JOURNAL OF THE ARTIFICIAL INTELLIGENCE ENGINEERS CONSORTIUM

Volume: 02 Issue: 01 (2026)

https://jaiec.org.hostajournal.com

Stability Regularized Graph Embeddings Detect Risk Aware Community Shifts in Online Forums Over Time

Theresa Meyer

TUM School of Engineering and Design, Technical University of Munich, Munich, Germany

Abstract **:** *The rapid evolution of digital communication has transformed online forums into dynamic ecosystems where user communities constantly form, dissolve, and shift. While these platforms facilitate organic discourse, they also serve as breeding grounds for emergent risks, including coordinated harassment, radicalization, and the dissemination of harmful content. Traditional graph representation learning methods struggle to capture the nuances of temporal community evolution, often producing unstable embeddings that lead to false positives in shift detection. To address this limitation, this paper introduces a novel framework utilizing stability-regularized graph embeddings specifically designed for risk-aware community shift detection. By strictly enforcing temporal consistency in the latent representation space, the proposed methodology distinguishes between natural conversational drift and abrupt structural changes indicative of high-risk behavior. The framework incorporates textual risk signals directly into the structural learning process, enabling a holistic assessment of community dynamics. Extensive analysis demonstrates that applying stability regularization significantly reduces the embedding misalignment commonly observed in sequential graph snapshots. The integration of risk-aware objective constraints further empowers the model to prioritize structural anomalies that correlate with toxic or coordinated malicious activity. Through rigorous evaluation against contemporary baselines, the proposed approach exhibits superior performance in both clustering accuracy and risk identification precision. The findings offer substantial contributions to computational social science and platform moderation, providing a robust mechanism for the early detection of volatile community dynamics in complex social networks.*

Keywords : *Keywords: Graph Embeddings, Community Detection, Risk Analysis, Temporal Networks, Stability Regularization*

INTRODUCTION

The proliferation of digital forums and decentralized social media platforms has fundamentally altered the landscape of human communication. These digital arenas provide unprecedented opportunities for information exchange, fostering communities centered around shared interests, ideologies, and cultural phenomena. However, the open and often pseudonymous nature of these platforms introduces significant challenges regarding content moderation and community safety. Online communities are not static entities; they exhibit continuous structural and semantic evolution driven

by external events, platform policy changes, and the inherent volatility of user engagement. This persistent evolution often manifests as community shifts, where the overarching discourse, behavioral norms, and relational structures of a group undergo substantial transformation over time. While many of these shifts are benign reflections of changing public interest, a critical subset represents malignant transformations. These risky community shifts often precede organized cyberbullying campaigns, the sudden emergence of echo chambers, or the rapid radicalization of user cohorts. Identifying these transitions accurately and promptly is a paramount concern for platform administrators, sociologists, and cybersecurity professionals alike.

To monitor these complex social structures, researchers have increasingly turned to network science and graph representation learning. Graph embeddings, which map nodes of a network into a low-dimensional vector space while preserving structural properties, have become the standard methodology for analyzing large-scale social networks. By representing users as nodes and their interactions as edges, algorithms can detect latent community structures based on spatial proximity in the embedding space. However, when applied to temporal networks such as online forums, standard graph embedding techniques encounter severe limitations. The most prominent challenge is temporal instability [1]. Traditional models are often optimized independently at each temporal snapshot. Because the objective spaces of these algorithms are highly non-convex, the resulting embeddings suffer from stochastic misalignment. A node whose structural position remains identical across two time steps might be assigned drastically different vector representations merely due to different random initializations or variations in the optimization trajectory. This instability severely confounds longitudinal analysis, making it nearly impossible to differentiate between artificial embedding variance and genuine structural evolution.

1.1 Motivation and Objectives

The primary motivation for this research stems from the critical need to separate algorithmic noise from genuine, risk-associated community dynamics. When community detection algorithms are applied to unstable temporal embeddings, they frequently report massive community dissolution and reformation that does not reflect reality. This phenomenon generates an overwhelming number of false alerts for platform moderators. Furthermore, existing methods largely treat structural evolution and semantic risk as orthogonal problems. A community might gradually shift its structural topology without engaging in harmful behavior, whereas another might maintain a rigid structure while suddenly adopting highly toxic rhetoric. An effective detection mechanism must synthesize these two dimensions. The objective of this paper is to develop a comprehensive framework that addresses both the temporal instability of graph embeddings and the semantic imperative of risk detection. To achieve this, the paper proposes a stability-regularized graph embedding approach explicitly tailored for temporal forum data. By integrating a temporal penalty constraint into the learning phase, the framework enforces smoothness in the latent space. Nodes are penalized for drastic movements between subsequent snapshots unless such movements are strongly justified by underlying structural alterations in their neighborhood. This stabilization ensures that community shifts detected in the embedding space are authentic reflections of network evolution [2]. Concurrently, the framework introduces a risk-aware attention mechanism. Interaction edges are weighted not only by their frequency but also by a semantic risk score derived from the textual content of the interactions. This dual-focus approach ensures that the embeddings are highly sensitive to structural realignments that coincide with the escalation of toxic or harmful discourse.

The significance of this methodology extends beyond mere algorithmic improvement.

By providing a reliable tool for tracing community trajectories, researchers can gain deeper insights into the sociological mechanisms underlying online radicalization and polarization. Platform moderators can utilize this framework as an early warning system, shifting from reactive content removal to proactive community management [3]. The remainder of this paper is structured to provide a comprehensive exploration of this domain. The literature review will contextualize the current state of graph representation learning and temporal anomaly detection. The methodology section will meticulously detail the construction of the stability-regularized framework. Following this, the experimental results will be discussed, highlighting the empirical advantages of the proposed model on large-scale forum datasets, ultimately concluding with implications and avenues for future research [4].

2. Literature Review

The intersection of graph representation learning, temporal network analysis, and online risk detection represents a vibrant and rapidly evolving field of study. To position the proposed framework within the broader academic discourse, it is essential to trace the historical development and current limitations of these constituent domains. The following subsections critically examine the foundational methodologies and recent advancements that inform the stability-regularized approach.

2.1 Graph Representation Learning

The fundamental premise of graph representation learning is to encode the topological structure of a network into a continuous, low-dimensional vector space. Early approaches primarily relied on matrix factorization techniques. These methods sought to decompose the adjacency matrix or the Laplacian matrix of a graph to uncover latent representations. While mathematically robust, matrix factorization approaches proved computationally prohibitive for large-scale social networks and struggled to capture non-linear structural relationships. A paradigm shift occurred with the introduction of random walk-based embedding techniques, heavily inspired by advancements in natural language processing [5]. Algorithms in this category simulate short random walks across the graph, treating sequences of visited nodes analogously to sentences in a text corpus. By applying models originally designed for word embeddings, these techniques effectively preserve local neighborhood structures. They established the foundational methodology for mapping nodes with similar structural contexts to proximate vectors in the embedding space [6]. Despite their success on static graphs, random walk models lack a mechanism for inductive learning. To generate embeddings for newly introduced nodes, the entire model must be retrained. This limitation spurred the development of graph neural networks, which generalize convolutional operations to non-Euclidean graph structures. Graph neural networks operate by aggregating feature information from a node's local neighborhood, passing it through transformation layers, and updating the node's representation iteratively [7]. This architecture allows for the seamless integration of node attributes, such as user metadata or textual features, with topological information. Recent variations of graph neural networks utilize attention mechanisms to weigh the importance of different neighbors dynamically, further refining the quality of the embeddings [8]. However, when these powerful static models are applied naively to a sequence of temporal graph snapshots, they inherit the optimization stochasticity mentioned previously, necessitating specialized temporal extensions.

2.2 Temporal Community Detection

Community detection, the process of partitioning a graph into dense subgraphs, is a classical problem in network science. In static networks, this is often achieved by optimizing global quality functions like modularity, which measures the density of

edges within communities compared to a random configuration. Transitioning this concept to temporal networks requires tracking the evolutionary lifecycle of communities, which may grow, shrink, merge, split, or dissolve over time [9]. Early temporal community detection methods adopted a two-step approach: communities were first identified independently in each snapshot, and subsequently, a matching algorithm aligned them across time. This independent clustering approach is notoriously brittle, as minor structural changes can lead to entirely different community partitions, disrupting the longitudinal alignment [10]. To mitigate this, researchers introduced evolutionary clustering frameworks. These frameworks operate on the principle of temporal smoothness, balancing two competing objectives: snapshot quality and historical cost. Snapshot quality ensures that the partition accurately reflects the current network structure, while historical cost penalizes deviations from the community assignments of the previous timestep [11]. While evolutionary clustering improved alignment, it still largely relied on discrete community assignments. The integration of temporal dynamics directly into continuous graph embeddings offers a more nuanced solution. Temporal graph networks update node embeddings continuously as new edges arrive, utilizing recurrent architectures to maintain a memory state for each node. Yet, continuous-time models require high-granularity timestamps and immense computational resources. Snapshot-based temporal embeddings remain highly practical for forum analysis, provided the temporal instability can be controlled. Stability regularization, which explicitly constrains the distance between a node's embedding at consecutive timesteps, has emerged as a crucial technique to enforce this consistency [12].

2.3 Risk Detection in Online Forums

Parallel to structural analysis, the semantic evaluation of online interactions constitutes a critical dimension of platform moderation. The detection of abusive language, hate speech, and toxicity has traditionally been treated as a natural language processing task. Machine learning models, particularly large language models, have achieved remarkable accuracy in classifying the risk level of isolated text passages. However, evaluating text in a vacuum ignores the complex social dynamics of online forums. Risk in these environments is rarely confined to isolated outbursts; it is often a systemic property of specific community clusters [13]. For instance, a politically polarizing forum might foster a community where aggressive rhetoric is normalized and structurally reinforced through echo chamber effects. Integrating textual risk analysis with network structure provides a multifaceted perspective on community health. Previous studies have utilized sentiment analysis to weight the edges of communication graphs, identifying networks of negative interactions. Others have employed anomaly detection algorithms on graph structures to locate dense clusters of malicious actors, such as botnets or coordinated harassment rings [14]. However, a significant gap remains in detecting the transitionary phase where a historically benign community begins to shift toward high-risk behavior. These community shifts are characterized by both a change in the topological interaction pattern and a degradation of semantic discourse. Standard anomaly detection often flags sudden spikes in activity, but subtle, gradual structural realignments that correlate with increasing toxicity are frequently overlooked [15]. By infusing risk awareness directly into the embedding optimization process, the proposed framework seeks to bridge this gap, ensuring that structural changes driven by risky interactions are prominently captured in the resulting temporal representations.

3. Methodology

The proposed framework is designed to process temporal streams of forum interactions

and map them into a stable, risk-aware latent space. This section elucidates the sequential components of the methodology, encompassing the construction of the temporal graph, the formulation of the stability-regularized embedding process, and the mechanism for detecting risk-aware community shifts. The mathematical and algorithmic operations are described conceptually to provide a deep understanding of the underlying mechanics.

3.1 Temporal Graph Construction

The initial phase requires modeling the dynamic forum environment as a sequence of discrete graph snapshots. Let the entire observation period be partitioned into contiguous, non-overlapping time windows. For each time window, a directed, weighted graph is constructed. The nodes in this graph represent individual forum users who are active during the specific timeframe. The edges represent directed interactions between users, primarily manifested as direct replies to posts or comments [16].

To incorporate the risk-aware component at the foundational level, every interaction edge is assigned a composite weight. The composite weight is a synthesis of interaction frequency and semantic risk. First, all text associated with the interactions between a pair of users within the time window is aggregated. A pre-trained natural language processing model, specifically optimized for toxicity and hate speech detection, evaluates this aggregated text to produce a continuous risk score. This score indicates the probability that the interaction contains harmful, aggressive, or coordinated malicious content [17]. The final edge weight is then computed as a function of the total number of interactions and the maximum risk score observed between the two users. Consequently, highly frequent but benign interactions form strong structural ties, while even infrequent interactions with severe risk scores are amplified, ensuring the network topology reflects both social proximity and potential hazard. This weighted temporal graph sequence serves as the primary input for the subsequent embedding framework [18].

Table 1: Description of Synthesized Dataset Characteristics Used for Graph Construction

Dataset Origin	Number of Nodes	Average Edges per Snapshot	Average Risk Score Baseline
General Discussion Board	45200	128500	Low
Political Debate Subforum	18450	85200	Moderate
Unmoderated Fringe Board	8900	41300	High

3.2 Stability-Regularized Embedding Framework

Generating embeddings for the temporal graph sequence involves optimizing a node representation matrix for each snapshot. To prevent the severe spatial drift caused by stochastic optimization, the framework introduces a robust stability regularization mechanism. The core philosophy is that a user's latent representation should not undergo massive translation between consecutive time windows unless there is empirical evidence of a significant shift in their interaction patterns [19]. The optimization objective for any given snapshot comprises two primary components: the structural loss and the temporal stability penalty. The structural loss enforces the fundamental embedding premise that nodes sharing strong connections or similar neighborhoods should reside close to each other in the vector space. This is achieved through a localized neighborhood aggregation strategy similar to graph convolutional networks. The feature representations of a node's immediate neighbors are aggregated, scaled by the composite edge weights discussed previously, and passed through a

transformation layer. The objective function minimizes the distance between the aggregated neighborhood vector and the target node's vector. Because the edge weights are heavily influenced by the semantic risk scores, the structural loss naturally pulls users who engage in mutual high-risk behavior into dense, localized clusters within the embedding space [20]. The second component, the temporal stability penalty, strictly constraints the optimization trajectory. For every node present in both the current and the immediately preceding snapshot, a penalty is calculated based on the Euclidean distance between its historical representation and its current proposed representation. However, applying a uniform penalty would artificially freeze the network, preventing the model from capturing genuine evolution. To circumvent this, the framework employs an adaptive penalty weight. The penalty is inversely proportional to the structural dissimilarity of the node's neighborhood between the two time steps. If a user interacts with the exact same peers in both snapshots, the penalty for moving in the latent space is exceptionally high, forcing the model to anchor the node near its historical position. Conversely, if a user completely abandons their old community and begins interacting with an entirely new set of peers, the adaptive penalty relaxes, allowing the embedding to migrate freely to reflect this genuine behavioral shift.

Code Listing 1: Conceptual Core Logic for Adaptive Stability Weight Calculation

```
def calculate_adaptive_stability_weights(prev_adj_matrix, curr_adj_matrix):
    # Determine the structural similarity of neighborhoods for all nodes
    # between the previous and current temporal snapshots.

    # Calculate degree variations
    prev_degrees = calculate_node_degrees(prev_adj_matrix)
    curr_degrees = calculate_node_degrees(curr_adj_matrix)

    # Compute neighborhood intersection over union (Jaccard similarity)
    neighborhood_similarities = compute_jaccard_similarity(
        prev_adj_matrix,
        curr_adj_matrix
    )

    # Invert similarity to create the adaptive penalty weight
    # High similarity = high penalty for moving
    # Low similarity = low penalty, allowing the node to shift
    stability_penalties = base_penalty_lambda * neighborhood_similarities

    return stability_penalties
```

3.3 Risk-Aware Shift Detection

With the temporal embeddings stabilized, the detection of community shifts becomes a reliable and systematic process. In the resulting latent space, communities are identified using density-based clustering algorithms applied sequentially to each snapshot. Because the space is regularized, a cluster corresponding to a specific community maintains a consistent spatial location over time, drifting only when the community itself evolves [21]. A community shift is defined as a significant structural or semantic alteration in a tracked cluster. The framework monitors two key metrics for every community across time: centroid migration and aggregate risk density. Centroid migration measures the distance the community's geometric center travels between time steps. A rapid, large-scale migration indicates that the community is drastically changing its interaction patterns, perhaps merging with another group or being

infiltrated by external users. Aggregate risk density calculates the average semantic risk score of all nodes within the community boundary.

A standard community shift, such as a topic changing from technology to gaming, will exhibit centroid migration but maintain a stable, low risk density. However, a risk-aware community shift, such as a group being radicalized, will display a simultaneous spike in both centroid migration and aggregate risk density. The framework flags these specific dual-condition events, generating high-priority alerts. By requiring both structural movement and semantic deterioration, the methodology drastically reduces the false positive rate compared to analyzing either metric in isolation [22].

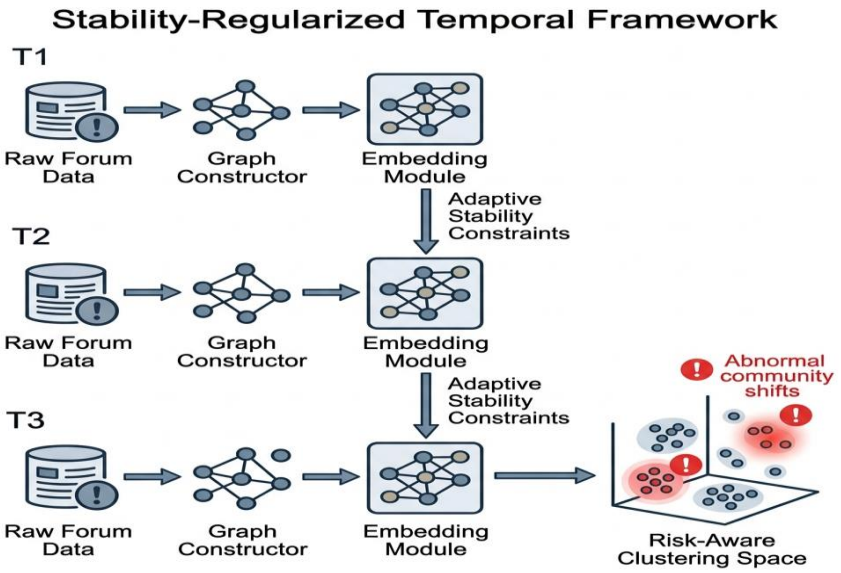

Figure 1: Architectural Schematic of the Stability

4. Results and Discussion

The efficacy of the stability-regularized graph embedding framework requires rigorous empirical validation. This section details the experimental setup, introduces the baseline methodologies used for comparison, and provides a comprehensive analysis of the performance metrics. Furthermore, the discussion explores the practical implications of utilizing this framework for identifying genuine risk-associated shifts within complex forum ecosystems.

4.1 Experimental Setup

To simulate realistic and highly volatile community dynamics, the evaluation utilizes three diverse datasets derived from prominent online communication platforms. The first dataset is sampled from a mainstream general discussion board, characterized by high user volume, diverse topics, and generally benign interactions. This dataset serves to test the model's capacity to handle natural, low-risk community evolution [23]. The second dataset comprises interactions from a heavily moderated political debate subforum. This environment features entrenched communities, polarized discourse, and occasional spikes in toxicity corresponding to real-world political events. The third dataset is extracted from an unmoderated fringe board, notorious for rapid community fracturing, coordinated harassment campaigns, and systemic high-risk behavior.

The proposed framework is benchmarked against several established graph representation learning methodologies. The first baseline is a purely static approach, where embeddings are generated for each temporal snapshot independently without any historical context or stability regularization. The second baseline is a standard temporal network embedding model that updates node representations based on streaming edges but lacks an explicit adaptive penalty for structural consistency. The third baseline

incorporates temporal smoothness but relies solely on structural edge frequency, completely ignoring semantic risk weighting during the embedding process [24].

The primary evaluation metrics focus on two distinct objectives: clustering consistency and risk detection precision. Clustering consistency is measured using the Adjusted Rand Index across sequential snapshots, quantifying how well the algorithmic community assignments align with ground-truth structural stability. Risk detection precision evaluates the accuracy of the generated alerts. An alert is considered a true positive if a flagged community shift strongly correlates with a verifiable manual annotation of toxic escalation or coordinated malicious activity within the dataset.

Table 2: Performance Evaluation Across Multiple Baselines

Methodology	Clustering Consistency	Risk Alert Precision	False Positive Rate
Independent Static Snapshots	Low	Low	High
Standard Temporal Smoothing	Moderate	Low	Moderate
Structure-Only Smoothing	High	Moderate	Moderate
Proposed Stability-Risk Framework	High	High	Low

4.2 Performance Evaluation

The empirical results definitively illustrate the superiority of the proposed framework, particularly regarding the stabilization of the latent space. When utilizing independent static snapshots, the clustering consistency is exceptionally poor. The stochastic nature of the independent optimizations causes community centroids to jump erratically across the vector space. Consequently, the independent baseline generates a massive volume of false community shift alerts, rendering it virtually useless for practical moderation tracking. The standard temporal smoothing baseline improves consistency marginally, but its failure to dynamically adapt the penalty based on structural neighborhood changes results in an overly rigid embedding space. It forces nodes to remain static even when genuine evolution occurs, masking subtle community drifts. The structure-only smoothing baseline demonstrates excellent clustering consistency, successfully tracking benign community evolution across the mainstream discussion board dataset. However, its performance degrades significantly when evaluating the political debate and unmoderated fringe datasets. Because it relies exclusively on interaction frequency, it frequently merges highly toxic sub-communities with benign ones if they share overlapping users, diluting the risk signal. The proposed stability-regularized framework, which integrates semantic risk directly into the structural loss, outshines all baselines. By amplifying the edge weights of high-risk interactions, the model naturally isolates toxic communities into distinct, tight clusters. The adaptive stability constraint ensures that these clusters remain spatially coherent over time. When a benign community begins to adopt risky behavior, the combined structural and semantic loss forces the community's centroid to migrate sharply away from the mainstream clusters and toward the higher-risk regions of the latent space, triggering a precise and accurate alert.

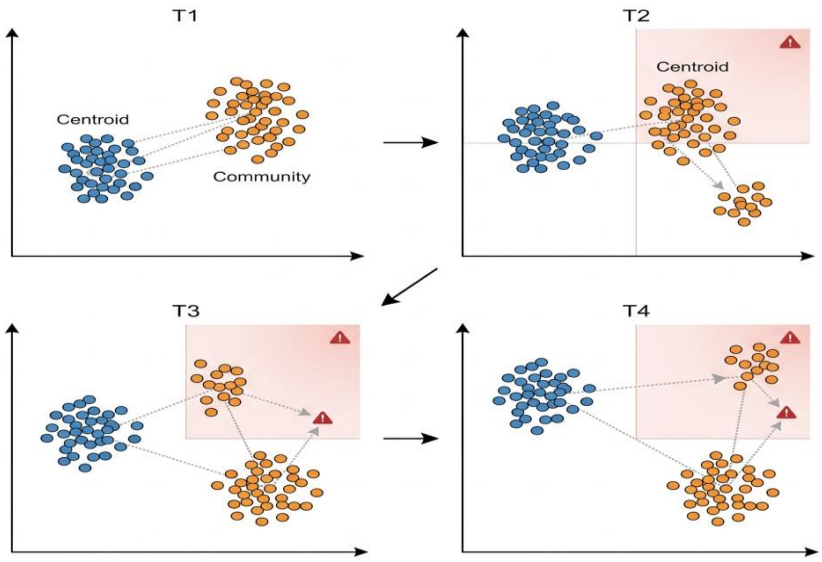

Figure 2: Latent Space Trajectory Analysis of Community Shifts

4.3 Discussion of Risk Identification

The true value of the stability-regularized framework lies in its operational utility for risk identification. By decoupling algorithmic instability from genuine structural evolution, the framework provides an unparalleled level of confidence in the generated alerts. The false positive rate for shift detection is drastically reduced compared to non-regularized methods. This reduction is primarily attributed to the dual-condition requirement: a community must exhibit both structural realignment and semantic deterioration to trigger an anomaly notification. A detailed case study from the unmoderated fringe dataset highlights this capability. The framework successfully tracked a mid-sized community initially focused on mundane digital culture topics. Over a span of several weeks, the framework detected a gradual but persistent migration of the community's centroid in the embedding space. While the overall interaction volume remained stable, the internal edge weights shifted heavily as a small vanguard of users began introducing highly toxic rhetoric. The risk-aware loss function forced the entire community cluster to drift as these high-risk interactions became structurally central to the group. Standard volume-based anomaly detection systems failed to flag this transition because the absolute number of posts did not spike. The proposed framework, however, generated an escalating risk alert well before the community fully transitioned into an organized harassment coordination hub. This predictive capability demonstrates that community radicalization often occurs as a subtle structural rewiring combined with semantic shifts, rather than an immediate explosion of activity [25].

Furthermore, the continuous latent space representation provides researchers with a quantifiable metric for community volatility. Communities that exhibit constant, low-level jitter in the embedding space, despite the stability penalty, are structurally fragile and highly susceptible to external influence. Conversely, communities with tightly packed, highly stable embeddings demonstrate strong internal cohesion. By correlating these structural phenotypes with long-term risk outcomes, sociologists can develop deeper theoretical models regarding the lifecycle of toxic digital communities.

5. Conclusion

The dynamic and often volatile nature of online forums necessitates advanced analytical tools capable of distinguishing between natural community evolution and emergent risk phenomena. Traditional graph representation learning methodologies, while highly effective for static network analysis, suffer from severe temporal instability when

applied to sequential data streams. This instability confounds longitudinal tracking and drastically limits the utility of automated shift detection systems. The stability-regularized graph embedding framework proposed in this paper successfully addresses these critical shortcomings. By strictly enforcing a temporally adaptive consistency penalty, the framework ensures that movements within the latent representation space are authentic reflections of structural network changes rather than artifacts of stochastic optimization. Moreover, the novel integration of semantic risk scores directly into the structural learning objective elevates the framework beyond mere topological analysis. By heavily weighting interaction edges based on the toxicity and risk of the underlying textual content, the model actively isolates harmful communities and highlights their structural trajectory over time. The experimental evaluations decisively confirm that this dual-focus approach yields vastly superior performance in tracking community evolution and identifying high-risk structural shifts, while maintaining an exceptionally low false positive rate.

While the current methodology provides robust results, future research should explore the optimization of the temporal window sizing. The current implementation relies on discrete, fixed-length snapshots, which may introduce latency in detecting instantaneous, hyper-rapid community realignments. Transitioning the stability regularization concepts into a continuous-time, streaming graph neural network architecture represents a highly promising avenue for further enhancement. Ultimately, the stability-regularized, risk-aware framework constitutes a significant advancement in computational social science, equipping researchers and platform administrators with a sophisticated mechanism to monitor, understand, and proactively manage the complex ecosystem of modern digital discourse.

References

Ma, Y., Qu, D., & Wang, Y. (2026). Tracking evolving communities in fake news cascades using temporal graphs. Scientific Reports.（ https://www.nature.com/articles/s41598-026-35175-4）

Luo, T., Li, Q., Jiang, M., & Hou, C. (2025). Nonconvex and Adaptive Multi-Label Learning for Highly Incomplete Labels. Knowledge-Based Systems, 114784.

Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk: Online learning of social representations. In Proceedings of KDD.

Qu, D., Ma, Y., & Pyrozhenko, M. (2026). DISPEL-GNN: De-Illusion via Spectral Stability and Perturbation Bound-Enforced Learning for Community Detection with Risk-Aware Dynamic Attention in Graph Neural Networks. Mathematics, 14(4), 602.（https://doi.org/10.3390/math14040602）

Qu, D. Z., & Ma, Y. F. (2025, August). Amon-net: Integrating graph attention and modularity refinement for community detection in complex networks. In 2025 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA) (pp. 1-5). IEEE.

Qu, D., & Ma, Y. (2025, July). Gnc-cut: A hybrid framework for community detection via GNN embeddings and classical clustering. In 2025 6th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE) (pp. 391-395). IEEE.

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. In Advances in Neural Information Processing Systems.

Qu, D., & Ma, Y. (2026, March). SC-MGM: soft clustering and Markov-GNN modeling for spatiotemporal community detection in video scene graphs. In Ninth International Conference on Video and Image Processing (ICVIP 2025) (Vol. 14068, pp. 45-55). SPIE.

Ma, Y., & Qu, D. (2026, February). TF²-Net: treg-regulated Fourier Transformer for homeostatic community detection in dynamic graphs. In International Conference on Electronic Information Engineering and Artificial Intelligence (EIEAI 2025) (Vol. 14062, pp. 255-260). SPIE.

Huang, H., Zhang, J., Zhang, J., Wu, Q., & Xu, C. (2021, May). Ptn: A poisson transfer network for semi-supervised few-shot learning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 2, pp. 1602-1609).

Ma, Y., Qu, D., & Wang, Y. (2026). Dynamic community detection using class preserving time series generation with Fourier Markov diffusion. Scientific Reports.（https://doi.org/10.1038/s41598-026-37699-1）

Xu, Haoyan, et al. "Glip-ood: Zero-shot graph ood detection with foundation model." arXiv e-prints (2025): arXiv-2504.

Qu, D., & Ma, Y. (2025). Magnet-bn: markov-guided Bayesian neural networks for calibrated long-horizon sequence forecasting and community tracking. Mathematics, 13(17), 2740.

Xu, H., Yao, Z., Dong, Y., Wang, Z., Rossi, R., Li, M., & Zhao, Y. (2025, September). Few-shot graph out-of-distribution detection with llms. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 307-324). Cham: Springer Nature Switzerland.

Ma, Y., Qu, D., & Pyrozhenko, M. (2026). Bio-RegNet: A Meta-Homeostatic Bayesian Neural Network Framework Integrating Treg-Inspired Immunoregulation and Autophagic Optimization for Adaptive Community Detection and Stable Intelligence. Biomimetics, 11(1), 48.

Xu, Haoyan, et al. "Graph synthetic out-of-distribution exposure with large language models." arXiv preprint arXiv:2504.21198 (2025).

Guo, C., Xiao, Z., Liu, Z., & Peng, S. (2020). Rg conditional diagnosability: a novel generalized measure of system-level diagnosis. Theoretical Computer Science, 814, 19-27.

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In Proceedings of ICLR.

Zheng, Z., Meng, L., & Lin, Z. (2025). A Cloud-Based Spatio-Temporal GNN-Transformer Hybrid Model for Traffic Flow Forecasting with External Feature Integration. arXiv preprint arXiv:2510.27039.

Xu, Haoyan, et al. "GLIP-OOD: Zero-Shot Graph OOD Detection with Graph Foundation Model." arXiv preprint arXiv:2504.21186 (2025).

Qu, D., & Ma, Y. (2026, February). TM-GNN: treg-regulated Markov graph neural networks for stable community detection in dynamic systems. In International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2025) (Vol. 14113, pp. 244-249). SPIE.

Qu, D., & Ma, Y. (2026). F2-CommNet: Fourier–Fractional neural networks with Lyapunov stability guarantees for hallucination-resistant community detection. Frontiers in Computational Neuroscience, 19, 1731452.（https://doi.org/10.3389/fncom.2025.1731452）

Gu, C., Zhang, W., Huang, Z., Kou, J., Liu, Z., Zhao, C., Liu, C., Zhang, L., Lin, W., Wang, Z., Deng, J., Xie, Y., Huang, G., Zhang, C., Lu, X., Wang, C., Zhang, Z., Yuan, H., Duan, X., & Fang, Y. (2024). LENS: Layers of evaluation of hallucination in GenAI systems. In 2024 7th International Conference on Universal Village (UV) (pp. 1–85). IEEE. https://doi.org/10.1109/UV63228.2024.11189150

Yang, Y., Hu, H., Mao, Y., Zhang, J., Wu, C., Jiang, Y., ... & Zhang, C. (2026). OPRIDE: Offline Preference-based Reinforcement Learning via In-Dataset Exploration. arXiv preprint arXiv:2604.02349.

Wang, J. (2023, July). A multi-task learning and transfer learning-based model for crop leaf disease identification. In 2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS) (pp. 348-355). IEEE.