which integrates three phases:

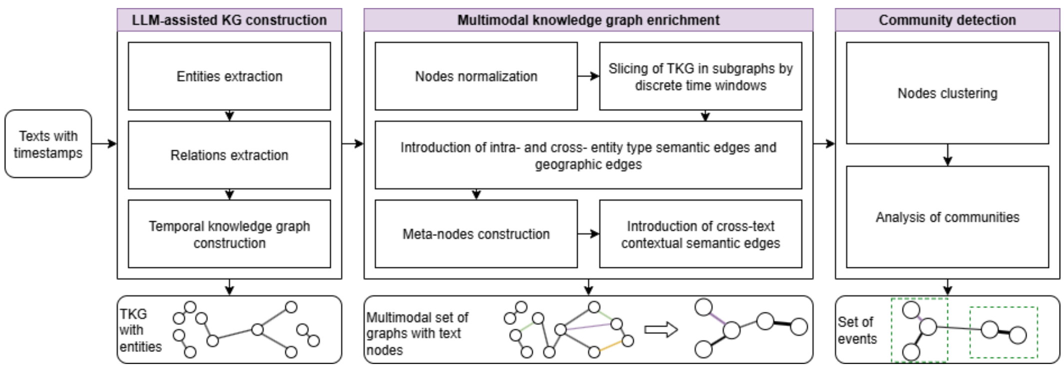

Figure 1: Text-to-event KED pipeline.

Phase I: LLM-assisted Knowledge Graph (KG) construction
Construct the initial knowledge graph (KG): Extract entities and relations from input texts using a custom system prompt and the LLM model (DeepSeek-32B).
Construct TKG: Build Temporal Knowledge Graph (TKG), where edges (u, v) are timestamped (u, v, t) edges from KG using metadata from source texts.
Phase II: Multimodal knowledge graph normalization and enrichment
Normalize Nodes: Apply lowercase conversion and remove hashtags. Use a union-find algorithm to merge redundant nodes, resulting in a simplified graph $G'(V', E')$.
Slice into Time Windows: Partition the TKG into subgraphs $G_t = (V_t, E_t)$, each covering a specific time window $[t, t + \Delta t]$.
Add Semantic Edges: For entities of specified types $(c1, c2)$, create edges if their embedding similarity $s(v_p, v_q)$ exceeds a threshold $\theta_{c1,c2}$. Weight is assigned as $w = \lambda_{c1,c2} \cdot s(v_p, v_q)$.
Add Geospatial Edges: Connect location nodes if their great-circle distance $d(v_p, v_q)$ is below a maximum d_{max}. Weight is assigned as $w = \beta(1 - \frac{d(v_p,v_q)}{d_{max}})$, parameter values are given below.
Construct Meta-Nodes V_M from nodes V of TKGs: Group entity nodes in TKG that co-occur in the same source text into higher-level meta-nodes, forming an aggregated meta-KG $G^* = (V_M, E_M)$.
Add Contextual Semantic Edges: Create edges between meta-nodes (texts) across different temporal subgraphs based on semantic similarity.
Phase III: Parametric Community Detection for Event Detection
Perform Clustering: Apply a community detection algorithm (e.g., Louvain, Leiden methods) to each TKG with meta-nodes to identify densely connected groups of text nodes.
Analyze Communities: Evaluate communities based on metrics like density and cohesion. The resulting partitions $C_{t,k}$ represent detected events.

We tune key hyperparameters (e.g., resolution, similarity thresholds) to probe robustness under distributional shift and class imbalance, a known challenge in ED [9]. The design is also compatible with KG embeddings for multi-modal fusion, enabling richer edge construction when appropriate [10].

3 Data Description

We chose Event2012 dataset [11] since it provides established benchmarks for fair comparison, allows testing imbalance handling, as well as the generalisation. The dataset was accessed using the SocialED library [12]. To study the proposed approach, standard datasets for event extraction algorithms were used. Evaluation utilized two Event2012 corpus derivatives: Dataset 1 ($20K$ chronologically sampled messages, highly imbalanced) and Dataset 2 ($4,833$ balanced messages, < 10 per event). Benchmarks included BERTopic, EventX [13] for real-time clustering, GNN-based methods (KPGNN [14], ETGNN [15]), and privacy-preserving models (ADPSEMEvent) [16]. Datasets are available via the SocialED repository [12]

4 Results

In Figure 2 we compare the evaluation metrics for our method and the baseline algorithms summarising the metrics averaged over temporal slices in the comparison with other baseline methods. Performance of the method was quantified via Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and Adjusted Mutual Information (AMI), with hyperparameters optimized through two-phase grid search.

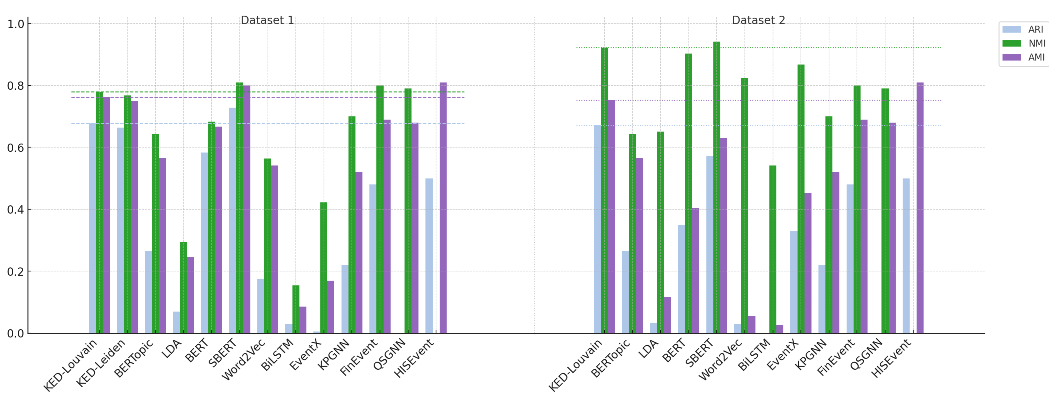

Figure 2: Comparison of average metrics between the proposed KED methods and other ED methods.

We re-evaluated the main methods on our reduced Event2012 dataset; the metrics for KPGNN [14], FinEvent [17], QSGNN [18], and HISEvent [19] are taken from the respective research papers that considered the same dataset. The final hyper-parameters for KED-Louvain are as follows: the temporal knowledge graph is segmented into disjoint weekly windows; the Louvain resolution parameter is set to 0.8; geospatial edges are included with a maximum distance threshold of 300 km and a weight β of 0.5; semantic edges between nodes are created when the semantic similarity $\theta_{c1,c2}$ exceeded 0.8; contextual edges are added with a $\mathrm{sim}_{\mathrm{thr}}$ of 0.5 and an α-weight of 0.6.

4.1 Key Findings

KED-Louvain achieved close to state-of-the-art results on imbalanced data (Dataset 1), attaining NMI 0.78, AMI 0.76 and ARI 0.68. Weekly disjoint windows outperformed daily slices (Δ NMI=+0.12) by mitigating sparse connectivity. On balanced data (Dataset 2), KED matched SBERT's [20] NMI (0.9) but exhibited superior structural coherence (ΔARI=+0.04). Both our method and the SBERT-based baseline exceed the 0.9 threshold on the balanced dataset, yet KED attains slightly higher scores on the other two evaluation metrics. The highest scores were obtained not for disjoint daily windows, but for disjoint weekly windows or for sliding windows 7–10 days wide with a 3-day stride. This dichotomy reflects Dataset 1's need for broader context to connect sparse event mentions, whereas Dataset 2 benefits from finer temporal resolution.

4.2 Ablation studies

The ablation framework systematically assessed changes in NMI under two scenarios: disabling one or several edge families; and removing individual node types when generating intra- and cross-type semantic edges. The key ablation findings are summarised in Table 1. Ablation studies revealed the importance of contextual edges (text-text similarity): their exclusion reduced NMI by 0.137 (for KED-Louvain) and 0.133 (for KED-Leiden). Geospatial edges provided secondary gains (ΔNMI $\approx +0.026$), with pronounced impact in localized events. Notably, toponyms and infrastructure nodes contributed more significantly to accuracy rather than actors or organizations.

Table 1: Key ablation results.

Ablation Variant	ΔNMI (Louvain)	ΔNMI (Leiden)
Contextual edges	-0.137	-0.133
Contextual and intra-type semantic edges	-0.112	-0.133
Contextual and geospatial edges	-0.106	-0.094
Contextual, geospatial and intra-type semantic edges	-0.096	-0.096
Semantic edges between Toponym type nodes	-0.028	-0.053
Geospatial edges	-0.026	-0.010
Semantic edges between Object type nodes	-0.020	-0.018
Semantic edges between Infrastructure type nodes	-0.019	-0.024

4.3 Case Studies

Week 1's even message distribution enabled near-perfect detection (NMI> 0.9), with clusters cleanly separating events. In contrast, Week 2's dominant Obama-Romney debate (40% of messages) caused semantic overlap, reducing NMI to 0.65. Here, KED identified Romney- and Obama-centric sub-events through centrality analysis, demonstrating its capacity to reveal nested event structures despite data being skewed. Visualizations for Week 1, Week 2 and Event 8 with Obama-Romney debates are presented in Figure 3, Figure 4 and Figure 5 respectively, Appendix A.

5 Discussion and Future Work

KED's TKG backbone delivers critical interpretability by explicitly modeling event constituents—resolving the "who, where, when" ambiguity present in embedding-based methods. The enrichment pipeline further enables knowledge augmentation through external sources. Applying the KED-algorithm to the unbalanced Dataset 1 we see that the majority topics dominate the graph structure. This is related to the Louvain method to merge with the minority topics into larger clusters, hence lowering the accuracy of the KED method in comparison with smaller balanced Dataset 2 which we presented in case study. Comparisons between ED methods are often based not only on absolute accuracy, but also on how performance depends on the amount of labeled data available [21]. While BERTopic/SBERT cluster texts by overall semantic similarity, KED's TKG structure explicitly preserves minority event entities. This prevents rare events from being drowned out by dominant topics.

At the same time there are three limitations in this method, first, LLM extraction demands significant computational resources. Second, global datasets dilute the impact of geospatial edges. Third, prominent entities (e.g., "Obama") can disproportionately influence graph topology, which we plan to address in the future work. Further directions include integrating Wikidata and OpenStreetMap to enhance relation extraction, optimizing graph construction for real-time streaming, and evaluating performance on domain-specific corpora like urban incident reports. The framework's modular design permits such extensions without the need to redeisgn the method itself.

6 Conclusion

KED method pioneers a structured paradigm for event detection by unifying temporal knowledge graphs with LLMs. KED framework demonstrates distinct advantages over three major classes of the event detection methods on imbalanced data while providing explainable, adaptable event representations. Its dynamic scaling and knowledge-augmentation capabilities establish a foundation for downstream tasks like causal inference and trend forecasting. Code implementation details are fully described for reproducibility (source code is demonstrated at https://github.com/Sandrro/KED-LLM).

Acknowledgments and Disclosure of Funding

This work supported by the Ministry of Economic Development of the Russian Federation (IGK 000000C313925P4C0002), agreement No139-15-2025-010

References

[1] Manuel Gilberto Garcia Alvarez. *GeoSmart cities: An event-driven approach for processing big geodata in smart cities*. Phd thesis, University of Twente, December 2024. URL `https://doi.org/10.3990/1.9789036564090`.

[2] Mikhail Kovalchuk, Anastasiia Filatova, Aleksei Korneev, Mariia Koreneva, Denis Nasonov, Aleksandr Voskresenskii, and Alexander Boukhanovsky. Semconvtree: Semantic convolutional quadtrees for multi-scale event detection in smart city. *Smart Cities*, 7:2763–2780, 09 2024. doi: 10.3390/smartcities7050107.

[3] S. Madeh Piryonesi. Incident management using SQL and textual data analytics. In *Proceedings of the CSCE Annual Conference 2024 (Niagara Falls, Canada)*, June 2024. URL `https://www.csce2024niagara.ca/`. Conference paper; official proceedings title/DOI not publicly available.

[4] Xiang Chen, Yi Li, Hao Zhang, and Lei Wang. Research review of the knowledge graph and its application in power system dispatching and operation. *Frontiers in Energy Research*, 10:896836, 2022. doi: 10.3389/fenrg.2022.896836.

[5] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure, 2022. URL `https://arxiv.org/abs/2203.05794`.

[6] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[7] Xenoss. Topic modeling for business (part 2): Comparing topic modeling approaches. `https://xenoss.io/blog/topic-modeling-for-business-part-2-comparing-topic-modeling-approaches`, 2025. Accessed 2025-08-20.

[8] Leland McInnes, John Healy, and Steve Astels. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11):205, 2017. doi: 10.21105/joss.00205.

[9] Ahmad Mousavi, Yeganeh Abdollahinejad, Roberto Corizzo, Nathalie Japkowicz, and Zois Boukouvalas. E-catch: Event-centric cross-modal attention with temporal consistency and class-imbalance handling for misinformation detection, 2025. URL `https://arxiv.org/abs/2508.11197v1`.

[10] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *The Semantic Web – ESWC 2018*, volume 10843 of *Lecture Notes in Computer Science*, pages 593–607. Springer, 2018. doi: 10.1007/978-3-319-93417-4_38.

[11] Andrew J. McMinn, Yashar Moshfeghi, and Joemon M. Jose. Building a large-scale corpus for evaluating event detection on twitter. In *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM)*, pages 409–418. ACM, 2013. doi: 10.1145/2505515.2505695.

[12] RingBDStack. Socialed: Social event detection. `https://github.com/RingBDStack/SocialED`, 2023.

[13] B. Liu, J. Liu, H. Wang, and X. Li. Extracting events accurately from vast news corpora and organize events logically: Eventx. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM)*, pages 457–465. ACM, 2020. doi: 10.1145/3377939.3377874. URL `https://doi.org/10.1145/3377939.3377874`.

[14] Yuwei Cao, Hao Peng, Jia Wu, Yingtong Dou, Jianxin Li, and Philip S. Yu. Knowledge-preserving incremental social event detection via heterogeneous gnns. In *Proceedings of The Web Conference 2021*

(WWW '21), pages 3383–3395, New York, NY, USA, 2021. Association for Computing Machinery. doi: 10.1145/3442381.3449834. URL `https://doi.org/10.1145/3442381.3449834`.

[15] Jiaqian Ren, Lei Jiang, Hao Peng, Zhiwei Liu, Jia Wu, and Philip S. Yu. Evidential temporal-aware graph-based social event detection via dempster-shafer theory. In *Proceedings of the 2022 IEEE International Conference on Web Services (ICWS)*, pages 637–644. IEEE, 2022. doi: 10.1109/ICWS55610.2022.00055. URL `https://ieeexplore.ieee.org/document/9885765`.

[16] Zhiwei Yang, Yuecen Wei, Haoran Li, Qian Li, Lei Jiang, Li Sun, Xiaoyan Yu, Chunming Hu, Hao Peng, and Philip S. Yu. Adaptive differentially private structural entropy minimization for unsupervised social event detection, 2024. URL `https://arxiv.org/abs/2407.18274`.

[17] Hao Peng, Ruitong Zhang, Shaoning Li, Yuwei Cao, Shirui Pan, and Philip S. Yu. Reinforced, incremental and cross-lingual event detection from social messages. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):980–998, 2023. doi: 10.1109/TPAMI.2022.3144993. URL `https://doi.org/10.1109/TPAMI.2022.3144993`.

[18] Jiaqian Ren, Lei Jiang, Hao Peng, Yuwei Cao, Jia Wu, Philip S. Yu, and Lifang He. From known to unknown: Quality-aware self-improving graph neural network for open set social event detection. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1696–1705. Association for Computing Machinery, 2022. doi: 10.1145/3511808.3557329. URL `https://doi.org/10.1145/3511808.3557329`.

[19] Yuwei Cao, Hao Peng, Zhengtao Yu, and Philip S. Yu. Hierarchical and incremental structural entropy minimization for unsupervised social event detection. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI)*, pages 8255–8264. AAAI Press, 2024. URL `https://ojs.aaai.org/index.php/AAAI/article/view/28666`. Accessed 2025-08-20.

[20] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Proceedings of EMNLP 2019*, pages 3980–3990. Association for Computational Linguistics, 2019. doi: 10.18653/v1/D19-1410.

[21] Shumin Deng, Ningyu Zhang, Luoqiu Li, Hui Chen, Huaixiao Tou, Mosha Chen, Fei Huang, and Huajun Chen. Low-resource event detection with ontology embedding. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*, Online, 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.220. URL `https://aclanthology.org/2021.acl-long.220`.

A Appendix / supplemental material

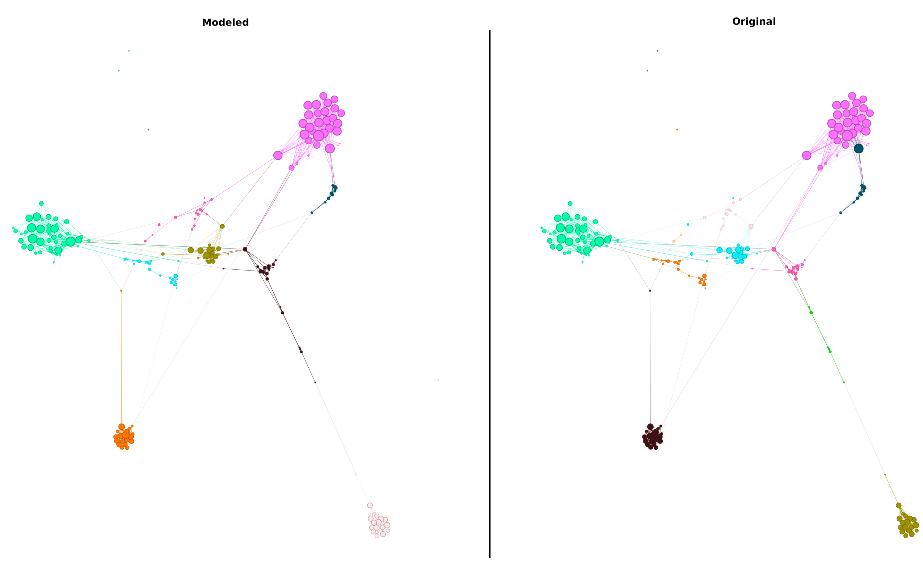

Figure 3: The illustration of the KGs constructed using KED event detection method applied to Dataset 1, Week 1 (left) and the ground-truth data labels (right). Various events are marked in different colors, yet each event is not associated with specific color.

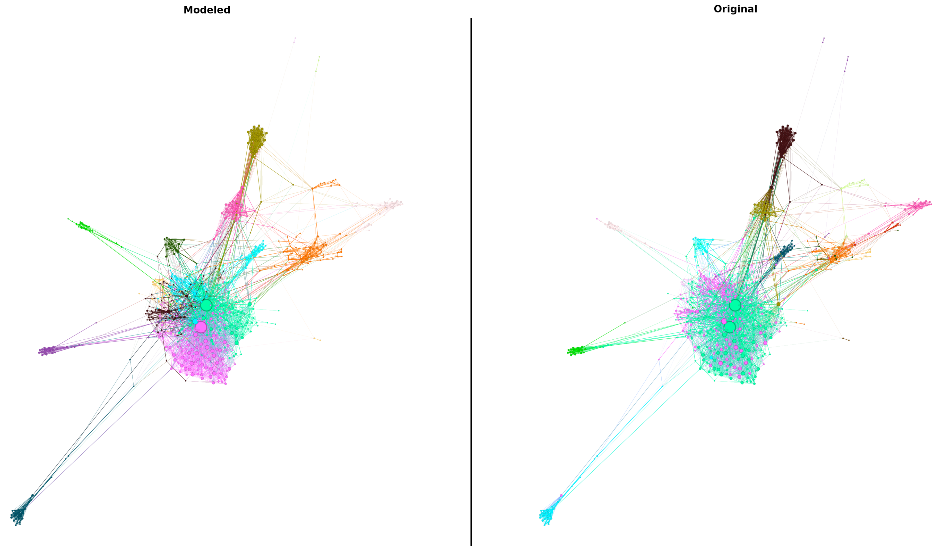

Figure 4: The illustration of the application of KED event detection method applied to the KG constructed from Dataset 1, Week 2.

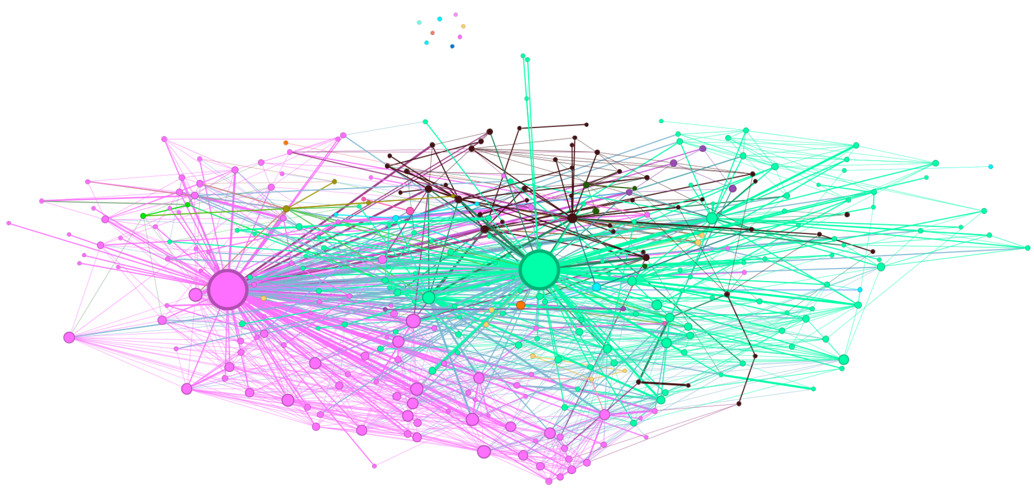

Figure 5: The illustration of the missclassification results by the KED method: some texts, such as Event 8 from Dataset 1, are attributed by KED to several events as the graph is split into several communities, hence lowering KED performance in comparison to SBERT method, Figure 2.