INTERNATIONAL JOURNAL OF MODERN COMPUTER SCIENCE AND IT INNOVATIONS (IJMCSIT)

eISSN: 3087-4289

Volume. 02, Issue. 04, pp. 01-07, April 2025"

LEVERAGING PERSISTENCE AND GRAPH NEURAL NETWORKS FOR ENHANCED INFORMATION POPULARITY FORECASTING

Dr. Erik G. Johansson
Department of Computer and Information Science, Linköping University, Sweden

Dr. Linnea K. Blomqvist
Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden

Article received: 16/02/2025, Article Revised: 08/03/2025, Article Accepted: 10/04/2025
DOI: https://doi.org/10.55640/ijmcsit-v02i04-01

ABSTRACT

Accurately forecasting the popularity of online information is critical for optimizing content delivery, recommendation systems, and network resource allocation. This paper introduces a novel framework that leverages temporal persistence patterns and graph neural networks (GNNs) to improve the prediction of information popularity. By modeling user-content interactions as dynamic graphs and incorporating historical popularity trends, our approach captures both structural and temporal dependencies. Extensive experiments on real-world social and content-sharing platforms demonstrate that the proposed method significantly outperforms traditional forecasting models in terms of accuracy and robustness. The results highlight the potential of combining graph-based learning with temporal analysis for intelligent information propagation modeling.

Keywords: Information popularity forecasting, graph neural networks, temporal persistence, dynamic graphs, user-content interaction, content recommendation, predictive modeling, social networks, deep learning, information diffusion.

INTRODUCTION

The rapid proliferation of online social networks and digital platforms has transformed how information is disseminated and consumed. Understanding and predicting the popularity of information, whether it be news articles, social media posts, or viral videos, has become a critical area of research with significant implications for marketing, public health, content recommendation, and even crisis management [1]. Information diffusion often manifests as cascades, where an initial piece of content spreads through a network as individuals adopt and re-share it [2, 3]. The ability to accurately forecast the trajectory and ultimate reach of these information cascades is invaluable, yet it presents considerable challenges due to the complex interplay of content attributes, user behaviors, and underlying network structures [4, 5, 6].

Early research in information popularity prediction primarily focused on identifying influential users and key features of content or cascades [7, 8, 9]. These approaches often relied on statistical models or machine learning techniques applied to handcrafted features such as initial propagation speed, user engagement metrics, or network topology characteristics [10, 11, 12, 13, 14, 15]. While these methods provided valuable insights, they often struggled to capture the intricate temporal dynamics and the inherent graph-like structure of information flow in social networks. The temporal evolution of a cascade, including its initial surge and subsequent "persistence" or sustained activity over time, is crucial for accurate long-term prediction [13, 17, 18, 19]. Models like self-exciting point processes have been employed to capture these temporal dependencies, treating information cascades as events that trigger subsequent events [19, 20, 23].

More recently, the emergence of deep learning, particularly Graph Neural Networks (GNNs), has opened new avenues for modeling complex relational data. Information cascades, by their very nature, form directed acyclic graphs where nodes represent users or events and edges represent the propagation of information. GNNs, especially Graph Convolutional Networks (GCNs), are adept at learning representations from graph-structured data by aggregating information from neighboring nodes [26]. This makes them particularly well-suited for analyzing information diffusion paths and predicting future popularity [24, 25]. However, effectively integrating the "persistence" aspect—the sustained influence and temporal patterns beyond initial bursts—into GCN frameworks remains an underexplored challenge. Existing GCN-based models for diffusion prediction often focus heavily on structural aggregation at a given snapshot or combine GCNs with Recurrent Neural Networks (RNNs) to handle sequences of events [27, 28, 29]. While effective for capturing immediate temporal dependencies, they may not fully leverage the long-term, non-linear persistence characteristics of information spread.

This article proposes a novel approach that augments Graph Neural Networks with a dedicated mechanism to capture information persistence for enhanced popularity forecasting. By integrating features that reflect the sustained activity and decaying influence of information over time directly into the graph learning process, our model aims to provide a more comprehensive understanding and prediction of information popularity. The remainder of this article is structured as follows: Section 2 details the proposed methodology, including the graph construction, persistence augmentation, and the GCN architecture. Section 3 presents the experimental results and comparative analysis. Section 4 discusses the findings, implications, limitations, and future research directions.

METHODS

Problem Formulation and Graph Representation

Information popularity prediction can be framed as forecasting the future engagement (e.g., number of retweets, likes, shares, or views) an item of information will receive within a specific timeframe, given its initial propagation history. We represent an information cascade as a directed graph G=(V,E), where V is the set of users involved in the cascade, and E is the set of directed edges representing the flow of information from one user to another (e.g., a retweet from user u to user v). Each node v∈V can be associated with features such as user attributes (e.g., follower count, activity level) and temporal features (e.g., time of interaction). The information item itself can also have content-based features (e.g., text, image, video characteristics) [10, 12, 22].

Given an observed partial cascade up to time T, the goal is to predict its total size or future popularity at a later time T′>T. The challenge lies in effectively encoding both the structural information (who influences whom) and the temporal dynamics (when these influences occur and how long they last) [18].

Persistence Augmentation

To capture the "persistence" of information, we introduce a set of features that quantify the sustained engagement and decay patterns of a cascade. Unlike models that rely solely on the raw sequence of events or simple aggregation of initial activity, our persistence augmentation aims to provide the GCN with a richer understanding of how long and how strongly information resonates within the network. These features are incorporated as node-level or edge-level attributes in the graph representation. Examples of persistence features include:

- Temporal Decay Rates: For each node (user interaction), we calculate a decay rate based on the time elapsed since the initial post and the time interval between successive interactions. This helps the GCN understand how "old" or "fresh" a piece of information is for a specific propagation path.

- Activity Windows: We define multiple time windows (e.g., 1 hour, 6 hours, 24 hours) and compute the number of interactions within each window for different parts of the cascade. This provides a multi-scale view of activity [13].

- Engagement Ratios: Ratios like retweets-per-follower or comments-per-view can indicate the intensity of engagement and potential for continued spread, reflecting the intrinsic "stickiness" of the content [21].

- Recency of Active Paths: For any given node in the cascade graph, we can compute the recency of the most recent interaction along the path leading to it. This highlights which parts of the cascade are still "alive" and propagating.

These persistence features are integrated into the initial feature vector of each node (hv(0)) alongside traditional content and user features. This enriches the input to the Graph Convolutional Layers, allowing the network to implicitly learn how persistence affects future popularity.

Graph Convolutional Network Architecture

Our model employs a multi-layer Graph Convolutional Network (GCN) to learn node embeddings that capture both local neighborhood information and global cascade structure, enhanced by persistence features. The core operation of a GCN layer is to aggregate feature vectors from a node's neighbors and transform them, generating a new representation for the node.

The propagation rule for a single GCN layer is defined as:
H(l+1)=σ(D~−21A~D~−21H(l)W(l))

where:

- H(l) is the matrix of node features at layer l, with H(0) being the initial feature matrix including persistence features.
- A~=A+IN is the adjacency matrix A of the graph with added self-loops (IN is the identity matrix).
- D~ is the diagonal degree matrix of A~.
- W(l) is the weight matrix for layer l, which is learned during training.
- σ is an activation function (e.g., ReLU).

In our architecture, we use multiple GCN layers to allow each node to aggregate information from increasingly distant neighbors, effectively capturing the global structure of the information cascade [26]. The output of the final GCN layer provides a rich, low-dimensional embedding for each node in the cascade graph.

Prediction Layer

After obtaining the node embeddings from the GCN, these embeddings are pooled to form a single, fixed-size representation of the entire information cascade. Various pooling strategies can be employed, such as sum pooling, average pooling, or attention-based pooling, which assigns different weights to node embeddings based on their importance [28, 29]. For instance, a simple sum pooling can be represented as:

Hcascade=v∈V∑Hv(L)

where Hv(L) is the final embedding for node v from the last GCN layer L.

This cascade-level representation (Hcascade) is then fed into a fully connected neural network (MLP) with one or more hidden layers, followed by an output layer. For regression tasks (predicting the exact number of future engagements), the output layer will have a single neuron with a linear activation. For classification tasks (e.g., predicting if popularity will exceed a certain threshold), a sigmoid or softmax activation would be used.

The model is trained using standard optimization techniques like Adam, minimizing a suitable loss function such as Mean Squared Error (MSE) for regression or Binary Cross-Entropy for classification.

Experimental Setup

Datasets

To evaluate the proposed model, we utilized two public datasets commonly used in information cascade prediction research:

1. Twitter Dataset: A collection of retweet cascades from Twitter, encompassing diverse topics. Each cascade includes the retweet graph, timestamps of interactions, and user metadata. The task is to predict the total number of retweets within a 24-hour or 7-day window after the initial post, given the first few hours of observation (e.g., first 1-2 hours) [4, 5, 6].

2. Weibo Dataset: A microblogging dataset from Sina Weibo, similar to Twitter, providing cascades of reposts. This dataset offers rich content features alongside structural and temporal information [9].

Baseline Models

We compared our Persistence Augmented GCN (PA-GCN) against several state-of-the-art and representative baseline models:

- Feature-Based Regression (FBR): A traditional approach using hand-crafted features (e.g., initial spread rate, number of unique users, user influence scores) fed into a linear regression or gradient boosting model [14, 15].

- SEISMIC: A self-exciting point process model that captures temporal dynamics by modeling how each event in a cascade triggers subsequent events [19].

- DeepCas: A deep learning model that leverages a Recurrent Neural Network (RNN) to process sequences of cascade events and predicts future popularity [26].

- Topological RNN (TRNN): A model that combines topological features with an RNN to predict diffusion [24].

- CasSeqGCN: A recent GCN-based model that combines network structure and temporal sequence for prediction [27].

Evaluation Metrics

For popularity prediction, which is typically a regression task, we used the following metrics:

- Mean Absolute Error (MAE): Average absolute difference between predicted and actual values.

- Root Mean Squared Error (RMSE): Measures the square root of the average of the squared errors, penalizing larger errors more heavily.

• R-squared (R2): Indicates the proportion of the variance in the dependent variable that is predictable from the independent variables.

RESULTS

Our experimental evaluations demonstrate the significant performance gains achieved by the proposed Persistence Augmented GCN (PA-GCN) compared to the baseline models across both Twitter and Weibo datasets. The inclusion of persistence features proved crucial in enhancing the model's ability to capture the long-term popularity dynamics of information cascades.

Performance Comparison

Table 1 summarizes the performance of PA-GCN and baseline models on the Twitter and Weibo datasets, focusing on predicting the total cascade size within a 7-day window.

Table 1: Performance Comparison of PA-GCN with Baseline Models (Lower MAE/RMSE, Higher R2 are better)

Model	Twitter (MAE)	Twitter (RMSE)	Twitter (R2)	Weibo (MAE)	Weibo (RMSE)	Weibo (R2)
FBR	125.6	289.1	0.52	88.2	195.3	0.48
SEISMIC	102.3	234.5	0.61	75.1	170.8	0.55
DeepCas	98.7	221.9	0.65	71.4	162.7	0.58
TRNN	95.1	215.2	0.67	69.8	158.9	0.60
CasSeqGCN	91.2	208.7	0.69	67.5	153.1	0.63
PA-GCN	**82.9**	**188.4**	**0.75**	**61.8**	**139.5**	**0.69**

As shown in Table 1, PA-GCN consistently outperforms all baseline models across all evaluation metrics on both datasets. Specifically, PA-GCN achieved a 9.1% reduction in MAE and a 9.7% reduction in RMSE compared to the next best model, CasSeqGCN, on the Twitter dataset. Similar improvements were observed on the Weibo dataset, with a 8.4% reduction in MAE and a 8.9% reduction in RMSE over CasSeqGCN. The higher R2 values indicate that our model explains a greater proportion of the variance in information popularity, suggesting a more robust predictive capability.

Ablation Study of Persistence Features

To validate the impact of persistence features, we conducted an ablation study on the PA-GCN model, removing the persistence features and comparing its performance against the full model.

Table 2: Ablation Study of Persistence Features (Twitter Dataset)

Model	MAE	RMSE	R2
PA-GCN (w/o Pers.)	96.5	218.0	0.66
PA-GCN (Full)	82.9	188.4	0.75

Table 2 clearly demonstrates that the inclusion of persistence features significantly contributes to the superior performance of PA-GCN. Without these features, the model's performance drops considerably, approaching that of other GCN-based methods like CasSeqGCN. This highlights the effectiveness of explicitly encoding the temporal decay, multi-scale activity, and engagement ratios within the graph representation.

Analysis of Convergence and Efficiency

The training process for PA-GCN exhibited stable convergence, typically within 50-70 epochs for both datasets. The computational overhead introduced by the persistence feature calculation was minimal and outweighed by the gains in predictive accuracy. The GCN architecture itself, while more complex than simple feature-based models, is designed for efficient parallel processing on GPUs, making it suitable for large-scale social network data.

DISCUSSION

The results presented unequivocally demonstrate the effectiveness of leveraging persistence features within a Graph Convolutional Network framework for information popularity forecasting. Our proposed PA-GCN model consistently outperformed established baselines, including advanced temporal models and other GCN-based approaches, across diverse real-world social media datasets. This superior performance can be attributed to the model's ability to holistically capture both the structural propagation paths and the nuanced temporal decay and sustained engagement patterns inherent in information cascades.

Traditional feature-based methods, while providing initial insights, often oversimplify the complex dynamics of information spread by relying on aggregate statistics that may not fully represent the micro-level interactions and their temporal evolution [14, 15]. Point process models like SEISMIC [19] excel at modeling sequential events but may not fully leverage the topological structure of the underlying social network. Deep learning models, such as DeepCas [26] and TRNN [24], have made strides by using RNNs to process cascade sequences, but they can struggle to integrate the non-sequential, graph-level interactions effectively.

Our approach addresses these limitations by augmenting the graph representation with specifically engineered persistence features. These features, such as temporal decay rates and activity window summaries, provide explicit signals to the GCN about how information "lives" and "dies" over time, beyond just its initial spread. By incorporating these directly into the node feature vectors, the GCN layers can learn to propagate and aggregate information that includes rich temporal context. This allows the model to differentiate between cascades that experience a rapid, short-lived burst of popularity and those that exhibit sustained, long-term engagement, leading to more accurate long-term popularity predictions.

The ablation study confirmed the critical role of these persistence features. The substantial drop in performance when these features were removed underscores that the gains are not merely from using a GCN but from the synergistic combination of the graph learning capabilities with an intelligent encoding of temporal persistence. This suggests that for dynamic graph problems like information diffusion, enriching the graph's static or semi-static features with indicators of dynamic behavior is highly beneficial.

Despite its strengths, the PA-GCN model has certain limitations. The engineered persistence features require careful domain knowledge and may not be universally optimal across all types of information or social networks. While effective, they still represent an explicit feature engineering step rather than a fully end-to-end learning of persistence directly from raw temporal signals. Furthermore, the model's interpretability, while better than some end-to-end black-box deep learning models, could be further enhanced to understand which specific persistence features contribute most to the prediction for a given cascade.

Future research could explore several exciting directions. First, integrating more advanced temporal graph neural networks or dynamic graph embedding techniques could allow for an even more intrinsic learning of persistence without explicit feature engineering. This might involve incorporating attention mechanisms that weigh temporal edges differently or using temporal convolution layers. Second, exploring the application of PA-GCN to other types of information diffusion, such as news propagation or meme spread, on different platforms could validate its generalizability. Finally, investigating methods to provide uncertainty estimates alongside popularity predictions would be valuable for practical applications, enabling more robust decision-making in areas like targeted advertising or virality assessment.

In conclusion, this work demonstrates that enhancing Graph Convolutional Networks with carefully designed persistence features significantly improves the accuracy of information popularity forecasting. By bridging the gap between static graph structure and dynamic temporal activity, the proposed PA-GCN offers a powerful and effective framework for understanding and predicting the complex phenomenon of information diffusion in online social networks.

REFERENCES

F. Zhou, X. Xu, G. Trajcevski, and K. Zhang, “A survey of information cascade analysis: Models, predictions, and recent advances,” ACM Comput. Surv., vol. 54, no. 2, pp. 1–36, 2021.

R. A. Ibrahim, H. A. Hefny, and A. E. Hassanien, “Controlling social information cascade: A Survey,” in Big Data Analytics, 1st ed. CRC Press, 2018, pp. 196–212.

A. Guille, H. Hacid, C. Favre, and D. A. Zighed, “Information diffusion in online social networks: A survey,” ACM Sigmod Rec., vol. 42, no. 2, pp. 17–28, 2013.

T. R. Zaman, R. Herbrich, J. Van Gael, and D. Stern, “Predicting information spreading in twitter,” in Proc. Workshop Comput. Social Sci. Wisdom Crowds, vol. 104, no. 45. 2010, pp. 17599–601.

K. Lee, J. Mahmud, J. Chen, M. Zhou, and J. Nichols, “Who will retweet this? automatically identifying and engaging strangers on twitter to spread information,” in Proc. 19th Int. Conf. Intell. User Interfaces, 2014, pp. 247–256.

Z. Luo, M. Osborne, J. Tang, and T. Wang, “Who will retweet me? finding retweeters in twitter,” in Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2013, pp. 869–872.

T. Zaman, E. B. Fox, and E. T. Bradlow, “A Bayesian approach for predicting the popularity of tweets,” Ann. Appl. Statist., vol. 8, no. 3, pp. 1583–1611, 2014.

S. Jamali and H. Rangwala, “Digging digg: Comment mining, popularity prediction, and social network analysis,” in Proc. IEEE Int. Conf. Web Inf. Syst. Mining, 2009, pp. 32–38.

P. Bao, H.-W. Shen, J. Huang, and X.-Q. Cheng, “Popularity prediction in microblogging network: A case study on sina weibo,” in Proc. 22nd Int. Conf. World Wide Web, 2013, pp. 177–178.

J. Chen, X. Song, L. Nie, X. Wang, H. Zhang, and T.-S. Chua, “Micro tells macro: Predicting the popularity of micro-videos via a transductive model,” in Proc. 24th ACM Int. Conf. Multimedia, 2016, pp. 898–907.

M. X. Hoang, X.-H. Dang, X. Wu, Z. Yan, and A. K. Singh, “GPOP: Scalable group-level popularity prediction for online content in social networks,” in Proc. 26th Int. Conf. World Wide Web, 2017, pp. 725–733.

J. Lv, W. Liu, M. Zhang, H. Gong, B. Wu, and H. Ma, “Multi-feature fusion for predicting social media popularity,” in Proc. 25th ACM Int. Conf. Multimedia, 2017, pp. 1883–1888.

B. Wu, T. Mei, W.-H. Cheng, and Y. Zhang, “Unfolding temporal dynamics: Predicting social media popularity using multi-scale temporal decomposition,” in Proc. 13th AAAI Conf. Artif. Intell., 2016, pp. 272–278.

S. Gao, J. Ma, and Z. Chen, “Effective and effortless features for popularity prediction in microblogging network,” in Proc. 23rd Int. Conf. World Wide Web, 2014, pp. 269–270.

C. Yi, Y. Bao, and Y. Xue, “Mining the key predictors for event outbreaks in social networks,” Physica A: Stat. Mechanics Appl., vol. 447, pp. 247–260, 2016.

S. Bourigault, S. Lamprier, and P. Gallinari, “Representation learning for information diffusion through social networks: An embedded cascade model,” in Proc. 9th ACM Int. Conf. Web Search Data Mining, 2016, pp. 573–582.

Q. Kong, M.-A. Rizoiu, and L. Xie, “Modeling information cascades with self-exciting processes via generalized epidemic models,” in Proc. 13th Int. Conf. Web Search Data Mining, 2020, pp. 286–294.

L. Yu, P. Cui, F. Wang, C. Song, and S. Yang, “Uncovering and predicting the dynamic process of information cascades with survival model,” Knowl. Inf. Syst., vol. 50, no. 2, pp. 633–659, 2017.

Q. Zhao, M. A. Erdogdu, H. Y. He, A. Rajaraman, and J. Leskovec, “Seismic: A self-exciting point process model for predicting tweet popularity,” in Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2015, pp. 1513–1522.

H. Shen, D. Wang, C. Song, and A.-L. Barabási, “Modeling and predicting popularity dynamics via reinforced poisson processes,” in Proc. AAAI Conf. Artif. Intell., vol. 28, no. 1, pp. 291–297, 2014.

N. Garg, M. Sellathurai, V. Bhatia, B. N. Bharath, and T. Ratnarajah, “Online content popularity prediction and learning in wireless edge caching,” IEEE Trans. Commun., vol. 68, no. 2, pp. 1087–1100, Feb. 2020.

G. Ver Steeg and A. Galstyan, “Information-theoretic measures of influence based on content dynamics,” in Proc. 6th ACM Int. Conf. Web Search Data Mining, 2013, pp. 3–12.

Q. Cao, H. Shen, K. Cen, W. Ouyang, and X. Cheng, “Deephawkes: Bridging the gap between prediction and understanding of information cascades,” in Proc. 2017 ACM Conf. Inf. Knowl. Manage., 2017, pp. 1149–1158.

J. Wang, V. W. Zheng, Z. Liu, and K. C.-C. Chang, “Topological recurrent neural network for diffusion prediction,” in Proc. IEEE Int. Conf. Data Mining, 2017, pp. 475–484.

X. Chen, F. Zhou, K. Zhang, G. Trajcevski, T. Zhong, and F. Zhang, “Information diffusion prediction via recurrent cascades convolution,” in Proc. IEEE 35th Int. Conf. Data Eng., 2019, pp. 770–781.

C. Li, J. Ma, X. Guo, and Q. Mei, “DeepCas: An end-to-end predictor of information cascades,” in Proc. 26th Int. Conf. World Wide Web, 2017, pp. 577–586.

Y. Wang, X. Wang, R. Michalski, Y. Ran, and T. Jia, “CasSeqGCN: Combining network structure and temporal sequence to predict information cascades,” Expert Syst. Appl., vol. 206, Nov., 2022, Art. no. 117693, doi: 10.1016/j.eswa.2022.117693.

M. Wang and K. Li, “Predicting information diffusion cascades using graph attention networks,” in Proc. Int. Conf. Neural Inf. Process., 2020, pp. 104–112.

Z. Huang, Z. Wang, R. Zhang, Y. Zhao, and F. Zheng, “Learning Bi-directional social influence in information cascades using graph sequence attention networks,” in Proc. Companion Web Conf., 2020, pp. 19–21.

R. Wang, “Dydiff-VAE: A dynamic variational framework for information diffusion prediction,” in Proc. 44th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2021, pp. 163–172.