Proceedings of the 4th International Conference on Applied Physics and Mathematical Modeling
DOI: 10.54254/2753-8818/2026.36905

Information Cascade Popularity Prediction Using Structural and Temporal Features

Hongfan Zhu

Phillips Exeter Academy, Exeter, USA
hzhu5@exeter.edu

Abstract. Prediction of the spread of information cascades can help explain how information spreads in society and can support multiple applications. Early models often used aggregated user data and linear approximations; more recent approaches use graph representations to capture complex spatio-temporal variations. This paper studies the development of popularity prediction models and introduces some types, such as feature-engineered statistical methods, advanced representation learning and generative frameworks. The basic ideas of online diffusion are introduced by formalising cascade and global graphs, and the extraction of structural and temporal features is described. An introduction to the new frameworks is provided here, including random walk models (DeepWalk, node2vec), graph convolutional networks (GCNs), and the newer CasDO (Cascade Diffusion and Neural ODEs), which introduces continuous-time dynamics to address irregular sampling and diffusion uncertainty. The above comparisons highlight the shift toward modelling stochastic cascade dynamics. Finally, the current limitations, such as cross-domain applicability, are mentioned, and the potential integration of large language models (LLMs) and context-aware generation in future research is also introduced.

Keywords: information cascade, popularity prediction, graph neural networks, temporal dynamics, structural features

1. Introduction

Information spreads rapidly across social networks in the digital age through posts, comments and shares. User-sharing behaviour generates a sharp peak in attention and thus viral content; however, the distribution of popularity is still heavily long-tailed, and only a small number of such works receive widespread attention [1]. The rapid, sequential transmission of information between users that affects decision-making is known as an information cascade. Accurate prediction can support trend monitoring, information management and public communication [2].

The basic problem in predicting the popularity of information cascades is that, although these pieces of information may start with similar initial popularities, they can have different ranges, depths and durations of spread later on [2]. Although some internet memes spread strongly within local communities, only a few spread across multiple communities globally [3].

Naturally, many researchers have focused on popularity prediction to date, building models based on user behaviour (either aggregated or individual) and observation windows. Given the

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.36905

shortcomings of the previous linear methods and metadata-dependent models, recent research has moved towards more advanced graph-learning models and dynamic continuous-time prediction [2, 4]. Based on the analysis of several previous predictive models, this paper will explore how structural and temporal features of information cascades can be used to predict them. The rest of this survey will present the basic concepts of network propagation, investigate structural and temporal indicators in greater detail, review some recent representative studies (such as Cascade Diffusion and Neural ODEs (CasDO), graph convolutional networks (GCNs), and node embedding techniques), and finally discuss the limitations and future directions.

2. Basic concepts of information cascades and online diffusion

The online social network and its underlying information cascades are mathematically represented as global graphs and cascade graphs, respectively. In global graphs, nodes are users, and edges represent the connections or associations among users, often weighted by the quantity of past interactions. Cascade graphs show the observed propagation structure of a piece of information and are generally modelled as directed trees or directed acyclic graphs, depending on the platform and data. The root of a cascade graph is the original source of information; nodes are users who have spread this information, and directed edges show how it was spread [2].

Cascade graphs are cumulative; they expand over time due to user-sharing behaviour, and dynamic propagation occurs as dissemination networks are gradually built from the initial publishers [3]. An adjacency matrix can represent the graph in a computationally feasible form; a non-zero value at row i and column j indicates a directed edge from node i to node j, and this value is the weight of the edge. A related transformation is the unnormalised graph Laplacian, calculated as the diagonal degree matrix minus the adjacency matrix. It is used in signal processing and spectral graph wavelets [2].

Prediction frameworks extract several structural indicators of cascade graphs to quantify diffusion behaviour. The primary measure of popularity is cascade size, which is the total number of nodes in a cascade tree. The vertical and horizontal extents of diffusion are shown by the maximum distance from the root node in hops (depth), and the width is the maximum number of nodes at any depth or level of the cascade. Depth and width are both measures of the cascade shape, and structural virality is generally expressed as a path-based quantity, such as the average pairwise distance between nodes.

At the node level, in-degree and out-degree are the number of incoming and outgoing edges, respectively. In addition, other centrality measures of a node in a network include degree, betweenness, eigenvector and closeness centrality.

At the same time, the elementary temporal indicators include publication time, repost interval, popularity growth rate, peak time and total lifecycle. Peak time refers to a period of high growth rate, and the end of a life cycle is characterised by a decline in popularity and a lower growth rate [2].

3. Structural and temporal features for cascade popularity prediction

Extraction of structural and temporal information for cascade prediction has moved from manual topological heuristics to automatic end-to-end representation learning, and recently, to hybrid dynamic models.

DOI: 10.54254/2753-8818/2026.36905

3.1. Evolution of structural features

In the traditional feature-engineering model, structural features were manually extracted scalar attributes of cascade and global graphs, such as cascade depth, width, and Wiener index. These numerical values served as the input "feature vectors" for the machine learning model. Manual feature extraction may not fully capture information about graph topology and node roles. For example, a cascade that is generally classified as a "broadcast configuration" can take very different local forms, and nodes with the same centrality score may have different functions during diffusion [2].

Therefore, representation-learning models use learned embeddings to incorporate cascade and global graph information. Given that extremely large graphs can be computationally expensive to process, sampling or embedding methods are used to obtain local representations of the model [2]. These methods produce more compact structural representations.

3.2. Temporal dynamics and the shift to continuous time

Temporal features describe how cascade popularity evolves over time. Traditional temporal analysis measured popularity at regular time intervals and obtained growth rates and curves for analysis. Early models generally assumed a linear or logistic relationship between the popularity observed in the initial time window and the popularity in the prediction window. Additionally, probabilistic point processes, such as the Hawkes self-exciting process, have been introduced to account for the fact that each past event contributes a triggering kernel which decays over time, thus modelling events at their exact times of occurrence [2].

However, the actual behaviour of users is highly irregular and random. Users post or comment at different times of the day, so the intervals between events are naturally irregular. Regular sampling may miss or distort the time of irregular occurrences and lose information about bursty diffusion dynamics. Therefore, the current models have shifted to continuous-time dynamics and neural ordinary differential equations (ODEs) to address time-series problems by maintaining a continuous hidden state rather than using only discrete-time updates [4, 5].

4. Representative graph-learning methods for cascade popularity prediction

Recently, more research has focused on graph learning frameworks trained with large-scale real-world data. The new models seek to incorporate the micro-level structural information and macro-level long-term effects of cascade graphs [6, 7].

4.1. Random walk graph embeddings with DeepWalk and node2vec

Early breakthroughs in network representation were driven by random walk models, such as DeepWalk and node2vec. These models generate embeddings by simulating random walks on the global or cascade graphs, and map nodes that appear together in these walks closer to each other in a continuous vector space [8, 9]. DeepWalk is a uniform random walk method that uses a Skip-Gram neural network to learn latent representations. Node2vec introduces a biased random walk strategy and uses return and in-out hyperparameters to balance between breadth-first sampling (capturing structural equivalence) and depth-first sampling (capturing homophily/proximity) [9]. Although these models reduce the requirement for hand-crafted features, they are inherently static; they can

DOI: 10.54254/2753-8818/2026.36905

extract structural information efficiently but cannot natively incorporate the temporal fluidity and continuous evolution of an active information cascade [2].

4.2. Graph convolutional networks and related models

Graph Convolutional Networks (GCNs) use graph topology and message-passing to aggregate features from the local neighbourhood of a node. GCN-based methods can handle graph-structured input and learn representations that consider the local connectivity pattern [2, 7].

Recently, many new graph-learning models for cascade prediction have been proposed. The Temporal-Spatial Cascade Convolutional Learning Framework (ViralGCN) was proposed by Xu and Qian and employs bi-directional spatial convolutional layers. ViralGCN obtains the local structural information of each node from both directions of information inflow and outflow in the cascade graph, and experiments have shown that both structural and temporal features can predict popularity [10].

Casformer is also a graph-based Transformer that uses an adaptive cascade sampling strategy [6]. A Cascade Attention Network (CAT) is used to capture fine-grained, time-aware structural details, and a Transformer learns the large-scale diffusion effect [6]. CasDacGCN is another typical type that uses a bidirectional cross-attention fusion module and a hypernetwork-based calibration strategy to address the decoupling of spatial and temporal signals in sparse or rapidly changing cascades [7].

4.3. CasDO for cascade popularity prediction

CasDO is a new hybrid model that combines cascade diffusion modelling with neural ordinary differential equations for information popularity prediction. Cheng et al. proposed CasDO to handle irregular sampling and diffusion uncertainty by integrating a probabilistic diffusion model with neural networks [5].

CasDO employs probabilistic diffusion mechanics in a conditioned latent space to address uncertainty in social-media sharing behaviour. It operates by injecting noise in the forward diffusion process and reconstructing cascade embeddings in the reverse process [5].

CasDO departs from discrete recurrent models by using a Temporal Neural ODE (T-ODE) network. This approach extends discrete state transitions to continuous-time dynamics [4, 5]. By solving an ODE between consecutive observations, CasDO maintains continuous hidden states and updates them at irregularly timed events. Structurally, it embeds the cascade graph using spectral graph wavelets derived from Chebyshev polynomials on the graph Laplacian. CasDO uses sparse matrix factorisation to generate scalable structural-proximity embeddings of the entire social network [5].

5. Comprehensive summary based on recent research

Recent research on information cascade popularity prediction has moved from macro-level approximations toward micro-level spatiotemporal and continuous-time modelling.

Early feature-based statistical models were mainly user-aggregated and post-publication. They framed the problem as a classification or regression problem and relied heavily on metadata, such as early views or subscriber counts. Feature-based models perform well in some environments but are very sensitive to the selection of features and prediction formulation [2]. Recurrent models later added sequential temporal information, but discrete-time formulations can struggle with irregularly timed events [4, 5].

DOI: 10.54254/2753-8818/2026.36905

Several studies have shown that structural and temporal features are highly predictive in some datasets, but the relative contribution of content varies depending on the platform, task and feature design [2, 10]. ViralGCN has also shown that structural and temporal cascade information improve the prediction performance on the test set [10]. Based on the above results, network structure and temporal changes can be used as features [2, 10].

Benchmark evaluations in these studies often use datasets from platforms such as Twitter and Sina Weibo, as well as citation-network data, and measure cascade popularity through retweets, reposts or citations [2, 6]. Many indicators are used to evaluate the model, such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and classification-oriented measures [2, 6, 7]. A typical problem is that the observation periods in the studies are not uniform [2].

6. Limitations and prospects

Although new models have been developed, some problems with cascade prediction still exist. First, there is no uniform standard for selecting observation windows and curating datasets; this can hinder fair comparison between models. Public social-media datasets may also be incomplete due to data-access and platform restrictions [2]. Platform-specific noise and changing community structures may also affect the observed diffusion pattern [11].

Future research will focus on cross-domain applicability to improve the generalisation ability of models trained for one set of users or platforms to others [1, 2]. In addition, tracking how information spreads across evolving communities remains relevant to cross-group diffusion [11].

Although recent deep learning models focus on topology and temporal dynamics, there is now growing interest in integrating large language models (LLMs) and rich semantic representations for popularity prediction [1, 12]. LLM-based approaches can obtain semantic representations of text and metadata to extend structural models [12]. Future frameworks should combine structural graph models with contextual embeddings from LLMs to better distinguish propagation mechanisms in different domains [1, 12].

Future evaluations should therefore use consistent data splits and observation windows, report prediction errors under comparable settings, and test whether results transfer across datasets, platforms and user groups. Anomalies and missing values should also be handled consistently so that reported gains can be compared more reliably.

7. Conclusion

Information cascade popularity prediction is at the intersection of graph theory, time-series analysis and behaviour modelling. Feature-based statistical models and representation-learning methods, such as GCNs, node2vec and DeepWalk, have all been used to study how to represent cascade structure and temporal evolution. Recently, frameworks such as ViralGCN, Casformer, CasDacGCN and CasDO have combined graph-based representations with temporal modelling, adaptive sampling, attention mechanisms or continuous-time dynamics. CasDO shows how neural ODEs and probabilistic diffusion can handle irregular event timing and uncertainty in information diffusion. At the same time, a single model strategy is not suitable for all platforms and cascade types. The error in the prediction may be due to the observation window, lack of structural and temporal information, incomplete data set, platform-specific noise, and evaluation method. Structure and time are good predictive indicators, and content and semantic information can be utilised in different ways in various datasets and fields. Therefore, the current deficiencies include irregular observation

DOI: 10.54254/2753-8818/2026.36905

conditions, incomplete social media data, cross-domain generalisation issues, and difficulty in interpreting complex deep models. Future work will use consistent data splits and observation windows to evaluate the model, test cross-platform and user group transferability, and report uncertainty along with prediction errors. Future work should add more semantic information to the graph structure, such as that provided by large language models, and differentiate between different diffusion modes when such data are available. Additional validation across datasets and environments is also needed; future studies should carefully handle anomalies or missing values and determine whether the reported gains can be generalised to other environments. In short, in order to achieve high-fidelity prediction of cascade propagation, structural, temporal and contextual features from multiple sources need to be integrated.

References

[1] Liu, Y., Lei, Q., Liu, Z., Zhao, M., Xu, Z., & Shi, L. (2026). A survey on social media popularity prediction: Concepts, technologies, applications, and challenges. *Journal of King Saud University Computer and Information Sciences, 38*, Article 501. https://doi.org/10.1007/s44443-026-00909-x

[2] Wang, Z., Wang, X., Xiong, F., & Chen, H. (2024). A survey of deep learning-based information cascade prediction. *Symmetry, 16*(11), Article 1436. https://doi.org/10.3390/sym16111436

[3] Weng, L., Menczer, F., & Ahn, Y.-Y. (2013). Virality prediction and community structure in social networks. *Scientific Reports, 3*, Article 2522. https://doi.org/10.1038/srep02522

[4] Jing, X., Jing, Y., Lu, Y., Deng, B., Yang, S., & Yang, D. (2025). On your mark, get set, predict! Modeling continuous-time dynamics of cascades for information popularity prediction. *IEEE Transactions on Knowledge and Data Engineering, 37*(9), 5436-5451. https://doi.org/10.1109/TKDE.2025.3583129

[5] Cheng, Z., Zhou, F., Xu, X., Zhang, K., Trajcevski, G., Zhong, T., & Yu, P. S. (2024). Information cascade popularity prediction via probabilistic diffusion. *IEEE Transactions on Knowledge and Data Engineering, 36*(12), 8541-8555. https://doi.org/10.1109/TKDE.2024.3465241

[6] Wang, B., Li, Z., Xu, Z., & Zhang, J. (2025). Casformer: Information popularity prediction with adaptive cascade sampling and graph transformer in social networks. *IEEE Transactions on Big Data, 11*(4), 1652-1663. https://doi.org/10.1109/TBDATA.2024.3524839

[7] Zhang, B., Zhu, Y., Zhang, Z., Liao, K., Niu, S., Li, B., & Li, H. (2025). CasDacGCN: A dynamic attention-calibrated graph convolutional network for information popularity prediction. *Entropy, 27*(10), Article 1064. https://doi.org/10.3390/e27101064

[8] Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 701-710). ACM. https://doi.org/10.1145/2623330.2623732

[9] Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 855-864). ACM. https://doi.org/10.1145/2939672.2939754

[10] Xu, Z., & Qian, M. (2023). Predicting popularity of viral content in social media through a temporal-spatial cascade convolutional learning framework. *Mathematics, 11*(14), Article 3059. https://doi.org/10.3390/math11143059

[11] Ma, Y., Qu, D., & Wang, Y. (2026). Tracking evolving communities in fake news cascades using temporal graphs. *Scientific Reports, 16*, Article 4952. https://doi.org/10.1038/s41598-026-35175-4

[12] Chen, T., Huang, J., Wu, X., & Shao, C. (2024). Predicting social media popularity with large language models: Transforming metadata into semantic-enriched and contextualized text. *IEEE Access, 12*, 192528-192538. https://doi.org/10.1109/ACCESS.2024.3484762