

Cracking the Snowflake: Recovering Propagation Rhythms for Fake News Detection in the Post-API Era

Chiao-Yu Li, Divya Chaudhary

Khoury College of Computer Sciences, Northeastern University
Seattle, WA, USA

{li.chiao, d.chaudhary}@northeastern.edu

Abstract

The proliferation of misinformation on social media has driven extensive research into automated fake news detection. However, the “Post-API” era presents a critical challenge: as major platforms implement restrictive data-access policies, traditional approaches relying on explicit social graphs and user metadata become increasingly impractical. This paper proposes RGCP-Snowflake, a structural-semantic fusion framework that leverages Snowflake IDs as structural surrogates to reconstruct temporal propagation patterns without requiring access to restricted social metadata. By extracting millisecond-level timestamps encoded within distributed identifiers, the proposed approach recovers the propagation rhythm, a latent veracity signal characterising the dissemination patterns typical of fake news. A comprehensive benchmarking across three model tiers (commercial large language models, traditional machine learning classifiers, and open-source language models) is conducted on the FakeNewsNet dataset. The experiments reveal two key findings: a scale threshold phenomenon, where smaller language models exhibit systematic label collapse, defaulting to classifying all inputs as misinformation; and the competitive performance of traditional classifiers against larger open-source models, indicating inherent limitations in purely semantic zero-shot reasoning. RGCP-Snowflake outperforms all baseline models by integrating non-manipulable temporal signatures with semantic headline features, achieving a mean accuracy of $89.53\% \pm 1.58\%$ across 10 random seeds on the FakeNewsNet PolitiFact benchmark—a statistically significant margin over the strongest LLM baseline ($p < 0.01$)—demonstrating that structural surrogates provide a robust detection mechanism effective in restricted environments where explicit social metadata is unavailable.

Introduction

The digital landscape has entered a restrictive “Post-API” era, where the rapid proliferation of misinformation is met with increasingly paywalled data environments. As major social media platforms implement aggressive data-access policies, researchers face a significant technical hurdle: the high-fidelity discriminative features previously used for fake news detection, such as complete conversational threads, user history, and precise engagement metadata, are no

longer accessible to the academic community. This systemic scarcity necessitates a shift toward data-efficient models capable of operating in zero-context environments, moving away from a total reliance on massive, platform-dependent datasets.

It is important to distinguish between Post-API data collection and Post-API real-time analysis. While the initial acquisition of tweet IDs still necessitates upstream API access—or reliance on established pre-API era benchmarks such as FakeNewsNet—the primary innovation of the proposed framework lies in the Post-API analysis phase. By exclusively utilising the metadata embedded within Snowflake IDs, the model eliminates the need for subsequent, resource-intensive API calls to hydrate tweet objects or retrieve full-text content during the detection process.

This research is motivated by the inherent limitations of purely semantic analysis, a gap highlighted by the shift from early supervised learning models like BERT to general-purpose large language models. While landmark studies have shown that pre-trained transformers excel at text classification, they often struggle with the adversarial nature of misinformation, which is specifically designed to mimic the linguistic style of legitimate journalism. This paper addresses this gap by introducing structural surrogates, a novel approach that leverages the physical architecture of Snowflake IDs to reconstruct propagation blueprints. The latent signals embedded in the rhythm and topology of information spread carry veracity markers that semantic-only models fail to capture, particularly as model scale decreases.

The scale of a language model is not simply a proxy for its detection capability. The experiments presented here show that models below a certain parameter threshold exhibit systematic label collapse, defaulting to conservative predictions that flag most content as fake regardless of actual veracity. This behaviour is particularly problematic in deployment settings where the cost of false positives is high, and it underscores the need for complementary signals that do not rely on parametric knowledge alone. By integrating temporal surrogate features derived from Snowflake identifiers, it is shown that it is possible to partially compensate for the reasoning limitations of smaller models and provide a more stable detection signal across the full range of model scales evaluated.

In this work, a comprehensive benchmarking of thirteen

distinct models is presented, bridging the gap between traditional statistical classifiers and the latest frontier of generative AI. By evaluating traditional machine learning algorithms alongside commercial and open-source large language models on the FakeNewsNet dataset, an empirical mapping of the scale threshold required for effective zero-shot reasoning is provided. The contribution lies in demonstrating how structural reconstruction can augment semantic detection, offering a robust framework for identifying misinformation in an era of limited data transparency.

Related Work

Fake news detection has evolved from a primarily text-based classification problem into a multi-faceted research area that increasingly incorporates social context, propagation behaviour, temporal dynamics, and large-scale language modelling. Early work treated the task as determining whether a claim or article was true or false based on linguistic and stylistic content alone, but later studies demonstrated that deceptive and truthful content may look similar at the lexical level while still differing substantially in how they spread online (Vosoughi, Roy, and Aral 2018). As a result, the field has gradually shifted from purely content-driven modelling toward approaches that also leverage propagation structure, user context, and, more recently, the reasoning capabilities of large language models.

Benchmark datasets shaped much of this evolution. Wang (Wang 2017) introduced LIAR to formalise veracity classification over short political statements, while FakeNewsNet (Shu et al. 2020) expanded the task by combining news content with social context and spatiotemporal information. Nakamura et al. (Nakamura, Levy, and Wang 2020) further broadened the landscape with Fakeddit, a large-scale multimodal dataset for fine-grained veracity classification. Early systems relied on classical ma-

n-gram counts, and stylometric patterns, paired with classifiers including SVMs and logistic regression (Shu et al. 2017). Deep learning reduced dependence on manual engineering: CNN-based models captured local lexical patterns, while LSTMs modelled sequential dependencies (Ma et al. 2016), and hybrid architectures combining both generally outperformed traditional baselines (Mridha et al. 2021; Ruchansky, Seo, and Liu 2017). Transformer-based models further raised the performance ceiling, with BERT establishing deeply bidirectional contextual representations as a standard backbone (Devlin et al. 2019). Specialised variants such as FakeBERT (Kaliyar, Goswami, and Narang 2021) and exBAKE (Jwa et al. 2019) demonstrated that BERT-based representations, optionally augmented with external knowledge, can be effectively adapted for fake news classification.

To capture evidence beyond semantic content, later work turned to graph-based modelling, representing users, posts, and replies as nodes and encoding interaction patterns as edges. BiGCN (Bian et al. 2020) modelled both forward propagation and backward user responses, while FANG (Nguyen et al. 2020) and UPFD (Dou et al. 2021) incorporated heterogeneous social context and user preference into graph representations. DPSG (Jing et al. 2025) further captured the dynamic nature of news cascades by jointly modelling user–post interactions over time. These approaches establish that structural information often provides evidence that text-only models cannot recover, though they rely on complete propagation graphs that are frequently unavailable due to API restrictions or data sparsity.

The emergence of LLMs introduced a reasoning-driven paradigm in which models can be queried zero-shot or via chain-of-thought prompting without task-specific fine-tuning. Comparative studies show that while zero-shot LLMs can lag behind specialised encoder models, fine-tuning substantially closes this gap (Papageorgiou, Varlamis, and Chronis 2025; Raza, Paulen-Patterson, and Ding 2025). Tong et al. (Tong et al. 2025) improved upon standard prompting with a generate-then-sample framework combining LLM-generated explanations with reinforcement learning, while Das and Dodge (Das and Dodge 2025) highlighted that LLMs can also be exploited to paraphrase fake news in ways that evade existing detectors, a phenomenon they term LLM laundering. More broadly, the dual-use nature of generative AI, as both a tool for producing and detecting synthetic misinformation, creates an ongoing arms race that shapes current research priorities (Loth, Kappes, and Pahl 2024).

Multimodal approaches address the fact that misinformation frequently combines text with manipulated images. SpotFake (Singhal et al. 2019) and MVAE (Khattar et al. 2019) demonstrated that jointly modelling textual and visual features improves over unimodal baselines, while EANN (Wang et al. 2018) disentangles event-specific features for better generalisation. More recent models including BMLHF (Wu et al. 2025), KAMP (Zhang et al. 2025), and MHR (Feng et al. 2025) further address modality imbalance, factual grounding, and hierarchical social structure respectively. A persistent challenge across all paradigms is gen-

Model	Approach	Dataset	Acc. (%)
SVM (2017)	Traditional ML	FakeNewsNet	76.20
CSI (2017)	Traditional ML	Twitter15/16	89.73
CNN (2021)	Deep Learning	LIAR	85.30
BiLSTM (2016)	Deep Learning	Twitter15/16	90.40
BERT (2019)	Transformer	LIAR	86.30
FakeBERT (2021)	Transformer	Various	98.90
exBAKE (2019)	Transformer	LIAR	87.00
BiGCN (2020)	Graph Neural Net	Twitter15/16	88.10
FANG (2020)	Graph Neural Net	FakeNewsNet	76.40
UPFD (2021)	Graph Neural Net	FakeNewsNet	84.50
DPSG (2025)	Graph Neural Net	Weibo21	90.12
GPT-4 (2025)	LLM (zero-shot)	Various	79.40
GSFND (2025)	LLM + RL	FakeNewsNet	91.30
SpotFake (2019)	Multimodal	Twitter/Weibo	77.23
BMLHF (2025)	Multimodal	Fakeddit	91.47
MHR (2025)	Multimodal	FakeNewsNet	90.15

Table 1: Representative prior work across fake news detection paradigms. Accuracy values are as reported in the original papers on the cited dataset.

chine learning over handcrafted features such as TF-IDF,

eralisation to emerging events; EvolveDetector (Ding et al. 2025) addresses this through continual knowledge accumulation to avoid catastrophic forgetting as new events arise. Surveys of the field consistently identify robustness to distribution shift, cross-platform generalisation, and interpretability as persistent open problems (Hu, Mao, and Zhang 2025).

Table 1 summarises representative methods across all paradigms. This work contributes a systematic three-way comparison on FakeNewsNet PolitiFact across classical ML baselines, LLM-based zero-shot prompting, and RGCP-Snowflake, a graph neural network that reconstructs temporal and sequence-order features from Twitter Snowflake identifiers without requiring complete propagation graphs.

Methodology

RGCP-Snowflake is proposed as a fake news detection framework designed for restricted-information environments. Many effective misinformation detectors rely on propagation graphs or user interaction structures, but these signals are often unavailable in practical deployment settings. Rather than assuming access to reply trees, repost networks, or full social graphs, the proposed method extracts lightweight temporal and ordering signals from Snowflake-style tweet identifiers and uses them as structural surrogates that remain accessible even when richer interaction data is absent. An overview of the full pipeline is shown in Figure 1.

Evaluation is performed on the FakeNewsNet PolitiFact subset (Shu et al. 2020), which provides ground-truth veracity labels alongside news content and associated tweet IDs. After filtering entries with missing or excessively short titles, the usable subset contains 1,052 samples (624 real, 428 fake). Because the full article body is not always available in the raw dataset, news headlines are used as the sole textual input signal. This constraint makes the results more conservative relative to body-text approaches and reflects a realistic scenario in which only minimal content metadata is at hand. Samples are divided into an 80/20 train–test split using stratified sampling to preserve the real/fake class ratio across both partitions (train: 496 real, 345 fake; test: 125 real, 86 fake).

The central hypothesis of this framework is that even in the absence of an explicit social graph, temporal metadata embedded in tweet identifiers can serve as a meaningful proxy for propagation structure. Twitter Snowflake IDs are 64-bit integers whose high-order bits encode a millisecond-precision creation timestamp relative to a fixed platform epoch. Recovering this timestamp requires only a single bit-shift operation, converting an opaque integer into an absolute creation time and unlocking a temporal dimension that is otherwise invisible in the raw identifier. This recovery process does not require any API access or user profile data, making it viable in the restricted-information settings where conventional graph-based methods typically fail.

The significance of this recovery is that it reveals the rhythm of how a news item accumulated engagement over time. The hypothesis motivating this work is that misinformation tends to arrive in irregularly bursty bursts resembling coordinated cascades, whereas organic news attracts more gradual and sustained engagement.

To test this hypothesis empirically, inter-arrival times were computed from the recovered Snowflake timestamps for every news item in the dataset and the resulting distributions compared across veracity classes. Figure 2 presents the results. Fake news shows a stronger spike at ultra-short intervals ($\log \text{IAT} \approx 0$), a higher coefficient of variation ($\text{CV} = 18.8$ vs. 15.0), and a shorter median IAT (263.8s vs. 390.5s for real news), all consistent with coordinated burst activity. Real news, by contrast, exhibits a heavier right tail, reflecting more sustained and distributed engagement. Notably, while these differences are statistically highly significant (Mann-Whitney U, $p = 6.30 \times 10^{-304}$), the two distributions overlap substantially, indicating that the temporal signal is subtle rather than immediately obvious. The fact that RGCP-Snowflake achieves 89.53% accuracy by leveraging these micro-temporal patterns demonstrates that graph-based relational learning can extract discriminative structure from signals that are not recoverable through simple statistical inspection alone. By decoding every tweet ID and sorting the resulting timestamps chronologically, a fine-grained sequence of engagement events is obtained that captures this rhythm without accessing any social network structure. The claim is therefore not that Snowflake metadata replaces social graphs entirely, but that it preserves a useful subset of propagation-aware information under conditions where richer signals are unavailable.

Each news item is represented as a heterogeneous directed graph containing one root node for the news article and up to 50 tweet nodes representing the engagement events, ordered chronologically by recovered timestamp. This cap limits memory consumption while retaining the most informative portion of the propagation sequence for high-volume items. Items whose tweet ID field is missing or empty are represented by a single root node with no edges, allowing the model to handle the no-propagation case without any special-case logic.

Node features are 768-dimensional throughout. The root node is assigned a sentence embedding of the headline produced by all-distilroberta-v1, capturing the semantic content of the article. Tweet nodes do not have associated text, so their features are constructed by blending the root embedding with a sinusoidal positional encoding derived from each node’s normalised position within the observed propagation window. This weighting preserves the semantic signal of the headline as the dominant component while injecting a continuous, ordered representation of when each engagement event occurred relative to the full cascade.

Three typed edge relations encode structurally distinct aspects of the propagation process. Type-0 edges connect the root node to every tweet node, modelling the initial broadcast of the article. Type-1 edges form a directed temporal chain between consecutive tweet nodes in chronological order, encoding the sequential dynamics of how engagement unfolds. Type-2 edges are reverse aggregation edges from every tweet node back to the root, closing the information loop and allowing accumulated temporal evidence to flow back into the article representation before classification. Together, these three relation types provide access to broadcast, sequential, and aggregation signals without requiring

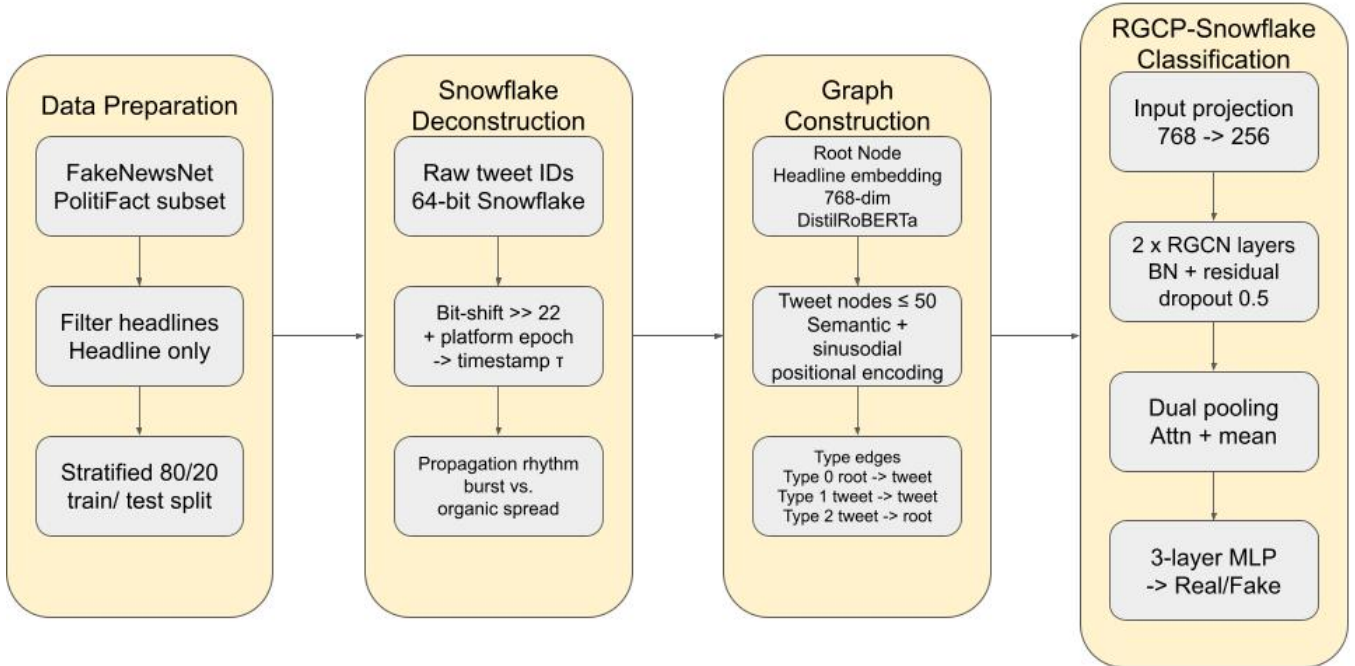


Figure 1: Overview of the RGCP-Snowflake pipeline across four stages: data preparation, Snowflake ID deconstruction, heterogeneous graph construction, and relational graph convolutional classification.

any explicit social network structure.

RGCP-Snowflake is a relational graph convolutional network with approximately 1.28 million parameters. An input projection MLP first compresses the 768-dimensional node features to a hidden dimension of 256. Two RGCN layers then propagate information across the three edge types, maintaining separate weight matrices per relation type so that broadcast, temporal, and aggregation messages are transformed distinctly. Each layer is followed by batch normalisation, a ReLU activation, dropout ($p = 0.5$), and a residual connection to mitigate over-smoothing. After message passing, node representations are aggregated into a fixed-size graph embedding via dual pooling: a learned global attention gate produces a weighted summary, concatenated with a mean pool to preserve broad context. A three-layer MLP classifier then maps the pooled embedding to a binary real/fake prediction.

The model is trained with AdamW (learning rate 5×10^{-4} , weight decay 10^{-3}) and a cosine annealing scheduler with warm restarts. Cross-entropy loss is weighted by the inverse class frequency of the training split to correct for class imbalance. Gradients are clipped to an ℓ_2 norm of 1.0, the batch size is 32, and training halts via early stopping if validation accuracy does not improve for 20 consecutive epochs, at which point the best checkpoint is restored. The full hyperparameter configuration is summarised in Table 2. On the PolitiFact subset, the model converged within approximately 27 epochs.

Table 2: Hyperparameter configuration of RGCP-Snowflake.

Hyperparameter	Value
Hidden dimension	256
Number of RGCN layers	2
Dropout rate	0.5
Max tweet nodes	50
Batch size	32
Optimiser	AdamW
Learning rate	5×10^{-4}
Weight decay	1×10^{-3}
Gradient clip (ℓ_2)	1.0
Early stopping patience	20 epochs
Total parameters	$\approx 1.28\text{M}$

To contextualise the contribution of Snowflake-derived structural information, two additional paradigm families are evaluated on the same benchmark. For classical ML baselines, TF-IDF features (vocabulary size 5,000; n -gram range (1, 2)) are extracted from headlines to train a linear SVM, Naive Bayes, Random Forest, and MLP. For LLM baselines, both closed-source models (GPT-4-Turbo, Claude-3.5-Sonnet, Gemini-2.0-Flash, Gemini-3-Flash) and open-source models (Mistral-7B-v0.3, Qwen2.5-7B-Instruct, Yi-1.5-6B-Chat, Qwen2.5-3B-Instruct) are evaluated in a zero-shot setting where each

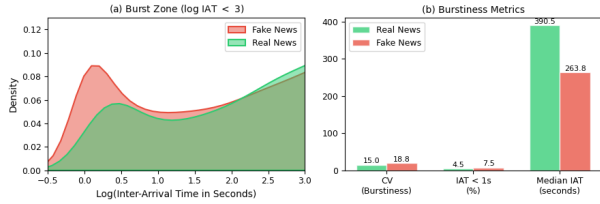


Figure 2: Empirical analysis of inter-arrival time (IAT) distributions on FakeNewsNet PolitiFact. (a) In the burst zone ($\log \text{IAT} < 3$), fake news exhibits a more pronounced short-interval spike, reflecting a higher proportion of near-instantaneous reactions ($\text{IAT} < 1\text{s}$: 7.5% vs. 4.5%). (b) Burstiness metrics confirm that fake news has a higher coefficient of variation ($\text{CV} = 18.8$ vs. 15.0) and a shorter median IAT (263.8s vs. 390.5s), consistent with compressed cascade behaviour. While the distributions overlap substantially, a Mann-Whitney U test confirms they differ significantly ($p = 6.30 \times 10^{-304}$), indicating that micro-temporal patterns carry a subtle but statistically reliable veracity signal that graph-based relational learning can exploit.

model receives only the headline and a binary classification prompt with no training examples or propagation features.

Results and Discussion

A comparative analysis of models evaluated on the FakeNewsNet PolitiFact benchmark ($n = 1,052$) is presented. The central question investigated is whether supervised statistical learning over surface-level features, zero-shot generative reasoning, or structure-aware graph learning provides the most reliable signal for fake news detection under constrained information access. Results are summarised in Table 3.

To ensure a rigorous head-to-head comparison with prior graph-based methods, BiGCN (Bian et al. 2020) and UPFD (SAGE variant) (Dou et al. 2021) were re-evaluated on the identical PolitiFact split used in all experiments. RGCP-Snowflake consistently outperforms both, achieving a mean accuracy of 89.53% compared to BiGCN (84.83%) and UPFD-SAGE (87.20%). Notably, both BiGCN and UPFD rely on joint textual-structural features derived from fully reconstructed social graphs, whereas RGCP-Snowflake achieves competitive performance while operating solely on the temporal rhythms recovered from Snowflake IDs. To validate the stability of these results, the model was evaluated across 10 independent random seeds, yielding a mean accuracy of $89.53\% \pm 1.58\%$ with individual runs ranging from 86.73% to 91.94%. A paired t -test against zero-shot Gemini-3-Flash ($86.97\% \pm 1.38\%$) yields $p = 0.0039$ ($p < 0.01$), confirming that the performance advantage is statistically significant and not attributable to favourable seed selection.

Within the commercial LLM tier, Gemini-3-Flash achieved the highest zero-shot accuracy at $86.97\% \pm 1.38\%$, defining the semantic performance ceiling for this class of

Table 3: Performance comparison on FakeNewsNet PolitiFact ($n = 1,052$). RGCP-Snowflake results are mean \pm std over 10 seeds. Graph baselines re-evaluated on the same split.

Model	Acc.	F1-Fake	F1-Real
SVM (Linear)	0.8578	0.8171	0.8837
MLP	0.8389	0.7901	0.8692
Naive Bayes	0.8341	0.7826	0.8659
Random Forest	0.7156	0.5082	0.8000
BiGCN (2020)	0.8483	0.8242	—
UPFD-SAGE (2021)	0.8720	0.8525	—
Mistral-7B-v0.3	0.7747	0.7228	0.8102
Qwen2.5-7B-Instruct	0.7614	0.7304	0.7860
Yi-1.5-6B-Chat	0.6958	0.6774	0.7122
Qwen2.5-3B-Instruct	0.5722	0.6318	0.4898
Gemini-3-Flash (0-shot)	0.8697 ± 0.014	—	—
Gemini-3-Flash (5-shot)	0.8346	—	—
Gemini-2.0-Flash	0.8565	0.8143	0.8830
Claude-3.5-Sonnet	0.8289	0.7941	0.8537
GPT-4-Turbo	0.8222	0.7983	0.8411
RGCP-Snowflake	0.8953 ± 0.016	0.8782	0.9135

model. Gemini-2.0-Flash followed at 85.65%, and Claude-3.5-Sonnet reached 82.89%. The competitive performance of these models is consistent with the expectation that large-scale pre-training exposes a model to a wide range of rhetorical patterns and factual inconsistencies associated with misinformation. That said, even the strongest commercial model falls several percentage points below RGCP-Snowflake, suggesting that semantic reasoning alone leaves meaningful headroom that structural signals can fill. To evaluate stronger LLM baselines beyond zero-shot, a 5-shot in-context learning setup was also assessed for Gemini-3-Flash. Counter-intuitively, performance declined to 83.46% in this setting, suggesting that explicit semantic exemplars may introduce contextual interference rather than useful grounding, further underscoring the resilience of the non-textual approach.

Among the traditional ML baselines, the linear SVM achieved the strongest result at 85.78%, with a well-balanced F1 across both classes (F1-Real: 0.8837, F1-Fake: 0.8171). Naive Bayes followed at 83.41%, while the MLP reached 83.89%. Random Forest dropped to 71.56%, with a substantially weaker F1 on the fake class (0.5082), indicating difficulty in reliably identifying deceptive content. The competitive performance of SVM and Naive Bayes relative to larger zero-shot models reflects the degree to which the PolitiFact subset contains consistent surface-level markers within the training distribution. However, this dependence on distributional overlap is also the source of their brittleness; such classifiers are unlikely to generalise to news cycles with different stylistic conventions or adversarially crafted content.

The open-source models revealed a clear relationship between parameter scale and detection reliability. Mistral-7B-v0.3 was the strongest in this group at 77.47%, followed by Qwen2.5-7B-Instruct at 76.14%. Both trail the SVM base-

line by a noticeable margin, indicating that zero-shot reasoning in 7B-class models does not yet match the statistical coverage that comes from supervised training on in-distribution examples. The situation becomes more acute in sub-7B models. Yi-1.5-6B-Chat achieved 69.58%, and Qwen2.5-3B-Instruct reached only 57.22%, barely above chance. Inspection of its confusion matrix revealed a systematic pattern: the model disproportionately predicted the fake class regardless of content, consistent with the label collapse phenomenon observed in under-parameterised models facing adversarial inputs. The finding that a 3B-parameter model collapses to a conservative heuristic, and that even a 7B-parameter model falls short of a well-regularised linear classifier, reinforces the view that scale alone is insufficient for reliable zero-shot misinformation detection without additional structural cues.

RGCP-Snowflake achieves a mean accuracy of $89.53\% \pm 1.58\%$ (F1-Real: 0.9135, F1-Fake: 0.8782) over 10 random seeds, outperforming all evaluated baselines including established graph neural network methods. This improvement is particularly meaningful given that both approaches have access to essentially the same textual content, with neither accessing user profiles, reply trees, or social network structure. The difference lies in the additional temporal layer introduced through Snowflake ID deconstruction. By recovering the rhythm of engagement from 64-bit identifiers, the model gains a non-semantic verification signal orthogonal to the stylistic and rhetorical cues exploited by language-based classifiers. Misinformation tends to accumulate engagement in bursty, compressed windows, while legitimate news attracts more gradual and sustained activity; this pattern is empirically confirmed by the Mann-Whitney U test ($p < 0.001$) and becomes visible once timestamps are recovered from the associated tweet IDs.

Taken together, the results indicate that no single paradigm dominates across all conditions. Commercial LLMs are strong zero-shot reasoners but plateau at a semantic ceiling. Traditional classifiers remain competitive when in-distribution lexical patterns are reliable but are vulnerable to distribution shift. Open-source models at 7B and below show increasing instability, with smaller models exhibiting outright classification collapse. RGCP-Snowflake achieves the highest accuracy by combining a semantically grounded node representation with temporal surrogate structure, supporting the broader claim that propagation-aware reasoning remains valuable for fake news detection even when access to explicit social graphs is unavailable.

Conclusion

This paper proposed RGCP-Snowflake, a fake news detection framework that reconstructs structural propagation signals from Snowflake tweet identifiers without requiring access to social network data or user profiles. The central argument is that the temporal rhythm encoded within distributed identifiers constitutes a meaningful veracity signal that is orthogonal to semantic content and remains accessible in restricted-information environments where conventional graph-based methods fail. A comprehensive benchmarking

against classical machine learning, established graph neural networks, and zero-shot large language models demonstrates that integrating non-manipulable micro-temporal signatures with semantic headline features yields a more robust detection signal than any single paradigm achieves in isolation. The results further show that purely semantic approaches, regardless of model scale, plateau at a ceiling that structural surrogates can consistently exceed. On the Fake-NewsNet PolitiFact benchmark, RGCP-Snowflake achieves a mean accuracy of $89.53\% \pm 1.58\%$ over 10 random seeds, outperforming all evaluated baselines including graph neural networks that rely on fully reconstructed social graphs, with the performance advantage confirmed statistically significant ($p < 0.01$) against the best-performing LLM baseline. Empirical analysis of inter-arrival time distributions further validates the burstiness hypothesis, confirming that fake news exhibits measurably distinct temporal engagement patterns (Mann-Whitney U, $p < 0.001$).

Several limitations bound these findings. The evaluation corpus of 1,052 samples from a single source is relatively small, and results may not generalise to larger collections, different outlets, or multilingual settings. The Snowflake surrogate method also assumes a predictable timestamp encoding, which would not hold on platforms using randomised or obfuscated identifiers. Additionally, the label collapse observed in sub-7B models was evaluated zero-shot only, and lightweight fine-tuning methods such as LoRA may recover reliable performance at smaller scales.

Future work will focus on evaluating across larger and more diverse datasets, exploring retrieval-augmented generation to provide LLMs with real-time factual grounding, and investigating hyperbolic graph representations to better model the hierarchical dynamics of rumour propagation.

References

- Bian, T.; Xiao, X.; Xu, T.; Zhao, P.; Huang, W.; Rong, Y.; and Huang, J. 2020. Rumor detection on social media with bi-directional graph convolutional networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 549–556.
- Das, R. K.; and Dodge, J. 2025. Fake news detection after LLM laundering: Measurement and explanation. *arXiv preprint arXiv:2501.18649*.
- Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186.
- Ding, Y.; Guo, B.; Liu, Y.; Jing, Y.; Yin, M.; Li, N.; Wang, H.; and Yu, Z. 2025. EvolveDetector: Towards an evolving fake news detector for merging events with continual knowledge accumulation and transfer. *Information Processing & Management*, 62(1): 103878.
- Dou, Y.; Shu, K.; Xia, C.; Yu, P. S.; and Sun, L. 2021. User preference-aware fake news detection. In *Proceedings of the 44th International ACM SIGIR Conference on Research and*

- Development in Information Retrieval (SIGIR '21)*, 2051–2055.
- Feng, S.; Yu, G.; Hu, H.; Liu, D.; Luo, Y.; Lin, H.; and Ong, Y.-S. 2025. MHR: A multi-modal hyperbolic representation framework for fake news detection. *IEEE Transactions on Knowledge and Data Engineering*.
- Hu, B.; Mao, Z.; and Zhang, Y. 2025. An overview of fake news detection: From a new perspective. *Fundamental Research*, 5: 332–346.
- Jing, P.; Gao, H.; Zhang, X.; Gao, T.; and Zhou, C. 2025. DPSG: Dynamic Propagation Social Graphs for multi-modal fake news detection. *Information Fusion*, 113: 102595.
- Jwa, H.; Oh, D.; Park, K.; Kang, J. M.; and Lim, H. 2019. exBAKE: Automatic fake news detection model based on bidirectional encoder representations from transformers (BERT). *Applied Sciences*, 9(19): 4062.
- Kaliyar, R. K.; Goswami, A.; and Narang, P. 2021. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia Tools and Applications*, 80(8): 11765–11788.
- Khattar, D.; Goud, J. S.; Gupta, M.; and Varma, V. 2019. MVAE: Multimodal variational autoencoder for fake news detection. In *Proceedings of the World Wide Web Conference (WWW '19)*, 2915–2921.
- Loth, A.; Kappes, M.; and Pahl, M.-O. 2024. Blessing or curse? A survey on the impact of generative AI on fake news. *arXiv preprint arXiv:2404.03021*.
- Ma, J.; Gao, W.; Mitra, P.; Kwon, S.; Jansen, B. J.; Wong, K.-F.; and Cha, M. 2016. Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, 3818–3824.
- Mridha, M. F.; Keya, A. J.; Hamid, M. A.; Monowar, M. M.; and Rahman, M. S. 2021. A comprehensive review on fake news detection with deep learning. *IEEE Access*, 9: 156151–156170.
- Nakamura, K.; Levy, S.; and Wang, W. Y. 2020. Fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection. In *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020)*, 6149–6157.
- Nguyen, V.-H.; Sugiyama, K.; Nakov, P.; and Kan, M.-Y. 2020. FANG: Leveraging social context for fake news detection using graph representation. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20)*, 1165–1174.
- Papageorgiou, E.; Varlamis, I.; and Chronis, C. 2025. Harnessing large language models and deep neural networks for fake news detection. *Information*, 16(4): 297.
- Raza, S.; Paulen-Patterson, D.; and Ding, C. 2025. Fake news detection: comparative evaluation of BERT-like models and large language models with generative AI-annotated data. *Knowledge and Information Systems*, 67: 3267–3292.
- Ruchansky, N.; Seo, S.; and Liu, Y. 2017. CSI: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM)*, 797–806.
- Shu, K.; Mahudeswaran, D.; Wang, S.; Lee, D.; and Liu, H. 2020. FakeNewsNet: A data repository with news content, social context and dynamic information for studying fake news on social media. *Big Data*, 8(3): 171–188.
- Shu, K.; Sliva, A.; Wang, S.; Tang, J.; and Liu, H. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1): 22–36.
- Singhal, S.; Shah, R. R.; Chakraborty, T.; Kumaraguru, P.; and Satoh, S. 2019. SpotFake: A multi-modal framework for fake news detection. In *Proceedings of the 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, 39–47.
- Tong, Z.; Gu, Y.; Liu, H.; Liu, Q.; Wu, S.; Shi, H.; and Zhang, X.-Y. 2025. Generate first, then sample: Enhancing fake news detection with LLM-augmented reinforced sampling. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.
- Vosoughi, S.; Roy, D.; and Aral, S. 2018. The spread of true and false news online. *Science*, 359(6380): 1146–1151.
- Wang, W. Y. 2017. “Liar, Liar Pants on Fire”: A new benchmark dataset for fake news detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 422–426.
- Wang, Y.; Ma, F.; Jin, Z.; Yuan, Y.; Xun, G.; Jha, K.; Su, L.; and Gao, J. 2018. EANN: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, 849–857.
- Wu, F.; Chen, S.; Gao, G.; Ji, Y.; and Jing, X.-Y. 2025. Balanced multi-modal learning with hierarchical fusion for fake news detection. *Pattern Recognition*, 164: 111485.
- Zhang, L.; Zhang, X.; Zhou, Z.; Zhang, X.; Yu, P. S.; and Li, C. 2025. Knowledge-aware multimodal pre-training for fake news detection. *Information Fusion*, 114: 102715.