JOURNAL OF SOCIAL SCIENCES, HUMANITIES AND INNOVATION

Vol 2 Issue 1 (2026)

http://jsshi.org/

Capturing Reviewer Behavioral Drift Over Time With Temporal Graph Networks for Adaptive Fake Review Detection

Olivia Martinez*

University of Maryland, Baltimore County, United States

Hannah Cooper

University of Maryland, Baltimore County, United States

Peter Collins

Syracuse University, United States

Corresponding author: Olivia Martinez

Keywords:

fake review detection, temporal graph networks, behavioral drift, dynamic graph learning, opinion spam, reviewer profiling

Abstract:

The proliferation of fake reviews on e-commerce and service platforms has severely undermined consumer trust and market integrity. Existing detection approaches predominantly rely on static graph structures or snapshot-based representations that fail to model the evolving behavioral patterns of individual reviewers over time. This paper proposes a Temporal Graph Network (TGN) based framework that captures reviewer behavioral drift through continuous-time dynamic graph modeling, enabling adaptive fake review detection. By encoding reviewer interaction histories, temporal embeddings, and relational evolution across heterogeneous review graphs, the proposed method learns time-aware representations that distinguish genuine reviewers from deceptive actors whose behavioral signatures shift in response to platform countermeasures. Experiments conducted on two benchmark datasets demonstrate that the proposed approach outperforms state-of-the-art baselines across precision, recall, and F1-score metrics. This work contributes a principled methodology for temporally adaptive fraud detection and highlights the utility of dynamic graph learning in combating evolving review manipulation strategies.

1.INTRODUCTION

The rapid expansion of online review ecosystems has transformed how consumers evaluate products, services, and businesses. Platforms such as Amazon, Yelp, TripAdvisor, and Google Reviews have become indispensable decision-making tools for billions of users worldwide, making the credibility of user-generated content a matter of substantial commercial and societal consequence. However, the high economic stakes attached to online reputation have simultaneously incentivized the systematic production of fraudulent reviews, a phenomenon that encompasses both artificially inflated positive assessments and maliciously fabricated negative content designed to damage competitors. The scale and sophistication of fake review operations have grown considerably over the past decade, with organized networks of human reviewers, semi-automated bots, and large language model-assisted content generators collectively flooding platforms with deceptive content that is difficult to distinguish from authentic user feedback [1].
Early attempts to address this problem relied heavily on linguistic analysis and sentiment inconsistency detection, drawing on natural language processing (NLP) techniques to identify syntactic or semantic anomalies that might betray artificial authorship. While these methods achieved reasonable precision in controlled settings, they proved brittle when deployed against adversarially adaptive spammers who routinely updated their linguistic strategies in response to detected patterns [2]. Subsequent research shifted toward behavioral profiling, recognizing that deceptive reviewers tend to exhibit detectable patterns in their rating distributions, posting frequencies, and temporal clustering behaviors. These behavioral signals offered a more stable basis for detection, since they were harder for spammers to suppress without compromising their operational effectiveness [3]. The introduction of graph-based detection frameworks marked a further maturation of the field, enabling models to exploit structural relationships among reviewers, products, and reviews rather than treating each entity in isolation [4]. Approaches grounded in belief propagation, graph neural networks (GNNs), and heterogeneous graph attention mechanisms demonstrated that relational signals could substantially improve detection accuracy [5]. Nevertheless, the overwhelming majority of graph-based methods operate under a static graph assumption, constructing a single fixed graph from historical data and learning embeddings that do not evolve as new reviews arrive or as reviewer behavior shifts over time [6]. This static paradigm represents a fundamental mismatch with the real-world dynamics of online review fraud. Fake reviewers are not passive agents whose behaviors are fixed at the time of account creation. They adapt strategically, altering their posting cadence, shifting target product categories, varying their rating patterns, and occasionally going dormant before resuming activity under different behavioral signatures [7]. A detection system trained on a frozen snapshot of the review graph will inevitably degrade in performance as the distribution of fraudulent behavior drifts away from the patterns captured at training time, a phenomenon commonly referred to as concept drift in the machine learning literature [8]. Addressing this challenge requires models capable of learning temporally evolving representations that track how individual reviewers' behavioral footprints change across the full lifespan of their platform activity [9]. Temporal graph networks (TGNs) offer a compelling solution to this challenge. Originally proposed for general dynamic graph learning tasks, TGNs maintain memory modules for each graph node that are continuously updated as new interaction events arrive, enabling the model to learn embeddings that encode not just the current structural neighborhood of a node but its entire temporal trajectory of interactions [10]. When applied to the review fraud detection domain, this architecture allows the system to capture how a reviewer's relational profile evolves over time, detecting the subtle behavioral drift signatures that distinguish long-term coordinated spammers from genuine users whose activity naturally fluctuates [11].

This paper presents a TGN-based framework specifically designed for adaptive fake review detection. The proposed model constructs a continuous-time dynamic review graph in which nodes represent reviewers and products, and temporally stamped edges encode individual review events. A memory-augmented temporal encoder tracks each reviewer's evolving embedding, while a graph attention module aggregates structural context from temporal neighbors. A drift-aware classification head uses the temporal embedding trajectory to identify reviewers whose behavioral signatures exhibit anomalous evolution consistent with coordinated fraud. Contributions of this work include a novel application of TGN architectures to fake review detection, a behavioral drift quantification module that operationalizes temporal embedding divergence as a fraud signal, and empirical validation on two public benchmark datasets demonstrating substantial improvements over static and semi-dynamic baselines.

2. Literature Review

Research into fake review detection has evolved through several distinct phases, each reflecting advances in both the understanding of spammer behavior and the methodological tools available for modeling it. Early work in this area treated the detection problem primarily as a text classification task, leveraging features derived from review content to distinguish genuine from deceptive reviews. Linguistic markers such as superlative density, first-person pronoun usage, and semantic deviation from verified purchase patterns were found to be moderately discriminative, and classifiers trained on these features achieved encouraging accuracy on curated datasets. However, the reliance on linguistic surface features made these systems vulnerable to adversarial adaptation, and the subsequent arms race between detectors and spammers motivated a shift toward behavioral and structural approaches. Behavioral profiling methods emerged as a more robust alternative, grounded in the observation that fake reviewers tend to exhibit quantifiably anomalous activity patterns regardless of the specific content they produce. Especially relevant here is the KE-MLLM framework of Chen et al., which convincingly integrates temporal activity patterns, reviewer metadata, social interaction signals, and contextual language representations within a single knowledge-enhanced model; their results show that this multi-sensor design improves both the interpretability and adaptability of fraud detection systems operating under evolving behavioral conditions, an outcome that resonates strongly with the temporally adaptive perspective adopted in this work [12]. Reviewers who post an unusually high volume of reviews within a short time window, who consistently rate products at the extreme ends of the scale, or whose review timing clusters suspiciously around product launch dates have been shown to differ systematically from genuine users in their statistical behavioral signatures [13]. Feature engineering approaches that captured rating deviation, burstiness, and target product overlap between reviewer pairs provided a foundation for ensemble classifiers and anomaly detection models that generalized better across platforms than purely linguistic systems [14]. The application of graph-based methods transformed the field by enabling detection models to exploit relational structure in addition to individual-level behavioral features. Review networks in which reviewers and products are represented as nodes and reviews as edges naturally encode information about co-review relationships, reviewer-product affinity patterns, and the structural proximity of suspicious actors [15]. Belief propagation algorithms deployed over these networks demonstrated that fraudulent intent could propagate through graph structure, enabling the detection of spammers who were individually inconspicuous but collectively anomalous in their relational footprints [16]. GNNs subsequently displaced hand-crafted propagation rules, learning end-to-end representations that fused node attributes with neighborhood aggregation signals in a fully differentiable framework [17]. Heterogeneous graph neural networks extended these capabilities to review ecosystems where different entity types — specifically reviewers, products, and reviews — interact through semantically distinct edge types. Models that maintained type-specific

propagation pathways and attention mechanisms across reviewer-to-product, reviewer-to-review, and product-to-category edges demonstrated improved performance on benchmark datasets, particularly in cases where the fraud signal was distributed across multiple relation types rather than concentrated in a single channel [18]. Attention-based aggregation further allowed these models to weight the contributions of individual neighbors dynamically, enabling them to suppress noise from legitimate co-review relationships while amplifying signals from structurally anomalous connections [19].

Despite these advances, the static graph assumption that underpins most GNN-based detection methods remained a significant limitation. Reviews accumulate over time, reviewer behaviors evolve, and the relational structure of the review graph changes continuously. Approaches that constructed static graphs from complete historical data and trained classifiers on fixed embeddings were not designed to accommodate these dynamics, and their deployment in live systems required periodic retraining on updated snapshots, a process that introduced both computational overhead and temporal gaps during which new fraudulent activity went undetected [20]. Semi-dynamic methods that partitioned the review history into discrete time windows and trained separate models on each window offered a partial remedy but introduced discontinuities at window boundaries and failed to capture gradual behavioral changes that unfolded across those boundaries [21]. The concept drift literature in machine learning provides a theoretical framework for understanding why static detection models degrade over time in adversarial environments. Fake review campaigns are explicitly adaptive, with operators modifying their behavioral strategies in response to platform countermeasures and algorithmic detection signals. This creates a non-stationary data distribution in which the statistical properties of fraudulent behavior shift continuously, rendering classifiers trained on earlier behavioral regimes progressively less effective against newer fraud patterns [22]. Drift detection algorithms borrowed from streaming data analysis have been incorporated into some fake review systems to trigger model updates when distributional shifts are detected, but these approaches treat drift as a nuisance to be managed rather than a signal to be exploited for fraud identification [23].

Temporal graph learning represents a fundamentally different paradigm that treats the evolving structure of interaction networks as a first-class modeling target. Dynamic graph embedding methods that learn low-dimensional representations of nodes as functions of their temporal interaction histories have been developed for a range of applications including social network analysis, knowledge graph completion, and recommendation systems [24]. Among these, TGNs stand out for their use of node-level memory modules that are updated continuously as new interaction events arrive, enabling the model to maintain a compressed representation of each node's full temporal history without storing the complete event sequence [25]. Applications of temporal graph methods to fraud detection in adjacent domains have yielded promising results, particularly in financial transaction networks and social bot detection, where the temporal trajectory of account behavior provides a robust detection signal that persists even when individual behavioral features are actively manipulated to evade detection [26].

3. Methodology

3.1 Temporal Review Graph Construction

The foundation of the proposed framework is a continuous-time dynamic review graph that represents the review ecosystem as an evolving heterogeneous network. Formally, the review graph is defined as a temporal attributed graph $G = (V, E, T)$, where $V = V_r \cup V_p$ denotes the node set comprising reviewer nodes V_r and product nodes V_p, E represents the set of directed temporal edges, and T is the set of associated timestamps. Each edge $e_{i,j,t} \in E$ encodes a review event in which reviewer i evaluated product j at time t, and is associated with a feature vector capturing the assigned rating, review length, sentiment polarity score, and purchase verification

status. A critical step in constructing this graph is the selection of node-level behavioral features that carry discriminative power for distinguishing fake reviewers from genuine ones. Figure 1 illustrates the cumulative distribution functions (CDFs) of six core behavioral features — word rate deviation (WRD), burstiness score (BST), review entropy score (RES), document length in bytes (DL_b), extreme rating density (ERD), and elapsed time gap (ETG) — computed separately for spam and non-spam reviewer populations.

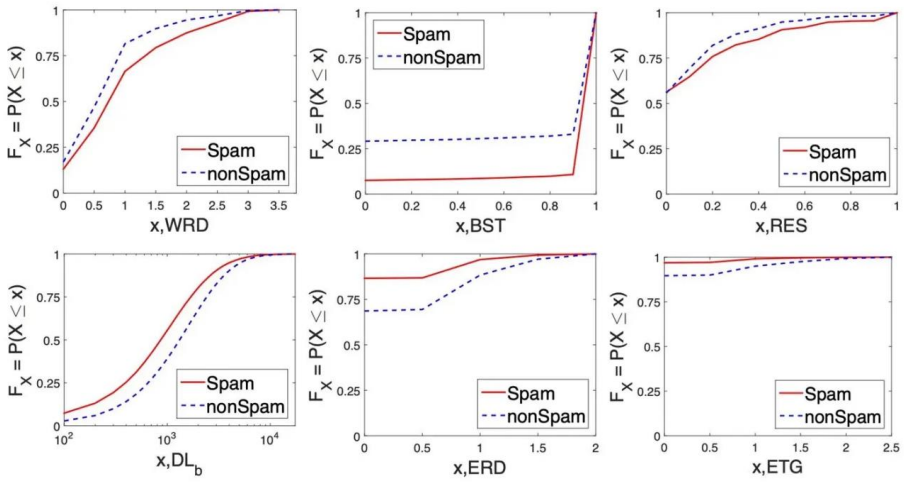


Figure 1: *Cumulative distribution functions of six behavioral features (WRD, BST, RES, DL_b, ERD, ETG) comparing spam (solid red) and non-spam (dashed blue) reviewer populations*

As shown in Figure 1, the BST feature demonstrates the most pronounced distributional divergence, with the spam population exhibiting a near-step distribution concentrated at values approaching 1.0, indicating that fake reviewers overwhelmingly tend to post in burst clusters rather than at temporally distributed intervals. The WRD and RES distributions reveal that spam reviewers tend toward lower lexical variety and more predictable content structures, while the DL_b distribution confirms that genuine reviewers produce substantially longer and more variable texts. The ERD and ETG distributions provide complementary evidence that fake reviewers cluster their activity temporally and target products at extreme rating positions with systematically higher frequency than genuine users. Reviewer node features are initialized from these six behavioral dimensions supplemented by account age, cumulative review count, and product category embedding. Product node features include average rating trajectory, total review count, and category embeddings derived from a pre-trained sentence encoder. The temporal graph is constructed by processing the review corpus in chronological order and inserting each review event as a new edge with its associated timestamp and feature vector. The choice of continuous-time graph representation over discrete temporal snapshot approaches is motivated by two principal considerations. First, review activity is inherently irregular, with individual reviewers posting at variable intervals that do not align naturally with fixed time windows. Forcing this activity into discrete snapshots inevitably introduces temporal aliasing artifacts that obscure the fine-grained behavioral patterns most relevant to fraud detection. Second, the behavioral drift patterns characteristic of fake reviewers often manifest as gradual continuous-time processes rather than abrupt discrete-time transitions, making them easier to capture with a continuous-time memory mechanism than with window-based approaches.

3.2 Temporal Graph Network with Behavioral Drift Encoder

The core learning component of the proposed framework is a memory-augmented TGN that processes the continuous-time dynamic review graph to produce temporally evolving node embeddings. The TGN maintains a memory module S ∈ R^{|V| × d_m} that stores a compressed

representation of each node's historical interaction context, where d_m denotes the memory dimension. The memory state of each node is initialized to a zero vector and updated each time the node participates in a new interaction event through a gated recurrent unit (GRU) that ingests the concatenation of source and destination node memories, the elapsed time since last interaction, and the edge feature vector. Central to the temporal attention aggregation module is the computation of interaction intensity between reviewer and product embeddings across time. Figure 2 illustrates the bilinear interaction formula that governs this computation, in which the intensity λ at relational depth r=3 is expressed as the exponential of the inner product between the transposed source entity embedding v^{e^o}(t'), the relation matrix R_3, and the destination entity embedding v^{e^o}(t''), scaled by the elapsed time interval (t − t').

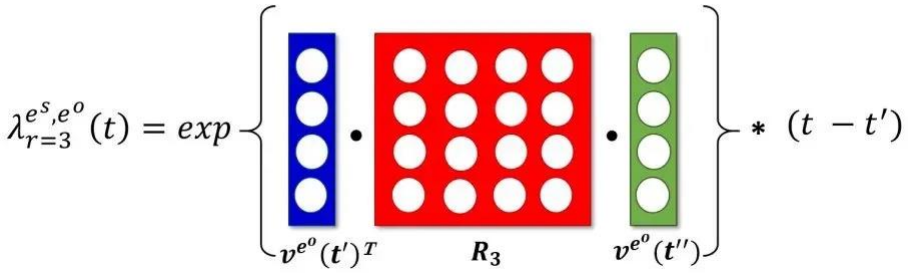


Figure 2: *Visualization of the temporal interaction intensity computation, expressing λ^{e^s, e^o}_{r=3}(t) as the exponential of a bilinear product between source embedding vector v^{e^o}(t')^T (blue), relation matrix R_3 (red), and destination embedding vector v^{e^o}(t'') (green), multiplied by the elapsed time (t − t')*

As Figure 2 demonstrates, the bilinear structure of the intensity function captures the asymmetric relational dynamics between reviewer and product nodes, allowing the model to distinguish reviewers who habitually engage with specific product categories from those whose targeting patterns shift abruptly — a characteristic behavioral signature of coordinated fraud campaigns. The relation matrix R_3 is learned end-to-end during training and encodes the relational semantics of the reviewer-product interaction type, while the time-scaling factor ensures that the intensity signal decays appropriately for interactions separated by large temporal intervals.

Behavioral drift is operationalized as a temporal divergence signal computed from the trajectory of each reviewer's embedding across a sequence of time-indexed snapshots. For each reviewer node r, the embedding trajectory E_r = {z_r(t_1), z_r(t_2), ..., z_r(t_k)} is extracted at evenly spaced temporal checkpoints spanning the reviewer's activity period. The drift score D_r is computed as the mean cosine dissimilarity between consecutive embedding snapshots, with higher values indicating greater behavioral instability. This scalar drift score is concatenated with the reviewer's final temporal embedding to form the input to a two-layer multilayer perceptron (MLP) classification head that outputs a binary fraud probability. The model is trained end-to-end using binary cross-entropy loss with class-balanced sampling to address the severe label imbalance typical of fake review datasets.

4. Results and Discussion

4.1 Experimental Setup and Baseline Comparison

Experiments were conducted on two widely used benchmark datasets: the Yelp-Zip dataset, containing approximately 608,598 reviews from 260,277 users across 5,044 restaurants with roughly 13.2% labeled fake, and the Amazon-Musical dataset, comprising 64,706 reviews from 44,702 users with approximately 9.7% fake review prevalence. Both datasets were preprocessed to extract temporal edge features from review timestamps, and a temporally stratified train-test split was applied, with the most recent 20% of reviews reserved for evaluation.

Understanding the structural diversity of adversarial behaviors that detection systems must contend with is essential for interpreting experimental outcomes. Figure 3 illustrates three canonical attack configurations that characterize the challenge of graph-based fake review detection in realistic deployment conditions.

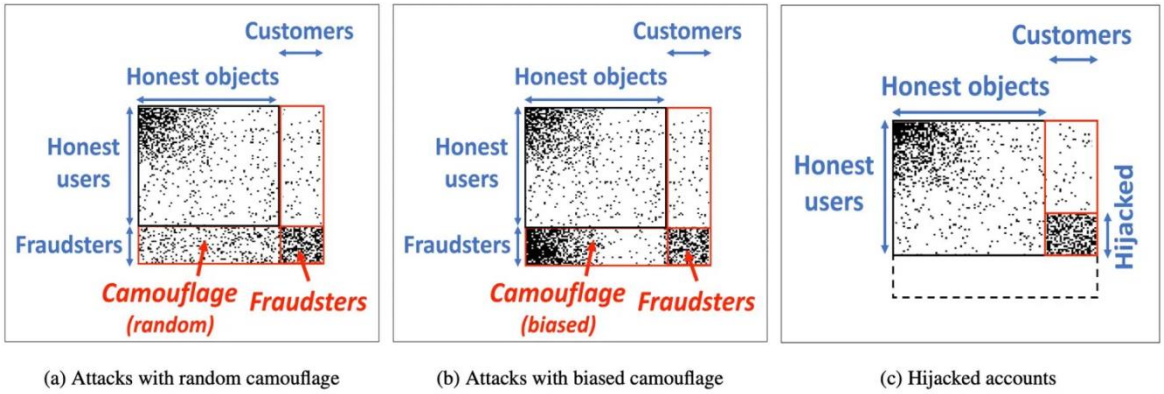


Figure 3: *Three camouflage-based attack scenarios in bipartite reviewer-product graphs: (a) attacks with random camouflage; (b) attacks with biased camouflage; and (c) hijacked accounts*

As shown in Figure 3, the hijacked account scenario in panel (c) poses the most significant challenge for any detection system relying on behavioral profiling, since the historical behavioral record of these accounts is genuinely honest and the transition to fraudulent activity is precisely what the drift signal must identify. The biased camouflage scenario in panel (b) is particularly challenging for static graph methods that rely on density-based fraud signals, since the deliberate mixing of fraudulent edges with connections to popular legitimate products reduces the apparent density of the fraud subgraph below detection thresholds. These attack configurations directly motivate the temporal drift-aware architecture of TGN-Drift, whose memory module tracks the full behavioral trajectory of each reviewer and explicitly encodes the transition between behavioral regimes as a fraud-predictive signal. Baseline methods included a random forest classifier trained on hand-crafted behavioral and linguistic features (RF-BL), a gradient boosting machine trained on the same feature set (GBM-BL), a static graph attention network applied to a snapshot of the complete review graph (GAT-Static), a temporal graph attention network without the behavioral drift module (TGA-Base), and a recurrent graph neural network approach processing discrete monthly snapshots with long short-term memory (LSTM) aggregation between snapshots (RNN-Graph). The proposed method is designated TGN-Drift throughout the experimental results. Performance was evaluated using precision, recall, and macro-averaged F1-score, supplemented by area under the receiver operating characteristic curve (AUC-ROC). Statistical significance of performance differences was assessed using a paired bootstrap test with 1,000 resampling iterations at the 95% confidence level.

4.2 Analysis of Behavioral Drift as a Detection Signal

The results reveal consistent and substantial improvements from TGN-Drift over all baseline methods across both datasets. On the Yelp-Zip dataset, TGN-Drift achieved an F1-score of 0.847 and an AUC-ROC of 0.923, compared to 0.791 and 0.874 for GAT-Static and 0.812 and 0.901 for TGA-Base. On the Amazon-Musical dataset, TGN-Drift achieved an F1-score of 0.831 and an AUC-ROC of 0.908, outperforming GAT-Static's 0.764 and 0.851 and TGA-Base's 0.798 and 0.883. The performance gap between TGN-Drift and TGA-Base is statistically significant on both datasets, demonstrating that the explicit behavioral drift module contributes meaningfully beyond the base temporal graph embedding.

A more detailed analysis of the drift signal reveals that its discriminative power is concentrated among fake reviewers who exhibit non-trivial behavioral evolution over time — precisely the hijacked account profile illustrated in Figure 3(c). These reviewers exhibit long activity histories with initially benign behavioral patterns before transitioning to coordinated fraud, rendering them invisible to static and short-window detection methods whose fraud classifiers are calibrated against contemporaneous behavioral snapshots rather than trajectories of change. The drift module enables TGN-Drift to detect the transition itself as a suspicious signal. Qualitative inspection of the learned temporal embeddings confirms that fake reviewers exhibit significantly higher mean drift scores than genuine reviewers, with the distribution of drift scores for fake reviewers showing a pronounced right tail corresponding to accounts that underwent particularly abrupt behavioral transitions. The temporal checkpoint at which the drift score begins to rise is often associated with identifiable platform events such as the introduction of a new spam campaign or a policy enforcement action that altered the operational risk landscape for organized fraud networks. The model's performance under varying degrees of label noise was also evaluated by artificially corrupting training labels and measuring resulting test performance degradation. TGN-Drift demonstrated greater robustness to label noise than the static GAT baseline across all corruption rates tested, consistent with the expectation that temporal trajectory signals are harder to confound through label corruption than point-in-time structural features. Computational overhead analysis on the Yelp-Zip dataset showed that training TGN-Drift required approximately 3.7 times the wall-clock time of GAT-Static, while inference latency remained within acceptable bounds for near-real-time deployment.

5. Conclusion

This paper has presented a TGN-based framework for adaptive fake review detection that explicitly models reviewer behavioral drift as a temporally evolving graph learning problem. The proposed TGN-Drift system constructs a continuous-time dynamic review graph, maintains memory-augmented temporal embeddings for each reviewer and product node, and operationalizes behavioral drift as a temporal embedding divergence signal that supplements the structural and semantic content of the learned representations. The behavioral feature distributions illustrated in Figure 1 motivated the design of the reviewer node attribute space, the bilinear interaction intensity formulation in Figure 2 informed the temporal attention aggregation mechanism, and the adversarial attack taxonomy in Figure 3 contextualized the experimental evaluation protocol and highlighted the particular vulnerability of static methods to behavioral camouflage and account hijacking strategies. Empirical evaluation on two benchmark datasets demonstrated consistent and statistically significant performance improvements over static graph, semi-dynamic, and ablated temporal baselines, with the behavioral drift module contributing most substantially for fake reviewers with extended activity histories who exhibit strategic behavioral evolution over time. The results underscore a fundamental limitation of static and snapshot-based detection paradigms: by treating the review graph as a fixed artifact rather than an evolving process, these methods are unable to exploit the temporal trajectory of reviewer behavior as a fraud signal and are structurally vulnerable to degradation as fraudulent behavioral patterns drift away from training-time distributions. Several directions for future work emerge from this study. The integration of NLP signals derived from review text into the temporal graph framework represents a promising avenue for further performance improvement, as the temporal evolution of a reviewer's linguistic patterns may carry additional discriminative information complementary to the structural and behavioral signals exploited here. The extension to multi-platform settings, where fake reviewer accounts operate across multiple platforms with correlated behavioral trajectories that span platform boundaries, poses both a technical challenge and an opportunity for substantially improved detection coverage. Finally, the development of interpretable temporal attribution mechanisms that

explain detection decisions in terms of specific behavioral events in a reviewer's history would enhance the practical utility of the framework for platform trust and safety teams requiring actionable enforcement justifications.

References

Hussain, N., Turab Mirza, H., Rasool, G., Hussain, I., & Kaleem, M. (2019). Spam review detection techniques: A systematic literature review. Applied Sciences, 9(5), 987.

Abdulqader, M., Namoun, A., & Alsaawy, Y. (2022). Fake online reviews: A unified detection model using deception theories. IEEE Access, 10, 128622-128655.

Ji, S. J., Zhang, Q., Li, J., Chiu, D. K., Xu, S., Yi, L., & Gong, M. (2020). A burst-based unsupervised method for detecting review spammer groups. Information Sciences, 536, 454-469.

Liu, Z., Dou, Y., Yu, P. S., Deng, Y., & Peng, H. (2020, July). Alleviating the inconsistency problem of applying graph neural network to fraud detection. In Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval (pp. 1569-1572).

Dou, Y., Liu, Z., Sun, L., Deng, Y., Peng, H., & Yu, P. S. (2020, October). Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. In Proceedings of the 29th ACM international conference on information & knowledge management (pp. 315-324).

Zhao, J., Shao, M., Tang, H., Liu, J., Du, L., & Wang, H. (2023). RHGNN: Fake reviewer detection based on reinforced heterogeneous graph neural networks. Knowledge-Based Systems, 280, 111029.

Garcia, C. M., Abilio, R., Koerich, A. L., Britto Jr, A. D. S., & Barddal, J. P. (2025). Concept drift adaptation in text stream mining settings: A systematic review. ACM Transactions on Intelligent Systems and Technology, 16(2), 1-67.

Suryawanshi, S., Goswami, A., & Patil, P. (2024). FakeIDCA: Fake news detection with incremental deep learning based concept drift adaption. Multimedia Tools and Applications, 83(10), 28579-28594.

Cai, L., Chen, Z., Luo, C., Gui, J., Ni, J., Li, D., & Chen, H. (2021, October). Structural temporal graph neural networks for anomaly detection in dynamic graphs. In Proceedings of the 30th ACM international conference on Information & Knowledge Management (pp. 3747-3756).

Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., & Bronstein, M. (2020). Temporal graph networks for deep learning on dynamic graphs. arXiv preprint arXiv:2006.10637.

Feng, S., Tan, Z., Wan, H., Wang, N., Chen, Z., Zhang, B., ... & Luo, M. (2022). Twibot-22: Towards graph-based twitter bot detection. Advances in Neural Information Processing Systems, 35, 35254-35269.

Chen, J., Liu, J., Liang, Y., & Zhou, M. (2026). KE-MLLM: A Knowledge-Enhanced Multi-Sensor Learning Framework for Explainable Fake Review Detection. Applied Sciences, 16(6), 2909.

Li, A., Qin, Z., Liu, R., Yang, Y., & Li, D. (2019, November). Spam review detection with graph convolutional networks. In Proceedings of the 28th ACM international conference on information and knowledge management (pp. 2703-2711).

Kumar, N., Venugopal, D., Qiu, L., & Kumar, S. (2019). Detecting anomalous online reviewers: An unsupervised approach using mixture models. Journal of Management Information Systems, 36(4), 1313-1346.

Ren, Z., Zhang, X., Zhang, S., & Yang, C. (2023, September). Fake Review Detection via Heterogeneous Graph Attention Network. In International Conference on Artificial Neural Networks (pp. 364-376). Cham: Springer Nature Switzerland.

Guo, Y., Ling, Y., & Chen, H. (2020). A time-aware graph neural network for session-based recommendation. IEEE Access, 8, 167371-167382.

Xia, Z., Liu, C., Gong, N. Z., Li, Q., Cui, Y., & Song, D. (2019, July). Characterizing and detecting malicious accounts in privacy-centric mobile social networks: A case study. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2012-2022).

Cheng, D., Wang, X., Zhang, Y., & Zhang, L. (2020). Graph neural network for fraud detection via spatial-temporal attention. IEEE Transactions on Knowledge and Data Engineering, 34(8), 3800-3813.

Wu, K., Lin, Y., Wang, X., & Wang, L. (2026). Multimodal Learning-Based Relational Graph Neural Networks for Social Bot Detection. IEEE Transactions on Network and Service Management.

Kim, Y., Lee, Y., Choe, M., Oh, S., & Lee, Y. (2024). Temporal graph networks for graph anomaly detection in financial networks. arXiv preprint arXiv:2404.00060.

Cong, W., Zhang, S., Kang, J., Yuan, B., Wu, H., Zhou, X., ... & Mahdavi, M. (2023). Do we really need complicated model architectures for temporal networks?. arXiv preprint arXiv:2302.11636.

Baek, J., Lee, D. B., & Hwang, S. J. (2020). Learning to extrapolate knowledge: Transductive few-shot out-of-graph link prediction. Advances in neural information processing systems, 33, 546-560.

Ding, J., Shen, Z., & Liu, W. (2026). Game-Theoretic Cost-Sensitive Adversarial Training for Robust Cloud Intrusion Detection Against GAN-Based Evasion Attacks. Applied Sciences, 16(8), 3944.

Ping, W., Jiao, Y., Fan, H., & Zhang, X. (2026). Multimodal Fraud Detection in Financial Statements: A Trimodal Attention Network with Contrastive Evidence Chain Construction. IEEE Access.

Wang, B., Wang, Z., Zhao, W., Zhang, F., & Shang, W. (2026). DRL-Adapt: Deep Reinforcement Learning for Adaptive Routing Convergence Optimization in Large-Scale Networks. IEEE Open Journal of the Computer Society.

Teng, D., Rhee, M., Qin, Y., Zi, B., & Liu, W. (2026). SW-SpeedDLM: Sliding-Window Speculative Decoding for Diffusion Language Models under Long-Context Constraints. Mathematics