Review

A Survey of Deep Learning-Based Information Cascade Prediction

Zhengang Wang [1], Xin Wang [2], Fei Xiong [1,*] and Hongshu Chen [3]

[1] School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China; 22120145@bjtu.edu.cn
[2] Beijing Institute of Computer Technology and Application, Beijing 100039, China; 15010899917@163.com
[3] School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China
[*] Correspondence: xiongf@bjtu.edu.cn

Abstract: Online social media have significantly boosted the creation and transmission of information, accelerating the dissemination and interaction of vast amounts of data, thereby making the prediction of information cascades increasingly important. In recent years, deep learning has been extensively applied in the domain of information cascade prediction. This paper primarily classifies, organizes, and summarizes the current research status and classic algorithms of information cascade prediction methods based on deep learning. According to the different focuses on characterizing information cascade features, studies on deep learning-based information cascade prediction are classified from two perspectives, i.e., prediction targets and prediction methods. Each category is explained in detail, along with its principles, advantages, and disadvantages, and the commonly used datasets and evaluation metrics in this field are introduced. Additionally, this paper explores the role of symmetry in the structural patterns of information diffusion networks, analyzing how symmetry impacts the pathways and efficiency of information dissemination. Finally, this paper summarizes the potential future research directions and development trends in this domain.

Keywords: cascade prediction; neural network; deep learning; popularity prediction

1. Introduction

With the rapid development of the internet and continuous upgrades in wireless communication technology, social media networks have rapidly spread across the world, transforming how individuals acquire data and utilize it in their interactions. Social networks now serve as primary platforms for information gathering and engagement. Information dissemination on social networks often exhibits the characteristics of cascade propagation, meaning that once information is released, it may trigger a chain reaction across the network, affecting more users. In many cases, the structure of these networks exhibits a degree of symmetry, which plays a critical role in the efficiency of information dissemination. Symmetrical structures can influence the speed and scope of cascade propagation, making symmetry an important factor to consider in understanding and predicting information cascades. For example, retweets of Weibo posts [1], citations of academic papers [2], and ratings of movies in IMDB [3] are all specific manifestations of information cascade phenomena.

Information cascades can be regarded as an example of collective behavior, and predictive research on its dissemination and development has significant theoretical and practical value [1]. From user-generated content such as scientific papers [2] and Weibo posts [3–5] to online services such as viral marketing [6] and advertisements [7], the flood of digital information in daily life has provided unprecedented opportunities to explore and utilize the trajectories and structures of information cascade dissemination. This helps to reveal the social dynamics and mechanisms of information dissemination within social networks [8–12]. By studying information cascades, we can gain insights into how users

form interest groups, create information dissemination networks, and understand the interactions and influences within groups. At the same time, predicting information cascades can be used to monitor public sentiment on social media, quickly identify potential issues or crises, and take measures for management and intervention to safeguard reputation and public relations. Furthermore, social media platforms can utilize research findings on information cascades to improve personalized recommendation systems [13], providing users with more relevant and engaging content, thereby increasing user satisfaction and platform stickiness. Advertisers [14] can also use information cascade predictions to better target advertising audiences, selecting the best timing and content to improve advertising effectiveness. Therefore, researching models and algorithms for predicting the cascade diffusion of user-generated content in online social networks, and achieving popularity prediction for such content, has significant theoretical and practical value.

However, the complexity of social networks presents significant challenges for predicting information cascades. First, the nodes and edges in social networks are not only large in number but also structurally complex. The connections between users can be strong ties or weak ties, and the influence of these connections on information dissemination varies. Secondly, the heterogeneity of user behavior further exacerbates the unpredictability of information dissemination. Different users exhibit significant differences in their reactions to information and dissemination behavior; some users may actively spread information, while others are merely passive receivers. This heterogeneity makes it difficult to predict the path and speed of information dissemination using simple rules. Moreover, the multimodality of information presents additional challenges for information cascade prediction. Information in modern social networks comes in various forms, including text, images, videos, audio, and other modalities, with varying dissemination effectiveness among different user groups. For example, an attractive short video may spread more quickly on social networks than a lengthy text. How to integrate features of different modalities during the modeling process has become an important research direction in information cascade prediction. Finally, the temporal dynamics of information dissemination are another key challenge. Information cascades are often time-sensitive processes, and the propagation patterns may change significantly at different time points. In the early stages of information dissemination, information may spread mainly within a small range, but as time progresses, the speed of dissemination may accelerate, potentially exhibiting explosive growth trends. Capturing these temporal dynamics in the modeling process to accurately predict dissemination trends is another major difficulty in information cascade prediction.

Recently, the latest applications of deep learning in various fields have led to deep learning-based information cascade prediction. Although the existing literature has extensively discussed information cascade prediction methods [1], there is still a lack of application research based on deep learning models. In particular, the performance of modern deep learning models in real-time processing of large-scale, multimodal data has not yet been fully studied. This paper not only systematizes the current research findings but also fills this research gap through the classification and analysis of modern deep learning methods. To address the limitations of current taxonomies, we propose a new classification method for information cascade prediction. Existing taxonomies often focus on a single dimension, such as prediction targets or methods, which fails to fully capture the complexity of information cascades in modern social networks. These traditional classifications do not account for the growing importance of multimodal data integration, the use of large models, and the dynamic, multi-scale nature of information propagation across diverse platforms. Our new classification method incorporates both prediction targets and methods, offering a more comprehensive framework for understanding and predicting information cascades. Unlike a systematic literature review, which typically follows a fixed methodology to collect, analyze, and synthesize all relevant studies within a specific scope, this comprehensive review aims to cover a broader range of topics in information cascade prediction. It focuses on the application of deep learning, selectively highlighting studies that illustrate new methods or represent emerging trends. This approach allows for a more

flexible exploration of both classical and cutting-edge methods, offering insights into future research directions and practical implementations.

Our contributions: our paper has made significant contributions in the following aspects:

- **New taxonomy:** We propose a new classification method for information cascade prediction. Based on different prediction targets, information cascade prediction can be divided into microscopic information cascade prediction, macroscopic information cascade prediction, and multi-scale information cascade prediction. Additionally, based on different prediction methods, information cascade prediction can be classified into topological information cascade prediction, content-based information cascade prediction, and large model-based information cascade prediction.
- **Comprehensive review:** We provide the most comprehensive overview of information cascade prediction based on modern deep learning techniques.
- **Comprehensive Resource Compilation:** We provide a detailed compilation of state-of-the-art models, benchmark datasets, and practical applications that have been widely used in recent deep learning-based information cascade prediction research. We do not merely list resources, but critically analyze and categorize these assets based on their relevance to specific tasks. This survey can serve as a practical guide for understanding, using, and developing various deep-learning methods for different real-world applications.
- **Future directions:** We discuss the relevant theories of information cascade prediction, analyze the limitations of existing methods, and suggest four potential future research directions: multimodal information fusion, temporal dynamics and real-time prediction, interpretability of large models, and Heterogeneous structures in social networks.

2. Background and Definition

In this section, we outline the background of information cascade prediction and list commonly used notations.

2.1. Background

The problem of information cascade prediction is commonly treated as either a classification or regression task. In classification, the goal is to determine if the cascade size will surpass a specific threshold [15,16] at some point in the future, while in regression, the focus is on estimating the precise future size of the cascade [17–20].

In the domain of information cascade prediction, various prediction techniques show marked differences depending on their design principles and practical application performance. Existing work on the prediction of information cascade propagation can be classified into three groups: feature engineering-based methods, generative model-based methods, and deep learning-based methods.

2.1.1. Feature Engineering-Based Methods

Feature-based information cascade prediction methods involve constructing predictive models by leveraging various features from information cascade events. These features can cover multiple aspects such as event timing, user behavior, and content attributes to enhance the understanding and prediction of the information diffusion process. Suh et al. [21], as pioneers in this area, employed principal component analysis and generalized linear models to investigate factors influencing the likelihood of retweets on Weibo, discovering that content and user features are two significant factors in the pre-event prediction of information cascades. Yu et al. [22] sought to extract fluctuations in popularity directly from the temporal process using manually crafted "phases." However, the influence of external factors might span various ranges and durations. It is challenging to artificially hypothesize the number and shape of these fluctuations. The automatic extraction of short-term fluctuations remains an unresolved issue with this approach. Jenders et al. [23] found

that while content-related features do not outperform structural features in predictive accuracy, they still demonstrate relatively strong predictive capabilities. Cheng et al. [15], through experimentation, discovered that predictive models relying solely on temporal features achieve nearly equivalent performance compared to those augmented with content, user, and structural features, underscoring the dominant role of temporal features in cascade prediction. Carta et al. [24] employed temporal features, user attributes, and the semantic content of post titles, utilizing a gradient-boosting technique to predict the future popularity of posts on the Instagram platform. Feature-based prediction methods offer high interpretability and are easy to implement, making them broadly applicable. However, their ability to capture complex diffusion patterns is limited, and they depend on manually crafted features, which restricts their generalization capabilities.

2.1.2. Generative Model-Based Methods

Generative model-based methods play a crucial role in information cascade prediction. These methods simulate the information dissemination process within a network, allowing for a better understanding and prediction of diffusion behavior. Generative models generally depend on probabilistic or statistical principles, viewing information spread as a random process and predicting future paths and scales based on historical data. The essence of generative models is to capture the underlying mechanisms of information propagation by building a model that can replicate the diffusion process for prediction purposes. Common generative models include probabilistic graphical models such as Markov chains and Bayesian networks, and random point process-based methods such as the Hawkes process. A key feature of generative models is their capacity to adaptively modify prediction outcomes according to varying historical data, granting them a notable advantage in complex and dynamic social network settings.

Methods based on generative models usually consider the growth or popularity of information cascades in online social media as a retweet arrival process, and they model the intensity function of this arrival process for each message separately [25–31]. However, these methods typically do not optimize directly for future popularity, nor do they learn parameters individually for each message. Consequently, they fail to fully leverage the potential dynamic popularity information in information cascades, resulting in a lack of interpretability in the prediction models. Generative models typically depend on extensive historical data for training, which may cause a decline in predictive performance when data is scarce or of poor quality. Furthermore, generative models often struggle to handle multimodal data, which restricts their use in complex social network environments.

2.1.3. Deep Learning-Based Methods

With the rapid advancement of neural networks, an increasing number of methods are leveraging neural networks and deep learning models to learn and infer information dissemination patterns from large-scale social media data [32–37], aiming to predict which information will achieve greater spread and impact within social networks. Numerous studies have indicated that deep neural network models are generally more effective than linear models. For instance, RNN-based models [38] do not rely on specific assumptions regarding cascade diffusion; they can flexibly capture the temporal dependencies within cascades. Models based on graph representation learning [39–42] do not require laborious manual feature crafting from the underlying cascade graph; they are capable of effectively learning node and structural features within the cascade graph.

2.2. Definition

This paper uniformly uses the following notation to define cascades and cascade graphs: given M pieces of information, denoted as $M = \{m^i\}$ ($1 \leq i \leq M$), for each piece of information m^i, the cascade $C^i = \{(u^i_j, v^i_j, t^i_j)\}$ is used to record the diffusion process of m^i. This paper defines a cascade graph as $G = (V, E)$, where V is the set of nodes and $E \in V \times V$ denotes the set of all relationships between users.

3. Categorization

In this section, we present our taxonomy of information cascade prediction, as shown in Figure 1. In the following, we give a brief introduction of each category.
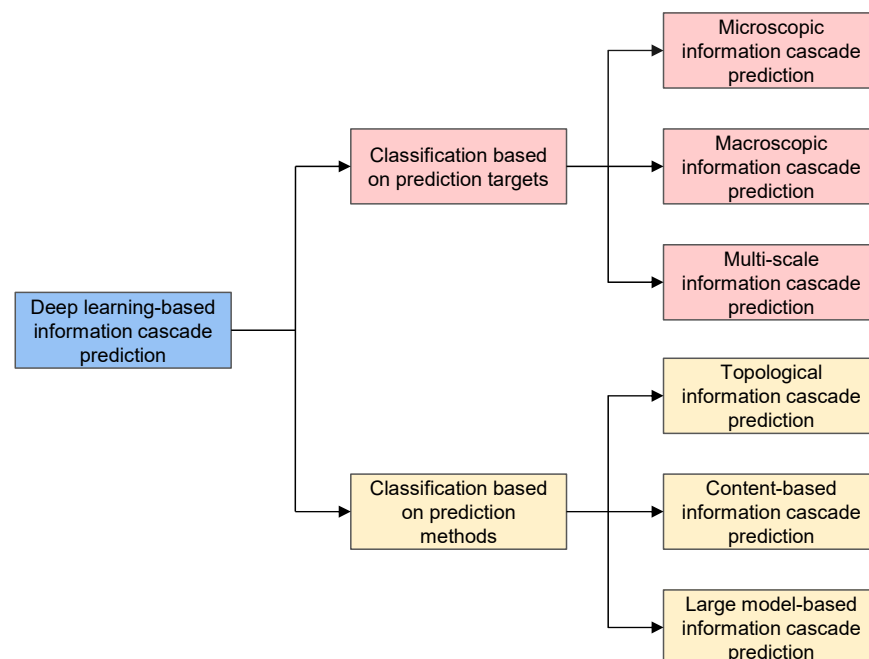

Figure 1. Deep learning-based information cascade prediction method classification.

3.1. Classification Based on Prediction Targets

In the research of information cascade prediction, methods can be divided into microscopic information cascade prediction, macroscopic information cascade prediction, and multi-scale information cascade prediction. Microscopic information cascade prediction primarily focuses on the dissemination behavior of individual nodes or small subgroups, aiming to predict the participation of specific users or nodes in the information dissemination network. Macroscopic information cascade prediction focuses on the dissemination patterns and scale of the entire network, typically used to predict global features such as the overall spread range of information and the final cascade size. Multi-scale information cascade prediction combines both microscopic and macroscopic perspectives, aiming to capture the dynamics of information dissemination at different levels. It considers both local node behaviors and global dissemination structures, striving for a more comprehensive understanding of the complexity and multidimensional characteristics of information cascades.

3.1.1. Microscopic Information Cascade Prediction

Microscopic cascade prediction focuses on predicting information dissemination behavior at the level of individual nodes or within a small scope. The goal is to identify how specific nodes participate in the information diffusion process or predict which node will be affected next [43–51]. This type of method typically uses deep learning techniques to capture complex interactions and dependencies between individual nodes. With the widespread use of social networks, understanding and predicting the role of individual behaviors in information dissemination has become increasingly important, as these behaviors can have a profound impact on the ultimate outcome of information spread. The specific model framework is illustrated in Figure 2.
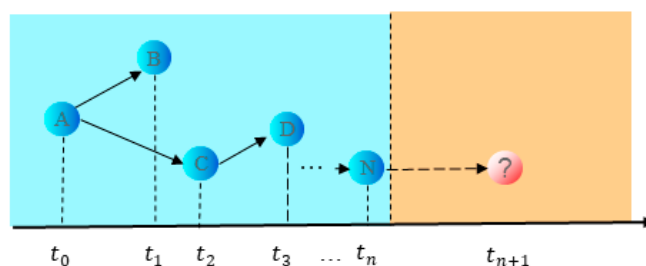
Figure 2. Example of microscopic information cascade prediction.

Microscopic cascade prediction methods usually rely on deep learning models, especially Graph Neural Networks (GNNs) and Recurrent Neural Networks (RNNs), to model complex node relationships and time-series data. Graph Neural Networks (GNNs) are among the most commonly used methods; they can capture the characteristics of nodes and their neighborhoods and predict information dissemination paths through multi-level node representation learning. These models iteratively aggregate the features of nodes and their neighbors, learning information that effectively represents the nodes' positions and roles within the network, thereby predicting their roles in the information dissemination process. Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory Networks (LSTMs) and Gated Recurrent Units (GRUs), are mainly used for handling time-series data and capturing the temporal dynamic characteristics between nodes. For example, in the information dissemination process, the frequency of user interactions, time intervals, and other factors may affect the dissemination path of the information. RNN models can effectively predict information dissemination at specific time points by learning these temporal dynamic relationships.

Zhao et al. [52] proposed a microscopic cascade prediction model combining an LSTM and Graph Convolutional Networks (GCN). This model simultaneously leverages cascade sequence information and social graph structure to capture the temporal dynamics and topological features of information dissemination. It employs an LSTM to model the cascade sequence, representing each node as an embedding vector, which is combined with the previous hidden state and fed into the LSTM. The LSTM regulates the retention of current input content and the incorporation of new memory through its forget gate and input gate. Subsequently, an enhanced GCN processes the social graph structure to aggregate features from the two-step neighbor nodes of each user, resulting in structured information for every node. The initial embeddings of user nodes are initialized using the DeepWalk [53] algorithm, and a random walk method is used to generate user node sequences. Finally, the outputs of LSTM and GCN are combined, and a softmax function is used to compute the multinomial probability distribution of all users to predict the next user to be activated. The GRASS model [54] brings a new perspective to microscopic cascade prediction tasks, primarily addressing the "position jump" and "spatial independence" issues. "Position jump" refers to the ambiguity in the dissemination relationship between infected users in cascade data, while "spatial independence" means that not all infected users affect the prediction of the next infected user. The model is composed of two main components: the GRU-like Attention Unit (GRAU) and Structural Spreading (SS). The GRAU extends the receptive field of the Recurrent Neural Network (RNN) module by introducing an attention mechanism into the Gated Recurrent Unit (GRU) to address the "position skipping" problem. The GRAU considers the current user embedding, the hidden state from the previous step, and all previously infected users to capture the actual first-order infection relationships between non-continuous infected users.

$$r_j = \sigma\left(W_1^r x_{u_j}^{seq} + W_2^r h_{j-1}\right) \tag{1a}$$

$$z_j = \sigma\left(W_1^z x_{u_j}^{seq} + W_2^z h_{j-1}\right) \tag{1b}$$

$$\varphi_j = Sequential\,Attention\left(Q_j^{seq}, K_j^{seq}, V_j^{seq}\right) \tag{1c}$$

$$\widetilde{h}_j = tanh\left(W_1^n \varphi_j + r_j \odot W_2^n h_{j-1}\right) \tag{1d}$$

$$h_j = z_j \odot h_{j-1} + \left(1 - z_j\right) \odot \widetilde{h}_j \tag{1e}$$

where $x_{u_j}^{seq}$ is the sequence feature vector of user u_j, h_j is the hidden state at the current time, and $Q_j^{seq}, K_j^{seq}, V_j^{seq}$ are parameters.

SS, on the other hand, solves the "spatial independence" problem by filtering relevant users through structural features and controlling the generation of cascade hidden states. The SS mechanism infers social closeness between users based on a "many-to-many" co-occurrence relationship and uses these structural features to generate context-aggregated representations. The GRASS model uses the GRU to extract time series features, capturing the dissemination patterns of information at different time points. At the same time, it focuses on key nodes and paths in chained cascade data through an attention mechanism, learning the spatial diffusion features of information. In this way, the GRASS model can not only accurately predict the microscopic diffusion paths of information but also effectively consider the impact of the global network structure on information dissemination. In experiments conducted on large-scale datasets such as Douban, Twitter, and Memetracker, GRASS consistently outperformed state-of-the-art models. Compared to the DeepHawkes [55] and T-Deephawkes [56] models, the GRASS model pays more attention to the spatiotemporal dynamic changes in the dissemination path, considering not only user influence and time decay but also integrating the spatiotemporal features of chained cascade data. It provides a robust solution for handling spatiotemporal dynamics in large-scale cascade data, significantly improving predictive performance over traditional models. The GRASS model demonstrates higher predictive accuracy when addressing the complexity of information dissemination in large networks. Moreover, GRASS reduces over-reliance on a single dissemination path by jointly learning temporal and spatial features, enhancing robustness against random dissemination behaviors.

Microscopic information cascade prediction methods can model and predict at the individual node level, focusing on dissemination paths and temporal features between nodes. This approach is particularly suitable for applications that require an in-depth understanding of node-level dissemination behaviors, such as viral marketing, personalized recommendation systems, and social media sentiment monitoring. However, they often neglect the influence of global network structure and macroscopic diffusion patterns. This limitation may result in the suboptimal performance of the model when dealing with global dissemination processes.

3.1.2. Macroscopic Information Cascade Prediction

Macroscopic information cascade prediction aims to understand and predict the diffusion behavior of information across the entire social network from a global perspective. Its goal is to capture the global dissemination patterns, the final coverage scale, dissemination speed, and the potential influence of information [15,55]. The specific model framework is illustrated in Figure 3. This prediction method is critical for many practical applications, such as social media sentiment monitoring, infectious disease outbreak warnings, marketing campaign optimization, and crisis management. Macroscopic information cascade prediction analyzes the process of information spreading outward from a few seed nodes in the network, helping researchers and decision-makers identify which nodes hold significant influence, predict which nodes may become key paths for information dissemination in the future, and estimate the eventual spread range. Unlike microscopic cascade prediction, which focuses on the node level, macroscopic prediction emphasizes global dissemination patterns, allowing it to reveal the overall characteristics and dynamic changes in the information dissemination process.
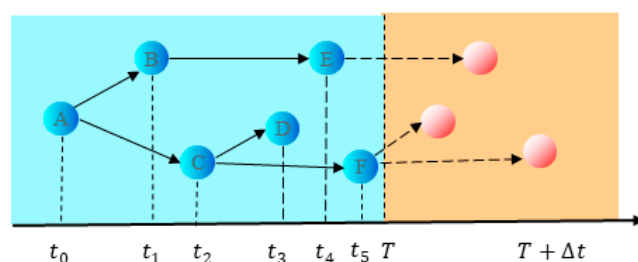
Figure 3. Example of macroscopic information cascade prediction.

Macroscopic information cascade prediction typically examines how information spreads across the entire network, with random walk serving as a stochastic method for simulating information propagation. Conducting random walks within the network allows for capturing the dynamic features of information spread, facilitating the prediction of the diffusion extent and velocity of information at a macroscopic scale. Random walk is a concept in statistics, referring to a stochastic process where, at each step, a particle or point moves in a specific direction with a certain probability. In graph theory, a random walk usually refers to a series of random moves between nodes in a graph, where, at each step, the current node moves to one of its neighboring nodes. The probability of selecting a neighbor can be uniform, meaning that each neighbor is chosen with the same probability, or it can be determined according to certain specific rules. This process is repeated until the sequence length satisfies a predetermined condition. Random walks are extensively applied in information cascade prediction to model and capture the paths and patterns of information diffusion. In the literature [53], sequences of nodes are generated by simulating random walks on a graph. These sequences are akin to sentences in natural language processing. It converts the graph embedding problem into a word embedding problem, employing the Skip-Gram model from Word2Vec to learn the low-dimensional embeddings of the nodes. Building on this, Grover and Leskovec introduced node2vec [57], which incorporates a more flexible random walk strategy. By adjusting the random walk strategy, they control the depth and breadth of node sampling, thus enabling the capture of higher-order or lower-order similarities among nodes. The process of information cascade sampling based on random walks can be summarized as follows. First, a random walk is performed on the cascade graph G, which includes the nodes and forwarding paths in the cascade sequence C, while maintaining the temporal order in C. Assuming that the random walk is currently at node v and is in state N of the Markov chain, the probability of diffusing from v to its neighboring nodes is:

$$\rho(u \in N_c(v)|v) = \frac{sc_i(u) + \alpha}{\sum_{s \in N_c(v)}(sc_t(s) + \alpha)} \tag{2}$$

where $N_C(V)$ denotes the set of neighboring nodes of node V in state N; $sc_t(u)$ is the transition scoring function of node u, generally calculated using the degree of node u in the cascade graph or the weight between u and v; and α is the smoothing parameter. Similarly, in the node set V_c of the cascade graph, the jump probability of node u can be defined using the jump scoring function $sc_j(u)$:

$$\rho(u) = \frac{sc_i(u) + \alpha}{\sum_{s \in V_c}(sc_j(s) + \alpha)} \tag{3}$$

In random walks, the jump and transition probabilities of nodes play a decisive role in determining the sampling outcomes. The probability p_o of deciding whether to make another random jump or to proceed to the terminal state essentially influences the expected number of sampled sequences. Meanwhile, the probability p_j of performing a random jump versus transitioning to neighboring nodes affects the sequence length. Both factors are crucial in shaping the representations of cascade graphs. In traditional random walk sampling, the node transition probabilities are manually set hyperparameters, leading

to the traditional random walk's inability to adapt to the common differences between different information cascade sequences. Accordingly, DeepCase [58] integrates random walk sampling into a deep learning framework to adaptively learn these two probabilities. It uses random walks to perform serialized sampling of the cascade graph, converting the graph structure into a sequence structure that is easier to handle. Subsequently, each random walk sequence is fed into the bidirectional gated recurrent unit (GRU) of a recurrent neural network to obtain the node embeddings. Finally, using an attention mechanism, different weights are assigned to each node and each walk sequence, followed by weighted aggregation to generate a representation of the information diffusion cascade graph, which is then used to predict information popularity. Based on graph neural networks and random walk sampling, the literature introduces a Hierarchical Attention Neural Network (HANN) [59] designed for information cascade prediction. The model aims to enhance the accuracy of information cascade prediction by learning both the network structure and temporal features. Specifically, the HANN generates node sequences using random walks and then applies a Hierarchical Attention Mechanism to focus on the key nodes and paths within these sequences. In the HANN, random walks are initially employed to explore node relationship paths within the graph, capturing both local and global structural features of the network. The HANN then introduces a dual-level attention mechanism, comprising node-level attention and path-level attention. The node-level attention mechanism identifies key nodes generated during the random walks, while the path-level attention mechanism focuses on critical propagation paths, thereby enriching the semantic information of the node representations. To address the temporal features of information cascades, the HANN incorporates a recurrent neural network (such as an LSTM) to capture the temporal dependencies during the information diffusion process. This integration allows the model to capture both the static structural features of the network and the dynamic temporal changes in information propagation. Moreover, the HANN utilizes an end-to-end joint optimization strategy during model training, enabling the simultaneous learning of network structure and temporal features, which enhances overall model performance and prediction accuracy. Through this innovative design, the HANN shows significant advantages in dealing with heterogeneous networks and capturing complex information diffusion patterns, giving it broad application potential in social network analysis and information cascade prediction.

The random walk method generates a series of node sequences by simulating the random spread of information within a network, but it may overlook certain key nodes or paths, making it difficult to fully reconstruct the true propagation route. Methods based on diffusion paths seek to utilize information propagation paths to capture and comprehend the dynamic features of information diffusion within a network. These methods analyze the propagation paths of information within the network to predict the scale and influence of information cascades. Cao et al. [55] attempted to integrate random point process models with deep learning frameworks, and they innovatively introduced the DeepHawkes model. The model captures three key aspects of the Hawkes process: user influence, self-excitation mechanisms, and time decay effects. Like DeepCas, this model also learns the representation and prediction of cascades in an end-to-end manner. The key to the Hawkes process is modeling the probability of new forwarding behavior, which can be expressed as:

$$\rho_t^i = \sum_{j:t_j^i \leq t} \mu_j^i \phi\left(t - t_j^i\right) \tag{4}$$

where ρ is the probability of a new forward of message m^i at time t, t_j^i is the time elapsed between the initial forward and the j-th forward, μ_j^i is the number of potential users directly influenced by the j-th forward, and $\phi(t)$ is the time decay function.

The DeepHawkes model combines the interpretability of the Hawkes process with the predictive strength of deep learning; however, it overlooks the impact of textual content on information diffusion. To compensate for this limitation, Wang et al. [56] incorporated a

topic classification model into the DeepHawkes framework, proposing the T-DeepHawkes model, which further accounts for the impact of textual content on information diffusion. The T-Deephawkes model employs User Embedding, Retweeting Path Encoding, and Time Decay as three components to simulate the interpretable factors in the Hawkes process. The model utilizes the LDA topic classification model to extract themes from message text content, integrating the cascade path representation and topic path representation via a pooling layer, which results in a comprehensive representation of the influence on information propagation, offering a more holistic modeling of the diffusion process. Yu et al. [60] connected the hierarchical attention mechanism with the Hawkes process to continuously model the flow of discrete events, thereby improving the modeling capacity and prediction accuracy for complex event sequences.

Macroscopic information cascade prediction often assumes that network nodes and connections are homogeneous, neglecting the heterogeneity among nodes, which may affect the representation of actual propagation processes. Furthermore, macroscopic predictions primarily focus on global trends, lacking multiscale analysis and failing to integrate important information from the micro and meso levels. Moreover, due to reliance on historical data, macroscopic models may struggle to adapt to rapidly changing propagation environments. Therefore, while macroscopic information cascade prediction offers a perspective on overall propagation trends, improvements are still needed in terms of details, heterogeneity, and dynamic adaptability.

3.1.3. Multi-Scale Information Cascade Prediction

Multi-scale prediction methods incorporate different scales of information dissemination into a single model, aiming to predict both individual node-level dissemination behaviors and network-wide information diffusion characteristics. This approach is particularly suitable for scenarios that require capturing the complex hierarchical structures and temporal dynamics of information dissemination. For example, on social media platforms, the dissemination of a specific event may go through multiple stages, from initial dissemination by a small number of users to widespread diffusion across the entire network. Multi-scale prediction models can more accurately capture and predict these multi-phase dissemination behaviors, thus enhancing the understanding of the overall diffusion process. In many application scenarios, single-scale prediction methods often fail to fully capture the complexity of information dissemination. For example, microscopic prediction methods, while capable of accurately predicting the dissemination behavior of individual nodes, often neglect the global network structure and macroscopic diffusion patterns of information. On the other hand, macroscopic prediction methods, while capable of analyzing the diffusion path of information across the entire network, often lack a detailed understanding of individual node behaviors. Multi-scale prediction methods address the shortcomings of single-scale methods by considering both scales simultaneously.

Yang et al. [61] proposed the Full-Scale Information Diffusion Prediction Model, which combines Reinforcement Learning (RL) and Recurrent Neural Networks (RNNs) to achieve multi-scale prediction of information diffusion. The model was evaluated on datasets such as Weibo and Twitter, showing an improvement of 15% in accuracy compared to traditional methods for predicting cascade growth, particularly in rapidly evolving information cascades. The core idea is to use reinforcement learning to coordinate microscopic and macroscopic cascade predictions, thereby enhancing the overall understanding of the information diffusion process. This model employs an RNN-based microscopic cascade prediction module to capture the fine-grained features of information dissemination. The module's goal is to predict the next node likely to be infected and the time of its dissemination. The model employs a structured context extraction algorithm, which is based on neighborhood sampling and effectively leverages underlying social graph information to enhance the model's predictive capabilities. The specific formula is:

$$f_v^{(1)} = \text{relu}\left(W \cdot \frac{1}{Z} \sum_{k=1}^{Z} f_{u_k}^{(0)} + b \right) \tag{5}$$

where u_k is the node uniformly sampled from user v and its neighboring nodes, and W, b are the weight matrix and bias vector.

To integrate macroscopic cascade information, the model introduces a reinforcement learning framework. The macroscopic cascade prediction task is regarded as a decision-making process, where the agent observes the historical state of microscopic cascades and takes actions to predict the overall scale and trend of information diffusion. The reward function R in the reinforcement learning framework is defined as the difference between the predicted result and the actual diffusion scale, and the model optimizes its performance by maximizing cumulative rewards. The model integrates microscopic and macroscopic cascade prediction modules into a unified optimization framework. In reinforcement learning, the state space S represents the current cascade state, while the action A corresponds to the model's prediction output. To predict the final scale of the cascade, the model performs cascade simulation based on the observed first K users. During the simulation process, the model recursively selects the next user based on the predicted probability distribution until a special termination signal ("<STOP>") is predicted, thereby estimating the final cascade size. In this way, the model can adaptively adjust its prediction strategies at both microscopic and macroscopic levels, gradually approaching optimal prediction. The model seamlessly integrates reinforcement learning with cascade prediction, creating a more adaptive and flexible multi-scale prediction framework. It significantly improves prediction accuracy at both micro and macro levels, making it highly valuable in applications such as social media monitoring and public opinion analysis. Furthermore, the model's ability to adapt to real-time data streams marks a significant advancement in the field of information diffusion prediction. On this basis, the multi-scale information diffusion prediction model (MSIDP) [62] proposed by Xu et al. further improves multi-scale information diffusion prediction methods by incorporating timestamp information and wide dispersion characteristics. The MSIDP model aims to dynamically integrate temporal and spatial features to capture the multi-scale characteristics in the information diffusion process. The MSIDP model introduces specific diffusion timestamp information and adjusts the temporal dynamics of model predictions through a time control module. MSIDP adopts Bidirectional Graph Convolutional Networks (BGCNs) to simultaneously learn the deep propagation and wide dispersion features in the information dissemination process. By performing convolution operations in both directions of the graph, the BGCN can capture the complex interactions between nodes and their neighborhoods in the information dissemination process. To better combine microscopic and macroscopic predictions, MSIDP performs information fusion at multiple scales. By using Multilayer Perceptrons (MLPs) and pooling layers, microscopic cascade features and macroscopic cascade features are combined to form a unified multi-scale representation.

In practical applications, multi-scale prediction models that combine reinforcement learning and timestamp information can be used in more complex scenarios, such as public opinion monitoring on social media, virus spread prediction, and risk assessment in financial markets. These models demonstrate strong capabilities in capturing the fine-grained and coarse-grained features of information dissemination. However, multi-scale information cascade prediction also faces challenges, such as high computational complexity, issues with model interpretability, and adaptability in large-scale dynamic networks. Future research could focus on developing more efficient algorithms to address these challenges and explore how to better integrate multimodal data to improve prediction accuracy and generalization capabilities.

3.2. Classification Based on Prediction Methods

Information cascade prediction methods, based on different prediction approaches, can be classified into topological information cascade prediction, content-based information

cascade prediction, and large model-based information cascade prediction. Topological information cascade prediction focuses on utilizing the structural features of the dissemination network [63], inferring the likelihood of information dissemination by analyzing the connections between nodes, community structures in the network, or dissemination paths. Content-based information cascade prediction focuses on the semantic features of the information itself, using content such as text, images, or other forms to assess the appeal of information in the network and predict its dissemination trend [64]. Large model-based information cascade prediction employs deep learning or pre-trained large-scale models, capturing complex spatiotemporal dependencies and multi-level features to enhance prediction accuracy, and is widely applied in more complex cascade scenarios.

3.2.1. Topological Information Cascade Prediction

Information propagation in social networks is generally a complex process, which is closely linked not only to the content of the information and user behavior patterns but also to the topological structure of the network. Topological structure refers to the connections between nodes (users) in the network, usually represented as a graph. Given the complexity and heterogeneity of social networks, topological information is essential in predicting the process of information dissemination.

In information cascade prediction, topological information can assist in understanding how information spreads through the network and identify which nodes play crucial roles in information diffusion. Thus, topology-based prediction methods emphasize nodes with high dissemination potential, community detection, and the path dependency characteristics of information spread. Such methods are effective in capturing how information spreads along network connections, especially in the context of large, complex social networks. Graph Neural Networks (GNNs) have become a crucial technology that has been applied to topological cascade prediction in recent years. GNNs capture local and global network features by transmitting information between nodes, aiding researchers in better understanding the dynamics of information propagation. For instance, GNN-based cascade prediction models can model the influence of individual nodes by aggregating the features of both the nodes and their neighboring nodes, enabling the prediction of whether a particular node will engage in the subsequent diffusion of information. A recent advancement in this area is the introduction of Temporal Graph Neural Networks (TGNNs), which extend traditional GNNs by incorporating time-dependent data. TGNNs allow the model to capture not only the structural dynamics of the network but also the temporal evolution of node interactions, making them particularly suited for dynamic, real-time social networks. TGNNs update node and edge representations over time, providing more accurate predictions of when and where information will spread across the network. This added temporal dimension significantly improves the model's ability to predict information cascades, especially in scenarios requiring real-time analysis, such as crisis management and viral content dissemination.

This approach of integrating topological information with temporal data allows the model to excel in handling complex information propagation tasks. The specific model framework is illustrated in Figure 4. Recent research has expanded this framework further. For instance, Graph Neural Networks (GNNs) incorporating self-attention mechanisms can dynamically adjust the weights between nodes, thus more accurately capturing information propagation patterns in heterogeneous networks. Moreover, research combining Graph Neural Networks with sequence models has shown even stronger predictive performance, especially in capturing long-term dependencies and complex propagation pathways. These recent advances offer substantial support for the further development of the model and pave new ways for applying information cascade prediction.
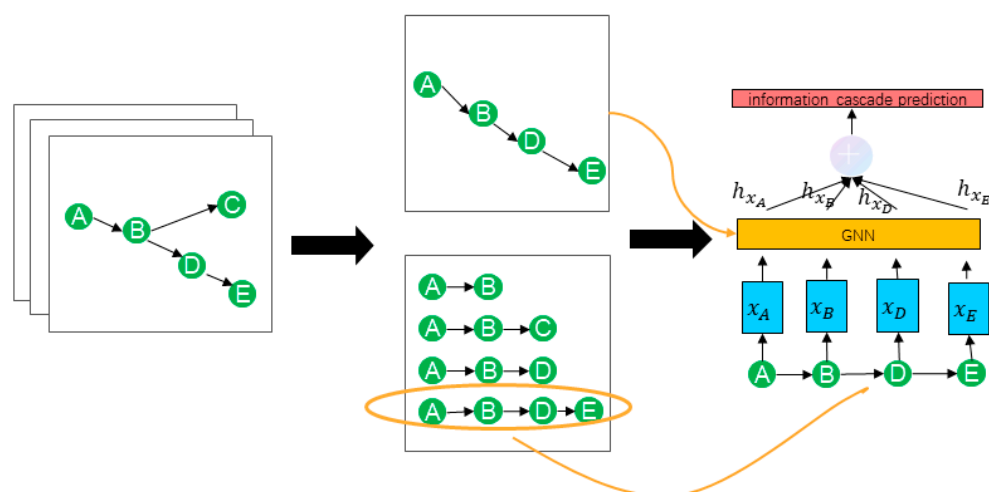
Figure 4. Example of topological information cascade prediction.

DEEPCASE uses random walks to perform serialized sampling of cascade graphs, converting the graph structure into a more easily processed sequence structure. This approach, however, unavoidably loses some of the graph's structural information. To compensate for this limitation, Wang et al. [49] retained the corresponding edges (u, v) in the user subgraph, where $t_u < t_v$ according to the sequence of user participation in the cascade, thus converting the participating user subgraph into a Directed Acyclic Graph (DAG). Furthermore, the study designed two distinct "user role" embedding vectors for each node: sender and receiver embedding vectors. The sender embedding allows the model to encode both the static tendencies of the nodes and the dynamic context of the information diffusion topology. In other words, this embedding captures not only the inherent characteristics of the node but also its specific context within the diffusion process and the network of activated nodes. Unlike the sender embedding, the receiver embedding primarily encodes the static attributes of the node, without accounting for the dynamic context of diffusion. This design choice is due to the fact that receiver nodes have not yet participated in the cascade, so their embeddings rely more on the inherent properties of the nodes. The model attempts to link active nodes with the sender embedding and inactive nodes with the receiver embedding. The paper [49] introduces the Topological Recurrent Neural Network (Topo-LSTM) model to effectively capture the structural characteristics of such DAGs, where each node in the subgraph is influenced solely by itself and the nodes directed toward it. The model takes the dynamic DAG as input and outputs topology-aware embeddings for each node in the DAG. Topo-LSTM utilizes the sender embedding of node v_t to ascertain which other nodes have been activated so far and how the topology has diffused to that node. It predicts the next diffusion node by learning the receiver embedding of each inactive node and determining its proximity to the sender embedding of the active nodes.

Approaches based on global topological structures emphasize the entire network's architecture and are well-suited for identifying critical nodes and paths across the network. However, they entail high computational complexity, particularly when managing large-scale dynamic networks. Neighborhood aggregation methods, on the other hand, primarily concentrate on the local features of nodes within the network. This approach generates node embeddings by aggregating information from each node's neighbors and is often used in Graph Neural Networks. Neighborhood aggregation techniques can dynamically capture changes in the local structure of nodes and adapt to more intricate local patterns. In the domain of information cascade prediction, frequently used Graph Neural Network frameworks include GraphSAGE [65], Graph Convolutional Network (GCN) [66], and Graph Attention Network (GAT) [67], among others.

The literature [40] introduced a model named CasCN, which links graph theory with deep neural networks by directly sampling data and learning features from subgraph structures. Classical GCNs are unsuitable for modeling information cascades because they are designed for undirected graphs, which do not account for the temporal dynamics inherent in cascade evolution. To tackle this issue, CasCN utilizes Diplacian [68] to resolve the problem of failing to capture the feature differences in random walks across different cascades. After acquiring the adjacency matrix representation of the sub-cascade graph sequence A_i^T and the Laplacian matrix Δ_c for each cascade, CasCN proceeds to capture both structural and temporal patterns by integrating a classical LSTM with the GCN. The model updates the memory cell by replacing the existing storage cell with a new cell c_t as shown below:

$$i_t = \sigma(W_i * \mathcal{G}X_t + U_i * \mathcal{G}h_{t-1} + V_i \odot c_{t-1} + b_i) \tag{6a}$$

$$f_t = \sigma\left(W_f * \mathcal{G}X_t + U_f * \mathcal{G}h_{t-1} + V_i \odot c_{t-1} + b_f\right) \tag{6b}$$

$$o_t = \sigma(W_o * \mathcal{G}X_t + U_o * \mathcal{G}h_{t-1} + V_o \odot c_{t-1} + o) \tag{6c}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c * \mathcal{G}X_t + U_c * \mathcal{G}h_{t-1} + b_c) \tag{6d}$$

where $*\mathcal{G}$ represents the graph convolution operation, $\sigma(\cdot)$ is the logistic sigmoid function, i_t, f_t, o_t, and b_* are, respectively, the input gate, forget gate, output gate, and bias vector, and W, U, and V are the weight matrices.

Previous research has predominantly concentrated on the structure or sequence of cascades, often neglecting the underlying social structures that, though not directly visible within the cascade itself, have a substantial impact on user behavior [49]. This omission makes predicting inactive users particularly challenging. To tackle this issue, FOREST [60] integrates GRU and GCN to simultaneously learn both cascade context and social networks. Since complete network structures are not always accessible, InfVAE [69] uses a variational autoencoder to model homophily, embedding unobserved social connections. DyHGCN [70] advances this approach by employing heterogeneous graphs to capture both user interactions and social relationships. While these approaches help account for temporal effects in cascades and social homophily, they generally fail to capture the dynamic nature of connections between users and cascades, which limits predictive accuracy. Increasingly, researchers have acknowledged that simply relying on traditional graph structures is insufficient for modeling more complex user interaction dynamics. To address this, Feng et al. [71] proposed a hypergraph neural network based on the Chebyshev expansion of the graph Laplacian. Following this, Bai et al. [72] introduced attention mechanisms into hypergraphs. Unlike standard graphs, hypergraphs allow hyperedges that connect multiple entities, making them particularly effective for modeling group interactions in real-world scenarios. Hypergraphs are especially valuable for representing multiple interactions between users and various cascades. In information diffusion tasks, hypergraphs excel at capturing how users influence different information sources and how these sources, in turn, collectively impact users. In 2022, Sun et al. [73] introduced the MS-HGAT model. Unlike traditional Graph Neural Networks (GNNs), MS-HGAT utilizes a hypergraph structure to model the complex interactions among users and incorporates a memory module to better capture users' historical behavior. The paper constructs sequential hypergraphs based on prior cascades and introduces a sequential hypergraph attention network to dynamically learn user interactions and the relationships between cascades at the cascade level. During each time interval, a Hypergraph Attention Network (HGAT) is employed to model the correlations among users. Given a hypergraph G_D^t, the first step of the HGAT is to learn the representation $o_{j,t}$ of the hyperedge e_j^t by aggregating the initial user representations $x_{j,t}$ of all connected nodes u_i^t, as follows:

$$o_{j,t}^{l+1} = \sigma \left(\sum_{u_i^t \in e_j^t} \alpha_{ij}^t W_1 x_{i,t}^l \right) \tag{7}$$

Since the root node can partially represent the content of the cascade, the paper retains the root information of all hyperedges for each hypergraph and calculates the attention scores of other nodes by measuring their distance from the root, represented as:

$$\alpha_{ij}^t = \frac{\exp\left(-dis\left(W_1 x_{i,t}^l, W_1 r_j^l\right)\right)}{\sum_{u_p^t \in e_j^t} \exp\left(-dis\left(W_1 x_{p,t}^l, W_1 r_j^l\right)\right)} \tag{8}$$

where r_j^l is the representation of root user of e_j^t at layer l, and $dis(\cdot)$ refers to the Euclidean distance.

Once the hyperedge representations are obtained, each hyperedge is assigned the same weight during aggregation.

$$x_{i,t}^{l+1} = \sigma \left(\sum_{e_j^t \in \varepsilon_i^t} W_2 o_{j,t}^{l+1} \right) \tag{9}$$

To further capture the connections between hyperedges, the hyperedge representations are updated using the embedding vectors of all users, which can be represented as:

$$o_{j,t}^{l+1\prime} = \sigma \left(\sum_{u_i^t \in e_j^t} \alpha_{ij}^t W_3 x_{i,t}^{l+1} \right) \tag{10}$$

Experimental results on four public datasets, including Twitter and Douban, indicate that MS-HGAT outperforms state-of-the-art models such as DyHGCN. It introduces a novel method for dynamically modeling user dependencies and preferences, capturing both short-term and long-term impacts on user behavior. Secondly, by leveraging memory-enhanced embedding lookups and attention mechanisms, the model excels in scenarios requiring the integration of user historical behavior and cascade context for more accurate diffusion prediction. MCDAN [74], building on MS-HGAT, introduces multi-scale diffusion hypergraphs that capture interactions at various time scales, enhancing the model's capability to learn both short-term and long-term diffusion patterns.

3.2.2. Content-Based Information Cascade Prediction

Prior research has shown that the content of information significantly influences the ability to forecast the future popularity of messages. User-generated content in online social networks comes in various modalities, including text [44,47,54,59], images [8,9], videos, and others. These different modalities often carry highly complex semantics. Content-based information cascade prediction involves analyzing the content features of the disseminated information to forecast its diffusion path and impact. This approach is distinct from methods that rely solely on network structure or temporal features, as it focuses on extracting information such as content semantics, sentiment, and themes to predict the scale and pattern of information diffusion. The specific model framework is illustrated in Figure 5.
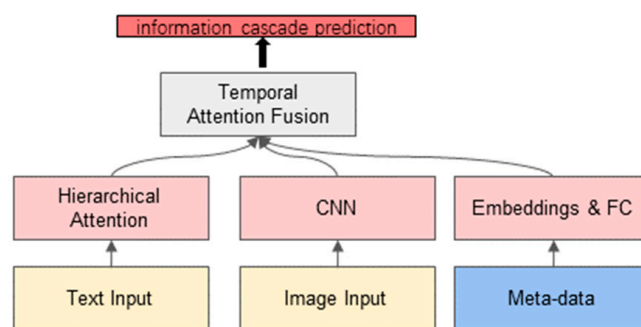
Figure 5. Example of content-based information cascade prediction.

For text-based disseminated information, researchers typically adopt common methods from natural language processing, representing each word in the text as a K-dimensional vector. This vector, usually known as a word embedding, allows semantically similar words to be mapped to adjacent positions in the vector space. Zhang et al. [75] employed an LSTM to characterize the sequential nature of text, feeding word vectors into the LSTM one by one according to their order in the text sequence, with the hidden vector of the last word representing the semantics of the entire sequence. Meanwhile, Wang et al. [76] utilized a one-dimensional convolutional neural network (CNN) to capture local sequence features in short text sequences by extracting phrase features with specific convolution kernel sizes using convolution operators and obtaining fixed-length global text semantic representations through pooling operations. Liao et al. [32] combined modeling of temporal processes with content features to predict the popularity of online articles at various stages of their lifecycle. They employed recurrent neural networks (RNNs) to model the temporal process and capture long-term trends in popularity growth. For short-term fluctuations, they used convolutional neural networks (CNNs) to automatically extract features indicating upward or downward trends. For content feature modeling, they employed a hierarchical attention network (HAN) [77] to capture text features, encoding the text at both the word and sentence levels to generate highly representative text vectors. Using this attention mechanism, they performed hierarchical weighted aggregation of words, sentences, and other granularities in the text, effectively capturing the structure and semantic characteristics of long texts. They also employed embedding techniques to integrate metadata features into corresponding dense vectors.

Previous approaches often assumed that the diffusion behavior and patterns of all information items were uniform, an assumption that may not hold in the real world. Intuitively, users generally have a variety of interests, and their dissemination behavior can vary significantly depending on the topic of the information item. For instance, users might follow and retweet different individuals depending on the topic, resulting in topic-specific dependencies. In models such as T-BERTSum [78], by combining BERT's pre-trained text embeddings with topic information, the model can generate topic-consistent and precise text summaries, providing a more accurate foundation for predicting information dissemination. In the study of Topic-Aware Information Coverage Maximization [79], by taking into account the interests of nodes in different topics within social networks, the model can better select seed nodes to maximize influence, thus optimizing the spread of information.

Content-based cascade prediction methods are particularly effective in addressing cold-start issues and predicting the spread of new information. These methods can directly extract features from the content, making them well-suited for content-rich social networks with diverse user behaviors. However, content-based cascade prediction methods also encounter several challenges. Firstly, content is not the sole factor influencing information dissemination; the structure of social networks, relationships between users, and temporal factors also play a role in information diffusion. Secondly, handling complex multimodal content demands substantial computational resources, and how to efficiently process this

content while incorporating user behavior and social network structure features remains a key research challenge.

3.2.3. Large Model-Based Information Cascade Prediction

With the advancement of deep learning and large-scale pre-trained models, large model-based information cascade prediction methods have shown remarkable advantages in processing complex social network data. Unlike traditional content-based prediction methods, large models are capable of processing the information content itself, as well as incorporating various contextual information, offering more comprehensive prediction results through the deep integration of multimodal data.

Transformers, originally designed for natural language processing, have demonstrated exceptional performance in time-series data processing due to their Multi-head Self-attention Mechanism. This architecture allows the model to capture both long-term dependencies and intricate temporal relationships between data points, making it highly suitable for predicting information cascades in real-time social networks. By dynamically modeling the evolution of information dissemination, Transformers can accurately identify critical moments and shifts in cascade behavior, providing more precise and timely predictions compared to traditional methods.

Moreover, large pre-trained models such as Transformers are particularly effective in handling multimodal data, allowing them to integrate textual, visual, and structural information from social networks. This enables more holistic predictions by considering not only the content of the information but also the timing and patterns of user interactions. The flexibility of Transformers in processing large-scale data streams makes them ideal for predicting fast-evolving cascades, such as viral content or rapidly spreading news across multiple platforms.

Large models, pre-trained on vast datasets, exhibit powerful semantic understanding and contextual capture abilities. Unlike content-based prediction methods, large model-based approaches are not restricted to processing single text or image data; they can integrate multimodal data, thereby improving prediction accuracy and adaptability. The CLIP model learns features from both images and text concurrently, mapping these into a shared embedding space, allowing the model to perform visual classification based on natural language descriptions. This capability allows the model to simultaneously process visual and linguistic information in social networks and predict the dissemination effects of multimodal content.

Large models have demonstrated remarkable advantages in processing time-series data and modeling user behavior. The TIME-LLM [80] model showcases how large language models can be reprogrammed to adapt to time-series prediction. This method can be applied to information cascade prediction by converting the time-series data of information dissemination into an input format that large models can handle, and then using pre-trained large models to predict future propagation paths and scale. Compared to traditional content-based prediction methods, large models are better equipped to capture the temporal dynamics of information dissemination. For instance, GPT models can not only generate text related to information dissemination but also simulate user interaction behaviors, predicting how these actions influence the diffusion path of information. CARE [81] (CAscade-REtrieved In-Context Learning) is a novel framework based on GPT-type large language models. CARE utilizes cascade retrieval prompts and in-context learning methods by creating a prompt pool from historical cascade data, combined with a social relationship-enhanced embedding module and prompt enhancement strategy, to enhance the accuracy of information diffusion prediction. The CARE model is inspired by prompt engineering in large models, utilizing patterns from historical cascade data to provide contextual support for current queries, thus improving the model's predictive performance. By incorporating temporal information and user behavior, large models can more comprehensively predict the patterns of information dissemination within social networks.

Compared to content-based prediction methods, approaches based on large models can capture key nodes and paths in the information dissemination process by modeling the network's topological structure. For instance, BERT can be utilized to extract text embeddings for nodes, while GNNs can analyze the connections between these nodes in the network. Wang et al. [82] introduced a topic-aware neural network model (TAN) to incorporate the topical context and diffusion history context into user representations for predictions. The specific model framework is illustrated in Figure 6.
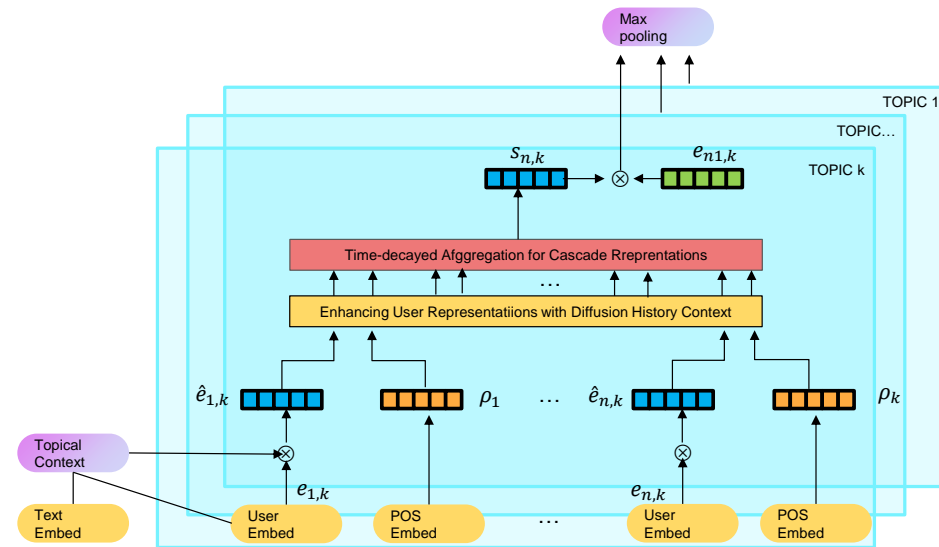

Figure 6. Example of TAN.

The paper employs pre-trained language models to encode the semantic information of diffusion texts. Additionally, the diffusion history context can be further divided into user dependency and position dependency. Inspired by multi-head attention [25], TAN views each topic as a specific head and independently executes attention mechanisms within each topic to extract user and position dependencies. Weights are assigned based on the cosine similarity between text and user embeddings to enhance the topic-related user embeddings, with the normalized cosine similarity function calculated as:

$$z_{j,k}^i = \frac{\exp\left(< y_i, e_{j,k}^i >\right)}{\sum_{l=0}^{K} \exp\left(< y_i, e_{j,l}^i >\right)} \tag{11}$$

where $k = 1, 2 \cdots K$ and $z_{j,k}^i$ is the weight for user u_j^i in the k-th topic, $e_{j,k}^i$ is the user embedding in the k-th topic, and y_i is the embedding of diffusion text. After acquiring the topic context representation, the paper employs an attention mechanism to capture the dependency between the target user and previously infected users, considering positional information. Furthermore, the time decay effect is integrated into the cascade representation, adjusting the user's influence weight via a non-parametric time decay model. This method enables the TAN to more precisely capture multi-topic dependencies, thus predicting the behavior and trends of information dissemination under different topics.

The strengths of large models in information cascade prediction are primarily reflected in their robust feature learning abilities and capacity to process complex data. By leveraging large-scale pre-training, these models can learn propagation patterns from abundant data without depending on traditional manual feature extraction. Compared to content-based prediction methods, large models excel at integrating multimodal data (such as text, images, and network structures), thus achieving a more comprehensive understanding of the information dissemination process. However, large models also face several challenges. For example, they typically require extensive computational resources and training data,

and their performance may lag behind traditional methods when handling long-tail distributions or small datasets. Moreover, improving the efficiency of large models in real-time prediction tasks remains an important focus for future research.

4. Datasets and Metrics

To evaluate the performance of information cascade prediction models, it is crucial to select appropriate datasets and metrics that reflect the dynamics and multimodal characteristics of social networks. In this section, we first introduce the key datasets commonly used in this field then outline the evaluation metrics for assessing model performance in various cascade prediction tasks.

The criteria for selecting datasets in this paper include their wide usage, representativeness, and diversity, ensuring coverage of different types of social networks. For example, Twitter and Weibo datasets not only have sufficient scale but also contain rich user interaction data, making them suitable for analyzing information cascade propagation. The Digg dataset also shares these characteristics, as it not only records users' "likes" but also contains precise time-series information and social network relationships between users, making it particularly suitable for studying the dynamic process of information dissemination and the role of social network influence in cascade propagation. We mainly sort datasets into two groups, namely citation networks and social networks. In Table 1, we summarize selected benchmark datasets.

However, these datasets can introduce biases due to the imbalanced representation of certain user groups, content types, or interaction patterns. To address dataset bias, several strategies can be applied:

- **Data balancing techniques**: Over-sampling underrepresented classes or under-sampling overrepresented classes helps mitigate bias from imbalanced data. Techniques such as SMOTE (Synthetic Minority Over-sampling Technique) or random under-sampling can help balance the dataset, making the model more robust to imbalances.
- **Adjusting loss functions**: In cases where data imbalances are prevalent, using weighted loss functions or focal loss can help the model focus on harder-to-predict samples, thereby reducing the bias that might arise from an overrepresentation of certain classes.
- **Data augmentation and cross-domain learning**: Augmenting data by synthetically generating new samples or incorporating data from other domains (e.g., cross-platform learning between Twitter and Weibo) can help improve model generalization and reduce bias from dataset limitations.

The evaluation of cascade prediction models requires metrics that can accurately capture the model's ability to forecast both the size and spread of information cascades.

Accuracy measures the overall correctness of predictions. It is a basic but important metric, particularly in predicting whether an information cascade will occur within a given threshold. However, accuracy alone is insufficient for imbalanced datasets, where the number of non-cascading events may vastly outnumber actual cascades. Precision evaluates the proportion of correctly predicted cascades out of all predictions, while recall assesses the proportion of actual cascades that were correctly predicted by the model. These metrics are particularly important when predicting the occurrence of rare, large-scale cascades, where false positives and false negatives carry different weights. The F1 score provides a balance between precision and recall, offering a single metric that reflects the model's robustness in handling imbalanced datasets. It is especially useful in scenarios where the prediction of small and large cascades needs to be treated equally. For regression tasks, Mean Squared Error (MSE) and its variants are the most commonly used metrics, while k-top coverage is also frequently used in certain specific scenarios. The choice of metrics is often subjective, even under well-defined problem statements. Previous research has found that a model may perform well on one metric but significantly worse on another, making it difficult to conduct fair comparisons across various methods. We summarize frequently used metrics as well as their adopters in Table 2.

Table 1. Summary of selected benchmark datasets.

Category	Dataset	Source	Nodes	Edges
Social Networks	Sina weibo	[55]	10,077	11,956
	Twitter	[83]	137,093	3,589,811
	BlogCatalog	[84]	10,312	333,983
	Flickr	[84]	80,513	5,899,882
Citation Networks	Digg	[85]	279,632	2,617,993
	Memetracker	[86]	5000	313,669
	HEP-PH	[87]	34,546	421,578
	APS	[29]	13,945	15,508

Table 2. Summary of selected metrics.

Metric	Formulation	Reference
Accuracy	-	[88–99]
Accuracy with tolerance τ	$\left(\lvert \frac{\hat{P}_i - P_i}{P_i} \rvert \leq \tau \right)$	[29,39,95,100–104]
Coefficient of Determination	-	[97,105–109]
Coefficient of Correlation	-	[4,92,109–115]
F1 Score	$\frac{2 Precision \cdot Recall}{Precision + Recall}$	[15,43,96–98,116–121]
Mean Absolute Error	$\frac{1}{M} \sum_i^M \lvert \hat{P}_i - P_i \rvert$	[3,97,112,122]
Mean Abs. Percent. Error	$\frac{1}{M} \sum_i^M \lvert \frac{\hat{P}_i - P_i}{P_i} \rvert$	[39,100–102,123–126]
Mean Square Error	$\frac{1}{M} \sum_i^M \left(\hat{P}_i - P_i \right)^2$	[119,124,127,128]
Precision	$\frac{TP}{TP+FP}$	[43,46,93,97,98,116–118,129–134]
Recall	$\frac{TP}{TP+FN}$	[135–140]
Root Mean Square Error	$\sqrt{\frac{1}{M} \sum_i^M \left(\hat{P}_i - P_i \right)^2}$	[109,124,126,141–144]

5. Applications

Information cascade prediction has wide-ranging applications across various fields, each leveraging different aspects of information diffusion to solve real-world problems. In this section, we detail some of the practical applications of information cascade prediction in public health, crisis management, and marketing. In these real-world applications, specific methods are applied to enhance the robustness of predictions and mitigate potential biases. Techniques such as cross-validation, data augmentation, and bias correction play key roles in ensuring models perform accurately across diverse datasets and dynamic scenarios.

Public Health: The ability to predict information dissemination, especially during a health crisis, is crucial for public health planning and intervention. During epidemics such as COVID-19, understanding how information about the disease spreads within communities is essential for managing public responses and organizing resources. Information cascade prediction models can be used to predict how public health information spreads on social media platforms.

Information cascade prediction models can be used to predict how public health information spreads on social media platforms. The process begins by collecting large-scale social media data, such as posts, tweets, and comments related to the crisis. The model is trained using historical data on how similar health information spread in past outbreaks. To ensure accuracy across different populations and platforms, cross-validation is used, where the model is repeatedly trained and tested on different subsets of the data. This helps reduce the risk of overfitting to specific trends or user groups. Additionally, temporal data augmentation is applied by introducing slight variations in the timing of information posts to simulate different real-world conditions, ensuring the model can predict information dissemination in rapidly changing crisis scenarios. Predictive models can analyze how information about virus outbreaks, such as COVID-19, spreads on platforms such as Twitter or Facebook [145]. By understanding how information about symptoms, treatments,

and regional outbreaks spreads, public health officials can gauge public awareness and preparedness, aiding in more effective resource allocation.

Crisis Management and Rumor Control: In emergencies such as natural disasters or political crises, controlling the spread of false or misleading information is crucial for maintaining public order. Information cascade prediction models are essential for monitoring and mitigating the spread of rumors and misinformation that could cause panic or unrest. In the process of applying these models, social media data on trending topics, shared articles, and comments are collected in real-time. The models are trained using historical data on rumor propagation during past crises, which includes variables such as user engagement patterns and the types of content that went viral. To reduce the risk of bias, ensemble methods are employed, combining several prediction models to improve the accuracy and robustness of the predictions. These models are then deployed to monitor real-time social media streams, continuously analyzing which new posts or articles are likely to go viral. Bias is further mitigated by calibrating the models across different social platforms and demographic groups, ensuring that predictions are not skewed towards specific user segments.

Models trained on social media data can predict which rumors are likely to spread widely, enabling governments or organizations to intervene early [11]. For example, during natural disasters, these models can identify posts or news articles spreading misinformation and alert authorities to issue timely corrections. These models combine real-time data to track the flow of information and predict whether a piece of content is likely to go viral, enabling rapid response. In crises such as wildfires or earthquakes, the ability to predict how crisis-related information will spread can help coordinate rescue efforts and public safety announcements. Prediction models analyze user interaction patterns to determine the most effective communication strategies for reaching a wide audience.

Marketing and Advertisement: In the field of marketing, information cascade prediction models are widely used to optimize advertising campaigns and understand consumer behavior. Predicting the viral potential of marketing content allows businesses to strategically target consumers, maximizing engagement and return on investment. The process begins by analyzing historical marketing data, including user engagement with previous campaigns, content-sharing patterns, and influencer performance. The model is trained using this historical data, identifying key features such as post timing, content type, and user demographics that predict successful engagement. To reduce bias, the model incorporates bias correction techniques, ensuring that certain demographic groups or content types are not overrepresented. After training, the model is applied to real-time data to predict which new marketing content is likely to go viral. By analyzing ongoing user interactions, the model helps businesses identify which social media influencers have the highest likelihood of sharing content with broader audiences. Businesses can use these predictions to allocate advertising resources more efficiently, focusing on content and influencers that maximize visibility. Bias mitigation is further enhanced through demographic balancing techniques, ensuring fair targeting across different consumer groups.

By predicting which marketing content is likely to be widely shared on social media, businesses can prioritize their advertising resources. Cascade prediction models identify key influencers and predict the spread of promotional content in social networks [146], enabling companies to enhance their marketing strategies by focusing on the most influential users. Social media influencers play a crucial role in information dissemination. Cascade prediction models can analyze past engagement data to determine which influencers are likely to spread content to a broader audience. Businesses can use this information to select influencers whose networks will bring the highest visibility to their products or services. This reduces guesswork in influencer marketing and improves the targeting of promotional campaigns.

6. Future Directions

Although significant progress has been made in the field of information cascade prediction, the rapid growth of social networks and big data technologies poses numerous challenges and unresolved issues in this domain. The following three directions present significant opportunities and potential for future research.

6.1. Multimodal Information Fusion

In the current social network environment, information dissemination is no longer confined to single-modal data but is instead a combination of multiple modalities such as text, images, videos, and audio. These multimodal data have different semantics and forms of expression, and they may also be influenced by factors such as user personality, platform characteristics, and information type during the dissemination process. Hence, effectively integrating multimodal information to predict information cascade dissemination more accurately has become a crucial direction for future research.

One of the key challenges in multimodal information fusion lies in balancing the contributions of different modalities. Information redundancy or noise may exist between different modalities; for instance, images and text might display different characteristics when conveying similar emotions or content. Current research primarily uses multimodal neural networks to process this information, but there is still room for optimization in dealing with the correlation and complementarity between multimodal data. For example, by introducing finer-grained attention mechanisms, the interdependencies between modalities can be more precisely captured, and modality weights can be dynamically adjusted according to context, thus enhancing the fusion effect. Multimodal information fusion must also address the challenge of data heterogeneity. In social networks, different modalities of data often possess unique spatiotemporal characteristics and data structures. For instance, video data encompasses extensive time-series information, whereas image data primarily focus on visual features. Future research could explore utilizing cross-modal alignment techniques and self-supervised learning methods to strengthen the correlations between different modalities of data and reduce prediction bias resulting from data heterogeneity. However, integrating these multimodal models into real-world information systems presents several practical challenges. Real-time processing of large volumes of multimodal data demands highly efficient algorithms to ensure scalability and low latency, especially in dynamic environments where social networks continuously evolve. In practical applications, such as social media monitoring, crisis management, and real-time content recommendation, information cascades often need to be predicted in real-time while processing data stream from multiple sources, including text, images, and videos. Synchronizing these modalities in real-time and maintaining consistency and accuracy poses significant computational challenges. Moreover, ensuring that models can dynamically adapt to changes in user behavior and the structure of social networks, while keeping computational costs manageable, remains a major hurdle. To tackle these challenges, future research could focus on developing more robust and scalable algorithms that leverage distributed computing technologies, such as Edge Computing and Federated Learning, to handle the demands of real-time, large-scale data processing. Additionally, advances in hardware acceleration, such as GPUs and TPUs, could further enable the deployment of multimodal models in real-world systems. Addressing these challenges will not only improve the integration of multimodal models into large-scale systems but also enhance their applicability in areas such as personalized content recommendations, public opinion analysis, and real-time crisis management.

6.2. Temporal Dynamics and Real-Time Forecasting

The spread of information cascades is essentially a temporally dynamic process. This time-based characteristic is reflected not only in the speed of information dissemination but also in the changes in user behavior and the evolution of propagation patterns. However, most existing prediction models assume that the information dissemination process is static

or predefined, overlooking the significant variations that might exist at different time points in the dissemination process [147]. Thus, future research should further explore refined temporal dynamic modeling to enhance the accuracy of information cascade prediction.

For temporal dynamic modeling, deep learning models based on time series, such as Transformers and Temporal Graph Neural Networks (TGNNs), can more effectively capture the temporal features of information dissemination. For instance, Transformer models can dynamically model the dissemination state of information at different moments using a Multi-head Self-attention Mechanism, identifying critical time points and turning points in the propagation process. Furthermore, Temporal Graph Neural Networks merge the strengths of Graph Neural Networks and time series analysis, enabling dynamic updates of node and edge states to more accurately simulate the dissemination paths and interaction patterns between nodes in social networks. At the same time, real-time prediction remains a pressing issue that needs to be resolved. In practical applications, the speed and scale of information dissemination often show nonlinear growth, and traditional batch processing prediction methods are inadequate for managing the challenges posed by real-time data streams. Future research could explore Incremental Learning and Online Learning methods, in conjunction with the latest Edge Computing and Federated Learning technologies, to enable real-time prediction and monitoring of information dissemination. This approach can not only enhance the model's response speed but also help identify potential dissemination trends and public opinion hotspots more effectively at the early stages of information outbreaks.

6.3. Interpretability of Large Models

In recent years, deep learning methods based on large-scale pre-trained models have shown significant performance advantages in information cascade prediction. These large models can handle large-scale multimodal data and possess strong contextual understanding and feature extraction abilities. However, their "black box" nature results in poor interpretability, which can lead to trust and ethical issues in various real-world applications. Thus, enhancing the interpretability and transparency of large models is a crucial direction for future research.

First, future research should develop new interpretability algorithms and techniques to help understand the decision-making process of large models in information cascade prediction. For instance, model-agnostic interpretability methods such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) can be integrated into cascade prediction models to analyze feature importance. SHAP values allow the calculation of feature contributions for each prediction, offering a global and local interpretability perspective. LIME can create surrogate models to approximate complex decisions, providing users with insights into how different data modalities influence the model's outputs. Furthermore, the use of attention mechanisms, especially self-attention and cross-attention layers, can provide additional transparency by visualizing which parts of the input data the model focuses on during the decision-making process. Incorporating these attention visualizations as part of the model's feedback system will enhance interpretability, making it easier for users to understand why certain predictions were made. Moreover, analysis methods based on causal inference can be employed to evaluate the influence of specific factors on information dissemination prediction outcomes, providing a deeper understanding of the model's behavior and prediction basis. Secondly, designing more transparent model architectures is an effective approach to improving the interpretability of large models. Researchers could explore combining traditional symbolic logic reasoning with deep learning methods to create "Hybrid Intelligence" models that combine strong data-driven prediction capabilities with rule-based reasoning and interpretative abilities. For instance, in the decision-making process of information dissemination, incorporating a rule-based reasoning module can offer a logical foundation and rationale for model decisions, thereby increasing user trust and acceptance. Additionally, developing new evaluation metrics to assess the interpretability and transparency of models

is an important future direction. Current model evaluation metrics typically focus only on prediction accuracy and computational efficiency, often overlooking the evaluation of model interpretability. In the future, new evaluation criteria such as Interpretability Score and Transparency Metrics could be introduced to provide a more comprehensive reference framework for model developers and users.

6.4. Heterogeneous Structures in Social Networks

In social networks, information dissemination typically occurs within highly heterogeneous network structures, characterized by complex interactions among multiple types of nodes and multiple types of edges. Many existing models have limitations in dealing with such complex heterogeneous structures, so intelligently modeling and optimizing heterogeneous structures in social networks is a critical future research direction for information cascade prediction.

The new types of Heterogeneous Graph Neural Networks (HGNNs) can be developed to specifically model the multiple types of nodes and relationships in social networks. For instance, graph neural network models based on meta-paths can learn the structural relationships in complex networks by defining connection paths between different types of nodes, allowing for better capture of potential patterns in information dissemination. Additionally, introducing dynamic attention mechanisms can dynamically adjust the model's learning strategies according to the importance of different node types and relationships, allowing the model to better adapt to various dissemination scenarios and data types. Then, it is essential to consider the dynamic and evolutionary characteristics of the network structure during the modeling of heterogeneous networks. Over time, the topology of social networks and the relationships between nodes may change significantly, which can greatly affect the paths and patterns of information dissemination. Future research could explore ways to integrate network evolution characteristics into heterogeneous graph neural network models, allowing them to capture network structure changes in real-time and dynamically update node and edge states. For example, a method combining Graph Convolutional Networks (GCNs) and Recurrent Neural Networks (RNNs) could be used to create a heterogeneous graph neural network with spatiotemporal dynamic modeling capabilities, enhancing the accuracy and robustness of information dissemination prediction.

7. Conclusions

In this survey, we conduct a comprehensive overview of information cascade prediction. We classify information cascade prediction from two perspectives, i.e., prediction targets and prediction methods. We conduct a detailed review and comparison of methods across various categories, summarizing their key aspects. Next, we explore a range of applications for information cascade prediction, along with an overview of relevant datasets and model evaluations. Additionally, we discuss the role of symmetry in information diffusion networks, highlighting how symmetrical structures can influence the efficiency and predictability of cascade propagation. Finally, we outline four potential directions for future research. While this survey offers valuable insights, it has some limitations. First, the focus on deep learning models excludes traditional and hybrid methods, which may provide complementary perspectives. Second, the datasets used mainly represent open social networks, limiting generalizability to closed or niche platforms. Lastly, the rapid development of deep learning may soon render some methods outdated. Addressing these limitations presents opportunities for future research to further enhance prediction models.

References

1. Zhou, F.; Xu, X.; Trajcevski, G.; Zhang, K. A Survey of Information Cascade Analysis: Models, Predictions, and Recent Advances. *ACM Comput. Surv.* **2022**, *54*, 1–36. [CrossRef]
2. Wang, D.; Song, C.; Barabási, A.-L. Quantifying Long-Term Scientific Impact. *Science* **2013**, *342*, 127–132. [CrossRef] [PubMed]
3. Kobayashi, R.; Lambiotte, R. TiDeH: Time-Dependent Hawkes Process for Predicting Retweet Dynamics. *arXiv* **2016**, arXiv:1603.09449. [CrossRef]
4. Zhao, Q.; Erdogdu, M.A.; He, H.Y.; Rajaraman, A.; Leskovec, J. SEISMIC: A Self-Exciting Point Process Model for Predicting Tweet Popularity. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10 August 2015; pp. 1513–1522.
5. Mishra, S.; Rizoiu, M.-A.; Xie, L. Feature Driven and Point Process Approaches for Popularity Prediction. Available online: https://arxiv.org/abs/1608.04862v2 (accessed on 25 September 2024).
6. Leskovec, J.; Adamic, L.A.; Huberman, B.A. The Dynamics of Viral Marketing. *ACM Trans. Web* **2007**, *1*, 5. [CrossRef]
7. Kempe, D.; Kleinberg, J.; Tardos, É. Maximizing the Spread of Influence through a Social Network. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 24–27 August 2003; Association for Computing Machinery: New York, NY, USA, 2003; pp. 137–146.
8. Wang, S.; Hu, L.; Cao, L.; Huang, X.; Lian, D.; Liu, W. Attention-Based Transactional Context Embedding for Next-Item Recommendation. *Proc. AAAI Conf. Artif. Intell.* **2018**, *32*, 2532–2539. [CrossRef]
9. Zhang, Z.-K.; Liu, C.; Zhan, X.-X.; Lu, X.; Zhang, C.-X.; Zhang, Y.-C. Dynamics of Information Diffusion and Its Applications on Complex Networks. *Phys. Rep.* **2016**, *651*, 1–34. [CrossRef]
10. Adamic, L.A.; Lento, T.M.; Adar, E.; Ng, P.C. Information Evolution in Social Networks. In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, San Francisco, CA, USA, 8 February 2016; pp. 473–482.
11. Vosoughi, S.; Roy, D.; Aral, S. The Spread of True and False News Online. *Science* **2018**, *359*, 1146–1151. [CrossRef]
12. Asur, S.; Huberman, B.A.; Szabo, G.; Wang, C. Trends in Social Media: Persistence and Decay. *arXiv* **2011**, arXiv:1102.1402. [CrossRef]
13. Wu, Q.; Gao, Y.; Gao, X.; Weng, P.; Chen, G. Dual Sequential Prediction Models Linking Sequential Recommendation and Information Dissemination. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 447–457.
14. Gao, H.; Kong, D.; Lu, M.; Bai, X.; Yang, J. Attention Convolutional Neural Network for Advertiser-Level Click-Through Rate Forecasting. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018; International World Wide Web Conferences Steering Committee: Geneva, Switzerland, 2018; pp. 1855–1864.
15. Cheng, J.; Adamic, L.A.; Dow, P.A.; Kleinberg, J.; Leskovec, J. Can Cascades Be Predicted? In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Republic of Korea, 7–14 April 2014; pp. 925–936.
16. Cui, P.; Jin, S.; Yu, L.; Wang, F.; Zhu, W.; Yang, S. Cascading Outbreak Prediction in Networks: A Data-Driven Approach. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; Association for Computing Machinery: New York, NY, USA, 2013; pp. 901–909.
17. Ahmed, M.; Spagna, S.; Huici, F.; Niccolini, S. A Peek into the Future: Predicting the Evolution of Popularity in User Generated Content. In Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, Rome, Italy, 4–8 February 2013; Association for Computing Machinery: New York, NY, USA, 2013; pp. 607–616.
18. Bakshy, E.; Hofman, J.M.; Mason, W.A.; Watts, D.J. Everyone's an Influencer: Quantifying Influence on Twitter. In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, Hong Kong, China, 9–12 February 2011; Association for Computing Machinery: New York, NY, USA, 2011; pp. 65–74.
19. Kupavskii, A.; Umnov, A.; Gusev, G.; Serdyukov, P. Predicting the Audience Size of a Tweet. *ICWSM* **2021**, *7*, 693–696. [CrossRef]
20. Szabo, G.; Huberman, B.A. Predicting the Popularity of Online Content. *arXiv* **2008**, arXiv:0811.0405. [CrossRef]
21. Suh, B.; Hong, L.; Pirolli, P.; Chi, E.H. Want to Be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. In Proceedings of the 2010 IEEE Second International Conference on Social Computing, Minneapolis, MN, USA, 20–22 August 2010; pp. 177–184.
22. Yu, H.; Xie, L.; Sanner, S. The Lifecyle of a Youtube Video: Phases, Content and Popularity. *Proc. Int. AAAI Conf. Web Soc. Media* **2021**, *9*, 533–542. [CrossRef]

23. Jenders, M.; Kasneci, G.; Naumann, F. Analyzing and Predicting Viral Tweets. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; Association for Computing Machinery: New York, NY, USA, 2013; pp. 657–664.

24. Carta, S.; Podda, A.S.; Recupero, D.R.; Saia, R.; Usai, G. Popularity Prediction of Instagram Posts. *Information* **2020**, *11*, 453. [CrossRef]

25. Crane, R.; Sornette, D. Robust Dynamic Classes Revealed by Measuring the Response Function of a Social System. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 15649–15653. [CrossRef] [PubMed]

26. Rizoiu, M.-A.; Xie, L.; Sanner, S.; Cebrian, M.; Yu, H.; Van Hentenryck, P. Expecting to Be HIP: Hawkes Intensity Processes for Social Media Popularity. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; pp. 735–744.

27. Xiao, S.; Yan, J.; Chu, S.M.; Yang, X.; Zha, H. Modeling The Intensity Function Of Point Process Via Recurrent Neural Networks. *arXiv* **2017**, arXiv:1705.08982. [CrossRef]

28. Wang, Y.; Shen, H.; Liu, S.; Gao, J.; Cheng, X. Cascade Dynamics Modeling with Attention-Based Recurrent Neural Network. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; AAAI Press: Melbourne, Australia, 2017; pp. 2985–2991.

29. Shen, H.-W.; Wang, D.; Song, C.; Barabási, A.-L. Modeling and Predicting Popularity Dynamics via Reinforced Poisson Processes. *arXiv* **2014**, arXiv:1401.0778. [CrossRef]

30. Gao, J.; Shen, H.; Liu, S.; Cheng, X. Modeling and Predicting Retweeting Dynamics via a Mixture Process. In Proceedings of the 25th International Conference Companion on World Wide Web, Montreal, QC, Canada, 11–15 May 2016; International World Wide Web Conferences Steering Committee: Geneva, Switzerland, 2016; pp. 33–34.

31. Gao, S.; Ma, J.; Chen, Z. Modeling and Predicting Retweeting Dynamics on Microblogging Platforms. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, Shanghai, China, 31 January–2 February 2015; Association for Computing Machinery: New York, NY, USA; pp. 107–116.

32. Liao, D.; Xu, J.; Li, G.; Huang, W.; Liu, W.; Li, J. Popularity Prediction on Online Articles with Deep Fusion of Temporal Process and Content Features. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; AAAI Press: Honolulu, HI, USA, 2019; pp. 200–207.

33. Liu, Y.; Bao, Z.; Zhang, Z.; Tang, D.; Xiong, F. Information Cascades Prediction with Attention Neural Network. *Hum. Centric Comput. Inf. Sci.* **2020**, *10*, 13. [CrossRef]

34. Shang, J.; Huang, S.; Zhang, D.; Peng, Z.; Liu, D.; Li, Y.; Xu, L. RNe2Vec: Information Diffusion Popularity Prediction Based on Repost Network Embedding. *Computing* **2021**, *103*, 271–289. [CrossRef]

35. Tang, S.; Li, Q.; Ma, X.; Gao, C.; Wang, D.; Jiang, Y.; Ma, Q.; Zhang, A.; Chen, H. Knowledge-Based Temporal Fusion Network for Interpretable Online Video Popularity Prediction. In Proceedings of the ACM Web Conference 2022, Lyon, France, 25–29 April 2022; Association for Computing Machinery: New York, NY, USA, 2022; pp. 2879–2887.

36. Xu, K.; Lin, Z.; Zhao, J.; Shi, P.; Deng, W.; Wang, H. Multimodal Deep Learning for Social Media Popularity Prediction With Attention Mechanism. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 4580–4584.

37. Zhang, Y.; Liu, J.; Guo, B.; Wang, Z.; Liang, Y.; Yu, Z. App Popularity Prediction by Incorporating Time-Varying Hierarchical Interactions. *IEEE Trans. Mob. Comput.* **2022**, *21*, 1566–1579. [CrossRef]

38. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

39. Cao, Q.; Shen, H.; Gao, J.; Wei, B.; Cheng, X. Popularity Prediction on Social Platforms with Coupled Graph Neural Networks. *arXiv* **2019**, arXiv:1906.09032.

40. Chen, X.; Zhou, F.; Zhang, K.; Trajcevski, G.; Zhong, T.; Zhang, F. Information Diffusion Prediction via Recurrent Cascades Convolution. In Proceedings of the 2019 IEEE 35th International Conference on Data Engineering (ICDE), Macao, China, 8–11 April 2019; pp. 770–781.

41. Feng, X.; Zhao, Q.; Liu, Z. Prediction of Information Cascades via Content and Structure Proximity Preserved Graph Level Embedding. *Inf. Sci.* **2021**, *560*, 424–440. [CrossRef]

42. Tang, X.; Liao, D.; Huang, W.; Xu, J.; Zhu, L.; Shen, M. Fully Exploiting Cascade Graphs for Real-Time Forwarding Prediction. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 582–590. [CrossRef]

43. Galuba, W.; Aberer, K.; Chakraborty, D.; Despotovic, Z.; Kellerer, W. Outtweeting the Twitterers—Predicting Information Cascades in Microblogs. In Proceedings of the 3rd Wonference on Online Social Networks, Boston, MA, USA, 22–25 June 2010; USENIX Association: Berkeley, CA, USA, 2010; p. 3.

44. Li, D.; Zhang, S.; Sun, X.; Zhou, H.; Li, S.; Li, X. Modeling Information Diffusion over Social Networks for Temporal Dynamic Prediction. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 1985–1997. [CrossRef]

45. Qiu, J.; Li, Y.; Tang, J.; Lu, Z.; Ye, H.; Chen, B.; Yang, Q.; Hopcroft, J. The Lifecycle and Cascade of WeChat Social Messaging Groups. *arXiv* **2016**, arXiv:1512.07831.

46. Qiu, J.; Tang, J.; Ma, H.; Dong, Y.; Wang, K.; Tang, J. DeepInf: Social Influence Prediction with Deep Learning. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 2110–2119.

47. Romero, D.M.; Meeder, B.; Kleinberg, J. Differences in the Mechanics of Information Diffusion across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter. In Proceedings of the 20th International Conference on World Wide Web, Hyderabad, India, 28 March–1 April 2011; Association for Computing Machinery: New York, NY, USA, 2011; pp. 695–704.

48. Tang, L.; Huang, Q.; Puntambekar, A.; Vigfusson, Y.; Lloyd, W.; Li, K. *Popularity Prediction of Facebook Videos for Higher Quality Streaming*; USENIX: Berkeley, CA, USA, 2017; pp. 111–123.

49. Wang, J.; Zheng, V.W.; Liu, Z.; Chang, K.C.-C. Topological Recurrent Neural Network for Diffusion Prediction. *arXiv* **2017**, arXiv:1711.10162.

50. Yang, C.; Sun, M.; Liu, H.; Han, S.; Liu, Z.; Luan, H. Neural Diffusion Model for Microscopic Cascade Prediction. *arXiv* **2018**, arXiv:1812.08933.

51. Zaman, T.R.; Herbrich, R.; Van Gael, J.; Stern, D. Predicting Information Spreading in Twitter. In Proceedings of the Workshop on Computational Social Science and the Wisdom of Crowds, NIPS 2010, Whistler, BC, Canada, 10 December 2010; Volume 104, pp. 17599–17601.

52. Jinghua, Z.; Jiale, Z.; Juan, F. Information Diffusion Prediction Based on Cascade Sequences and Social Topology. *Comput. Electr. Eng.* **2023**, *109*, 108782. [CrossRef]

53. Perozzi, B.; Al-Rfou, R.; Skiena, S. DeepWalk: Online Learning of Social Representations. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24 August 2014; pp. 701–710.

54. Li, H.; Xia, C.; Wang, T.; Wang, Z.; Cui, P.; Li, X. GRASS: Learning Spatial–Temporal Properties From Chainlike Cascade Data for Microscopic Diffusion Prediction. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *early access*. [CrossRef]

55. Cao, Q.; Shen, H.; Cen, K.; Ouyang, W.; Cheng, X. DeepHawkes: Bridging the Gap between Prediction and Understanding of Information Cascades. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, Singapore, 6–10 November 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 1149–1158.

56. Wang, S.; Zhou, L.; Kong, B. Information Cascade Prediction Based on T-DeepHawkes Model. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *715*, 012042. [CrossRef]

57. Grover, A.; Leskovec, J. Node2vec: Scalable Feature Learning for Networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 855–864.

58. Li, C.; Ma, J.; Guo, X.; Mei, Q. DeepCas: An End-to-End Predictor of Information Cascades. *arXiv* **2016**, arXiv:1611.05373.

59. Zhong, C.; Xiong, F.; Pan, S.; Wang, L.; Xiong, X. Hierarchical Attention Neural Network for Information Cascade Prediction. *Inf. Sci.* **2023**, *622*, 1109–1127. [CrossRef]

60. Yu, L.; Xu, X.; Zhong, T.; Trajcevski, G.; Zhou, F. Linking Transformer to Hawkes Process for Information Cascade Prediction (Student Abstract). *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 13103–13104. [CrossRef]

61. Yang, C.; Wang, H.; Tang, J.; Shi, C.; Sun, M.; Cui, G.; Liu, Z. Full-Scale Information Diffusion Prediction With Reinforced Recurrent Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 2271–2283. [CrossRef]

62. Xu, S.; Zhou, L.; Xu, J.; Wang, L.; Chen, H. MSIDP: Multi-Scale Information Diffusion Prediction with Timestamp Information and Wide Dispersion. In Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 18 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–10.

63. Zhao, Q.; Zhang, Y.; Feng, X. Predicting Information Diffusion via Deep Temporal Convolutional Networks. *Inf. Syst.* **2022**, *108*, 102045. [CrossRef]

64. Jiang, L.; Jia, F. Attention Based Information Cascade Prediction Model. In Proceedings of the 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 24–26 June 2022; pp. 990–996.

65. Hamilton, W.L.; Ying, R.; Leskovec, J. Inductive Representation Learning on Large Graphs. *arXiv* **2018**, arXiv:1706.02216.

66. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv* **2017**, arXiv:1609.02907.

67. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2018**, arXiv:1710.10903.

68. Li, Y.; Zhang, Z.-L. Digraph Laplacian and the Degree of Asymmetry. *Internet Math.* **2012**, *8*, 381–401. [CrossRef]

69. Sankar, A.; Zhang, X.; Krishnan, A.; Han, J. Inf-VAE: A Variational Autoencoder Framework to Integrate Homophily and Influence in Diffusion Prediction. In Proceedings of the 13th International Conference on Web Search and Data Mining, Houston, TX, USA, 20 January 2020; pp. 510–518.

70. Yuan, C.; Li, J.; Zhou, W.; Lu, Y.; Zhang, X.; Hu, S. DyHGCN: A Dynamic Heterogeneous Graph Convolutional Network to Learn Users' Dynamic Preferences for Information Diffusion Prediction. *arXiv* **2020**, arXiv:2006.05169.

71. Feng, Y.; You, H.; Zhang, Z.; Ji, R.; Gao, Y. Hypergraph Neural Networks. *Proc. AAAI Conf. Artif. Intell.* **2019**, *33*, 3558–3565. [CrossRef]

72. Bai, S.; Zhang, F.; Torr, P.H.S. Hypergraph Convolution and Hypergraph Attention. *Pattern Recognit.* **2021**, *110*, 107637. [CrossRef]

73. Sun, L.; Rao, Y.; Zhang, X.; Lan, Y.; Yu, S. MS-HGAT: Memory-Enhanced Sequential Hypergraph Attention Network for Information Diffusion Prediction. *AAAI* **2022**, *36*, 4156–4164. [CrossRef]

74. Wang, X.; Wang, L.; Su, Y.; Zhang, Y.; Liu, A.-A. MCDAN: A Multi-Scale Context-Enhanced Dynamic Attention Network for Diffusion Prediction. *IEEE Trans. Multimed.* **2024**, *26*, 7850–7862. [CrossRef]

75. Zhang, W.; Wang, W.; Wang, J.; Zha, H. User-Guided Hierarchical Attention Network for Multi-Modal Social Image Popularity Prediction. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018; International World Wide Web Conferences Steering Committee: Geneva, Switzerland, 2018; pp. 1277–1286.

76.   Wang, W.; Zhang, W.; Wang, J.; Yan, J.; Zha, H. Learning Sequential Correlation for User Generated Textual Content Popularity Prediction. In Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; AAAI Press: Stockholm, Sweden, 2018; pp. 1625–1631.

77.   Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; Hovy, E. Hierarchical Attention Networks for Document Classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; Knight, K., Nenkova, A., Rambow, O., Eds.; Association for Computational Linguistics: San Diego, CA, USA, 2016; pp. 1480–1489.

78.   Ma, T.; Pan, Q.; Rong, H.; Qian, Y.; Tian, Y.; Al-Nabhan, N. T-BERTSum: Topic-Aware Text Summarization Based on BERT. *IEEE Trans. Comput. Soc. Syst.* **2022**, *9*, 879–890. [CrossRef]

79.   Li, Z.; Du, H.; Li, X. Topic-Aware Information Coverage Maximization in Social Networks. *IEEE Trans. Comput. Soc. Syst.* **2024**, *11*, 1722–1732. [CrossRef]

80.   Jin, M.; Wang, S.; Ma, L.; Chu, Z.; Zhang, J.Y.; Shi, X.; Chen, P.-Y.; Liang, Y.; Li, Y.-F.; Pan, S.; et al. Time-LLM: Time Series Forecasting by Reprogramming Large Language Models. *arXiv* **2024**, arXiv:2310.01728.

81.   Zhong, T.; Zhang, J.; Cheng, Z.; Zhou, F.; Chen, X. Information Diffusion Prediction via Cascade-Retrieved In-Context Learning. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, Washington, DC, USA, 14–18 July 2014; ACM: Washington, DC, USA, 2024; pp. 2472–2476.

82.   Wang, H.; Yang, C.; Shi, C. Neural Information Diffusion Prediction with Topic-Aware Attention Network. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Online, 1–5 November 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 1899–1908.

83.   Hodas, N.O.; Lerman, K. The Simple Rules of Social Contagion. *Sci. Rep.* **2014**, *4*, 4343. [CrossRef]

84.   Tang, L.; Liu, H. Relational Learning via Latent Social Dimensions. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 28 June–1 July 2009; Association for Computing Machinery: New York, NY, USA, 2009; pp. 817–826.

85.   Hogg, T.; Lerman, K. Social Dynamics of Digg. *arXiv* **2012**, arXiv:1202.0031. [CrossRef]

86.   Leskovec, J.; Backstrom, L.; Kleinberg, J. Meme-Tracking and the Dynamics of the News Cycle. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 28 June–1 July; Association for Computing Machinery: New York, NY, USA, 2009; pp. 497–506.

87.   Gehrke, J.; Ginsparg, P.; Kleinberg, J. Overview of the 2003 KDD Cup. *SIGKDD Explor. Newsl.* **2003**, *5*, 149–151. [CrossRef]

88.   Luo, Z.; Liu, X. Real-Time Scholarly Retweeting Prediction System. In Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations, Santa Fe, NM, USA, 20–26 August 2018; Zhao, D., Ed.; Association for Computational Linguistics: Santa Fe, NM, USA, 2018; pp. 25–29.

89.   Kim, S.-D.; Kim, S.-H.; Cho, H.-G. Predicting the Virtual Temperature of Web-Blog Articles as a Measurement Tool for Online Popularity. In Proceedings of the 2011 IEEE 11th International Conference on Computer and Information Technology, Paphos, Cyprus, 31 August–2 September 2011; pp. 449–454. [CrossRef]

90.   Chen, G.; Kong, Q.; Mao, W. An Attention-Based Neural Popularity Prediction Model for Social Media Events. In Proceedings of the 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), Beijing, China, 22–24 July 2017; pp. 161–163.

91.   Hong, L.; Dan, O.; Davison, B.D. Predicting Popular Messages in Twitter. In Proceedings of the 20th International Conference Companion on World Wide Web, Hyderabad, India, 28 March–1 April 2011; pp. 57–58. [CrossRef]

92.   Jamali, S.; Rangwala, H. Digging Digg: Comment Mining, Popularity Prediction, and Social Network Analysis. In Proceedings of the 2009 International Conference on Web Information Systems and Mining, Shanghai, China, 7–8 November 2009; pp. 32–38.

93.   Khabiri, E.; Hsu, C.-F.; Caverlee, J. Analyzing and Predicting Community Preference of Socially Generated Metadata: A Case Study on Comments in the Digg Community. *Proc. Int. AAAI Conf. Web Soc. Media* **2009**, *3*, 238–241. [CrossRef]

94.   Bielski, A.; Trzcinski, T. Understanding Multimodal Popularity Prediction of Social Media Videos With Self-Attention. *IEEE Access* **2018**, *6*, 74277–74287. [CrossRef]

95.   Gupta, M.; Gao, J.; Zhai, C.; Han, J. Predicting Future Popularity Trend of Events in Microblogging Platforms. *Proc. Am. Soc. Inf. Sci. Technol.* **2012**, *49*, 1–10. [CrossRef]

96.   Gao, S.; Ma, J.; Chen, Z. Effective and Effortless Features for Popularity Prediction in Microblogging Network. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Republic of Korea, 7–11 April 2014; ACM: Seoul, Republic of Korea, 2014; pp. 269–270.

97.   Dong, Y.; Johnson, R.A.; Chawla, N.V. Can Scientific Impact Be Predicted? *IEEE Trans. Big Data* **2016**, *2*, 18–30. [CrossRef]

98.   Dong, Y.; Johnson, R.A.; Chawla, N.V. Will This Paper Increase Your H-Index? Scientific Impact Prediction. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, Shanghai, China, 2 February 2015; pp. 149–158.

99.   Chen, G.; Kong, Q.; Xu, N.; Mao, W. NPP: A Neural Popularity Prediction Model for Social Media Content. *Neurocomputing* **2019**, *333*, 221–230. [CrossRef]

100.   Bao, P.; Shen, H.-W.; Jin, X.; Cheng, X.-Q. Modeling and Predicting Popularity Dynamics of Microblogs Using Self-Excited Hawkes Processes. *arXiv* **2015**, arXiv:1503.02754.

101.   Bao, P.; Zhang, X. Uncovering and Predicting the Dynamic Process of Collective Attention with Survival Theory. *Sci. Rep.* **2017**, *7*, 2621. [CrossRef]

102. Ding, W.; Shang, Y.; Guo, L.; Hu, X.; Yan, R.; He, T. Video Popularity Prediction by Sentiment Propagation via Implicit Network. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Melbourne, Australia, 18–23 October 2015.

103. Xiao, S.; Yan, J.; Li, C.; Jin, B.; Wang, X.; Yang, X.; Chu, S.M.; Zhu, H. On Modeling and Predicting Individual Paper Citation Count over Time. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; AAAI Press: New York, NY, USA, 2016; pp. 2676–2682.

104. Yu, L.; Cui, P.; Wang, F.; Song, C.; Yang, S. From Micro to Macro: Uncovering and Predicting Information Cascading Process with Behavioral Dynamics. *arXiv* **2015**, arXiv:1505.07193.

105. Romero, D.M.; Galuba, W.; Asur, S.; Huberman, B.A. Influence and Passivity in Social Media. *arXiv* **2010**, arXiv:1008.1253. [CrossRef]

106. Martin, T.; Hofman, J.M.; Sharma, A.; Anderson, A.; Watts, D.J. Exploring Limits to Prediction in Complex Social Systems. *arXiv* **2016**, arXiv:1602.01013.

107. Lakkaraju, H.; Ajmera, J. Attention Prediction on Social Media Brand Pages. In Proceedings of the 20th ACM international Conference on Information and Knowledge Management, Glasgow, UK, 24–28 October 2011; Association for Computing Machinery: New York, NY, USA, 2011; pp. 2157–2160.

108. Stieglitz, S.; Dang-Xuan, L. Political Communication and Influence through Microblogging–An Empirical Analysis of Sentiment in Twitter Messages and Retweet Behavior. In Proceedings of the 2012 45th Hawaii International Conference on System Sciences, Maui, HI, USA, 4–7 January 2012; pp. 3500–3509.

109. Tatar, A.; Leguay, J.; Antoniadis, P.; Limbourg, A.; de Amorim, M.D.; Fdida, S. Predicting the Popularity of Online Articles Based on User Comments. In Proceedings of the International Conference on Web Intelligence, Mining and Semantics, Sogndal, Norway, 25–27 May 2011; Association for Computing Machinery: New York, NY, USA, 2011; pp. 1–8.

110. Ding, K.; Wang, R.; Wang, S. Social Media Popularity Prediction: A Multiple Feature Fusion Approach with Deep Neural Networks. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 2682–2686. [CrossRef]

111. He, X.; Gao, M.; Kan, M.-Y.; Liu, Y.; Sugiyama, K. Predicting the Popularity of Web 2.0 Items Based on User Comments. In Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, Gold Coast, Australia, 6–11 July 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 233–242.

112. Oghina, A.; Breuss, M.; Tsagkias, M.; de Rijke, M. Predicting IMDB Movie Ratings Using Social Media. In *Advances in Information Retrieval*; Baeza-Yates, R., de Vries, A.P., Zaragoza, H., Cambazoglu, B.B., Murdock, V., Lempel, R., Silvestri, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 503–507.

113. Samanta, B.; De, A.; Chakraborty, A.; Ganguly, N. LMPP: A Large Margin Point Process Combining Reinforcement and Competition for Modeling Hashtag Popularity. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 2679–2685.

114. Trzcinski, T.; Rokita, P. Predicting Popularity of Online Videos Using Support Vector Regression. *IEEE Trans. Multimed.* **2017**, *19*, 2561–2570. [CrossRef]

115. Wang, Y.; Ye, X.; Zhou, H.; Zha, H.; Song, L. Linking Micro Event History to Macro Prediction in Point Process Models. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR, Fort Lauderdale, FL, USA, 20–22 April 2017; pp. 1375–1384.

116. Alzahrani, S.; Alashri, S.; Koppela, A.R.; Davulcu, H.; Toroslu, I. A Network-Based Model for Predicting Hashtag Breakouts in Twitter. In *Social Computing, Behavioral-Cultural Modeling, and Prediction*; Agarwal, N., Xu, K., Osgood, N., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 3–12.

117. Bian, J.; Yang, Y.; Chua, T.-S. Predicting Trending Messages and Diffusion Participants in Microblogging Network. In Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, Gold Coast, Australia, 6–11 July 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 537–546.

118. Bora, S.; Singh, H.; Sen, A.; Bagchi, A.; Singla, P. On the Role of Conductance, Geography and Topology in Predicting Hashtag Virality. *Soc. Netw. Anal. Min.* **2015**, *5*, 57. [CrossRef]

119. Gao, X.; Cao, Z.; Li, S.; Yao, B.; Chen, G.; Tang, S. Taxonomy and Evaluation for Microblog Popularity Prediction. *ACM Trans. Knowl. Discov. Data* **2019**, *13*, 15:1–15:40. [CrossRef]

120. Gou, C.; Shen, H.; Du, P.; Wu, D.; Liu, Y.; Cheng, X. Learning Sequential Features for Cascade Outbreak Prediction. *Knowl. Inf. Syst.* **2018**, *57*, 721–739. [CrossRef]

121. Guo, R.; Shaabani, E.; Bhatnagar, A.; Shakarian, P. Toward Order-of-Magnitude Cascade Prediction. *arXiv* **2015**, arXiv:1508.03371.

122. Wang, S.; Yan, Z.; Hu, X.; Yu, P.S.; Li, Z. Burst Time Prediction in Cascades. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; AAAI Press: Washington, DC, USA, 2015; pp. 325–331.

123. Chu, Q.; Cao, Z.; Gao, X.; He, P.; Deng, Q.; Chen, G. *Cease with Bass: A Framework for Real-Time Topic Detection and Popularity Prediction Based on Long-Text Contents*; Chen, X., Sen, A., Li, W.W., Thai, M.T., Eds.; Springer International Publishing: Cham, Switzerland, 2018; Volume 11280, pp. 53–65.

124. Guo, R.; Shakarian, P. A Comparison of Methods for Cascade Prediction. In Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, 18–21 August 2016.

125. Hoang, M.X.; Dang, X.-H.; Wu, X.; Yan, Z.; Singh, A.K. GPOP: Scalable Group-Level Popularity Prediction for Online Content in Social Networks. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; International World Wide Web Conferences Steering Committee: Geneva, Switzerland, 2017; pp. 725–733.

126. Gursun, G.; Crovella, M.; Matta, I. Describing and Forecasting Video Access Patterns. In Proceedings of the 2011 IEEE INFOCOM, Shanghai, China, 10–15 April 2011; pp. 16–20. [CrossRef]

127. Kupavskii, A.; Ostroumova, L.; Umnov, A.; Usachev, S.; Serdyukov, P.; Gusev, G.; Kustarev, A. Prediction of Retweet Cascade Size over Time. In Proceedings of the 21st ACM International Conference on Information and Knowledge Management, Maui, HI, USA, 29 October 2012; Association for Computing Machinery: New York, NY, USA, 2012; pp. 2335–2338.

128. Rizos, G.; Papadopoulos, S.; Kompatsiaris, Y. Predicting News Popularity by Mining Online Discussions. In Proceedings of the 25th International Conference Companion on World Wide Web, Montreal, QC, Canada, 11–15 April 2016; International World Wide Web Conferences Steering Committee: Geneva, Switzerland, 2016; pp. 737–742.

129. Artzi, Y.; Pantel, P.; Gamon, M. Predicting Responses to Microblog Posts. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Monteal, QC, Canada, 3–8 June 2012; Fosler-Lussier, E., Riloff, E., Bangalore, S., Eds.; Association for Computational Linguistics: Montréal, QC, Canada, 2012; pp. 602–606.

130. Kong, S.; Ye, F.; Feng, L.; Zhao, Z. Towards the Prediction Problems of Bursting Hashtags on Twitter. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 2566–2579. [CrossRef]

131. Krishnan, S.; Butler, P.; Tandon, R.; Leskovec, J.; Ramakrishnan, N. Seeing the Forest for the Trees: New Approaches to Forecasting Cascades. In Proceedings of the 8th ACM Conference on Web Science, Hannover, Germany, 22–25 May 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 249–258.

132. Romero, D.M.; Tan, C.; Ugander, J. On the Interplay between Social and Topical Structure. *Proc. Int. AAAI Conf. Web Soc. Media* **2021**, *7*, 516–525. [CrossRef]

133. Shamma, D.; Yew, J.; Kennedy, L.; Churchill, E. Viral Actions: Predicting Video View Counts Using Synchronous Sharing Behaviors. *ICWSM* **2021**, *5*, 618–621. [CrossRef]

134. Tsugawa, S. Empirical Analysis of the Relation between Community Structure and Cascading Retweet Diffusion. *Proc. Int. AAAI Conf. Web Soc. Media* **2019**, *13*, 493–504. [CrossRef]

135. Zhao, Y.; Wang, C.; Chi, C.-H.; Lam, K.-Y.; Wang, S. A Comparative Study of Transactional and Semantic Approaches for Predicting Cascades on Twitter. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 1212–1218.

136. Yi, C.; Bao, Y.; Xue, Y. Mining the Key Predictors for Event Outbreaks in Social Networks. *Phys. A Stat. Mech. Its Appl.* **2016**, *447*, 247–260. [CrossRef]

137. Zhang, B.; Qian, Z.; Lu, S. Structure Pattern Analysis and Cascade Prediction in Social Networks. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Riva del Garda, Italy, 19–23 September 2016; Springer: Berlin/Heidelberg, Germany, 2016; Volume 985, pp. 524–539.

138. Yan, Y.; Tan, Z.; Gao, X.; Tang, S.; Chen, G. *STH-Bass: A Spatial-Temporal Heterogeneous Bass Model to Predict Single-Tweet Popularity*; Navathe, S.B., Wu, W., Shekhar, S., Du, X., Wang, S.X., Xiong, H., Eds.; Springer International Publishing: Cham, Switzerland, 2016; Volume 9643, pp. 18–32.

139. Weng, L.; Menczer, F.; Ahn, Y.-Y. Predicting Successful Memes Using Network and Community Structure. *arXiv* **2014**, arXiv:1403.6199. [CrossRef]

140. Kefato, Z.T.; Sheikh, N.; Bahri, L.; Soliman, A.; Montresor, A.; Girdzijauskas, S. CAS2VEC: Network-Agnostic Cascade Prediction in Online Social Networks. In Proceedings of the 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS), Valencia, Spain, 15–18 October 2018; pp. 72–79. [CrossRef]

141. Lerman, K.; Hogg, T. Using a Model of Social Dynamics to Predict Popularity of News. *arXiv* **2010**, arXiv:1004.5354.

142. Lymperopoulos, I.N. Predicting the Popularity Growth of Online Content: Model and Algorithm. *Inf. Sci.* **2016**, *369*, 585–613. [CrossRef]

143. Matsubara, Y.; Sakurai, Y.; Prakash, B.A.; Li, L.; Faloutsos, C. Rise and Fall Patterns of Information Diffusion: Model and Implications. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; Association for Computing Machinery: New York, NY, USA, 2012; pp. 6–14.

144. Yu, L.; Cui, P.; Wang, F.; Song, C.; Yang, S. Uncovering and Predicting the Dynamic Process of Information Cascades with Survival Model. *Knowl. Inf. Syst.* **2017**, *50*, 633–659. [CrossRef]

145. Wang, Y.; Hao, H.; Platt, L.S. Examining Risk and Crisis Communications of Government Agencies and Stakeholders during Early-Stages of COVID-19 on Twitter. *Comput. Human Behav.* **2021**, *114*, 106568. [CrossRef]

146. Morone, F.; Makse, H.A. Influence Maximization in Complex Networks through Optimal Percolation. *Nature* **2015**, *524*, 65–68. [CrossRef]

147. Wu, Y.; McAreavey, K.; Liu, W.; McConville, R. A Comparative Analysis of Information Cascade Prediction Using Dynamic Heterogeneous and Homogeneous Graphs. In *Complex Networks & Their Applications XII*; Cherifi, H., Rocha, L.M., Cherifi, C., Donduran, M., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 168–179.