applied sciences

MDPI

Article

UPI-LT: Enhancing Information Propagation Predictions in Social Networks Through User Influence and Temporal Dynamics

Zexia Huang [1,2], Xu Gu [3], Jinsong Hu [1,*] and Xiaoliang Chen [2,*]

1 School of Big Data and Artifcial Intelligence, Chengdu Technological University, Chengdu 611730, China; zexiahuang@mail.xhu.edu.cn
2 School of Computer and Software Engineering, Xihua University, Chengdu 610039, China
3 Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu 610041, China; guxu24@mails.ucas.ac.cn
* Correspondence: hjsong1@cdtu.edu.cn (J.H.); chenxl@mail.xhu.edu.cn (X.C.)

Abstract: The TEST pervasive use of social media has highlighted the importance of developing sophisticated models for early information warning systems within online communities. Despite the advancements that have been made, existing models often fail to adequately consider the pivotal role of network topology and temporal dynamics in information dissemination. This results in suboptimal predictions of content propagation patterns. This study introduces the User Propagation Influence-based Linear Threshold (UPI-LT) model, which represents a novel approach to the simulation of information spread. The UPI-LT model introduces an innovative approach to consider the number of active neighboring nodes, incorporating a time decay factor to account for the evolving influence of information over time. The model's technical innovations include the incorporation of a homophily ratio, which assesses the similarity between users, and a dynamic adjustment of activation thresholds, which reflect a deeper understanding of social influence mechanisms. Empirical results on real-world datasets validate the UPI-LT model's enhanced predictive capabilities for information spread.

Keywords: information dissemination; linear threshold; user propagation influence; homophily; time decay

check for updates

Citation: Huang, Z.; Gu, X.; Hu, J.; Chen, X. UPI-LT: Enhancing Information Propagation Predictions in Social Networks Through User Influence and Temporal Dynamics. *Appl. Sci.* **2024**, *14*, 9599. https://doi.org/10.3390/app14209599

Academic Editor: Luis Javier Garcia Villalba

Received: 22 September 2024
Revised: 10 October 2024
Accepted: 18 October 2024
Published: 21 October 2024

1. Introduction

In the era of social media, information sharing and dissemination have become integral to the way news, opinions, and ideas circulate online. One of the most significant mechanisms driving this phenomenon is the practice of sharing and reposting content, particularly on platforms such as Twitter. This form of digital interaction plays a critical role in generating cascades of information, where specific topics gain traction as users engage with them, leading to widespread dissemination across the network. Initially, a small group of users may show interest, but as these individuals share the content, it attracts more attention, ultimately resulting in large-scale propagation.

This cascading effect is not simply a reflection of user engagement but is deeply intertwined with the concept of social influence. On social media platforms, influence is not evenly distributed among users; instead, certain individuals—whether due to expertise, popularity, or network position—wield considerable influence over others. These influential users can shape the decisions and opinions of their followers, creating ripples throughout the online community. For instance, during critical events like the COVID-19 pandemic, public opinion was significantly influenced by the spread of both accurate information and misleading rumors. The actions of influential users during this time played a pivotal role in shaping public perception, sometimes helping to manage the flow of critical information while, at other times, amplifying harmful misinformation.

The impact of such influence highlights the need for effective identification and management of key users in social media networks. These users often hold the power to steer the conversation in particular directions, either curbing the spread of false information or contributing to its amplification. By identifying these users, it becomes possible to influence the overall direction of information flow [1,2] within a network—a task that has broad applications in areas such as topic detection [3–5], network monitoring [6], rumor control [7–9], and the simulation of information spread in critical situations like pandemics [10–12].

The spread of information on social media platforms has been extensively studied using models inspired by epidemic dynamics, such as SI, SIS, SIR, and SIRS [13–16]. These models simulate the process of information transmission between users and offer a theoretical basis for understanding how information diffuses across social networks. While these traditional models provide valuable insights, they often fall short in capturing the complexities of user behavior and the diverse interactions that characterize modern social networks. To address these limitations, researchers have developed more sophisticated models, such as the SEIRD [17] and 3SI3R [18] models, which take into account the cognitive states of users and additional factors that influence the spread of rumors and information.

In recent years, the role of social factors in information dissemination has garnered increasing attention. Models like CT-SFI [19] and MM-SIS [20] have explored the effects of cross-diffusion and the simultaneous spread of multiple pieces of information within a network. These approaches provide a more nuanced understanding of how different types of information propagate, but challenges remain. One of the primary obstacles is accurately identifying and quantifying the influence of key users in dynamic social networks. Highly influential users play a crucial role in information dissemination, yet existing models often struggle to fully capture their impact due to their reliance on static analyses. While methods such as gravity-based models and profitability-based influence ranking have introduced new perspectives on node importance, they tend to focus on static network snapshots and fail to reflect the dynamic nature of information flow. As a result, there is a pressing need for models that can more accurately represent the role of influential users in dynamic social networks, especially as information dissemination becomes increasingly rapid and complex.

This paper proposes a novel User Propagation Influence-based Linear Threshold (UPI-LT) model designed to overcome the limitations of existing information dissemination models. The UPI-LT model takes into account both the network topology and the social attributes of users to provide a more comprehensive understanding of how information spreads in dynamic environments. By incorporating metrics such as PageRank [21], the local clustering coefficient, and structural hole attributes, the model offers a holistic evaluation of a node's role in information dissemination. The UPI-LT model advances the state of the art by integrating the following three key components: the homophily ratio, time decay factor, and dynamic thresholds. These innovations aim to better reflect the real-world dynamics of social networks, where user interactions are influenced by similarities, influence diminishes over time, and network structures evolve continuously. By incorporating these factors, the UPI-LT model not only provides a more realistic representation of influence propagation but also offers superior performance in terms of both propagation effectiveness and computational efficiency. The homophily ratio builds on the work of McPherson et al. [22], emphasizing that users with similar attributes are more likely to influence one another. This concept allows the UPI-LT model to more accurately reflect the clustering tendencies seen in social networks. The time decay factor aligns with findings from previous studies on temporal dynamics in information spread [23], allowing the model to realistically diminish the influence of outdated information. Finally, the dynamic thresholds leverage insights from adaptive systems [24], enabling the model to adjust to real-time changes in network topology, thereby maintaining relevance in rapidly changing environments. Experimental results demonstrate that the UPI-LT model significantly improves both the scope and efficiency of information dissemination compared to existing models. By providing a more accurate method for identifying influential users and simulating information flow, this research offers valuable insights for a wide range of applications, from rumor control to topic detection and network monitoring.

To address the shortcomings of existing research in effectively identifying and quantifying the influence of key users in social networks, this study aims to develop a method capable of dynamically assessing user influence as network conditions evolve, thereby reflecting the inherent nature of social interactions. Furthermore, this study explores a more nuanced understanding of influence propagation by integrating the homophily ratio, time decay factor, and dynamic thresholds.

This paper is structured as follows: The Section 2 provides a detailed overview of related work, situating the UPI-LT model within the broader context of influence propagation research. Section 3 introduces the theoretical framework underlying the proposed model, outlining the centrality metrics and the rationale for incorporating homophily, time decay, and dynamic thresholds. Section 4 begins by outlining the methodology, including descriptions of the datasets, the experimental setup, and the evaluation metrics used in this study. This section then presents the experimental results, providing a comparative analysis of the UPI-LT model against several baseline models and assessing its performance across different datasets. Following the results, the implications of these findings are thoroughly discussed. Finally, in Section 5, we provide concluding remarks, highlighting the limitations of the current work and suggesting potential directions for future research.

2. Related Work

Significant advancements have been achieved in the scholarly domain of information dissemination models within online social networks. Scholars have delved into the multifaceted dynamics of this intricate mechanism, where each individual is conceptualized as a node with the capacity to propagate information via network connections and user interactions. The extant literature on information sharing in social networks can be systematically classified into the following three pivotal domains: those investigating the mechanisms of the sharing process, those analyzing the impact of social determinants, and those concentrating on the centrality of user influence.

2.1. Social Network Models

In social networks, many diffusion processes can be modeled as complex chain reactions. Kempe et al. [25] introduced two probabilistic diffusion models—the Independent Cascade (IC) and Linear Threshold (LT) models—to describe these dynamics. Among these models, the LT model is widely acknowledged as a representative framework for comprehending influence propagation mechanisms. This study further investigates the efficiency of seed node selection using the LT model [26]. DeepWalk [27] is inspired by the skip-gram model, which is used for word representations [28], and aims to learn latent node embeddings in social networks. Node2Vec [29] builds upon DeepWalk by incorporating a bias parameter, allowing for a flexible balance between depth-first and breadth-first search strategies, thereby improving the quality of the learned embeddings. Line [30] emphasizes capturing second-order proximity by focusing on common neighbors among indirectly linked nodes, which serves to enhance the understanding gained from first-order proximity data.

2.2. Process-Based Information Dissemination Research

Process-based information dissemination models have been widely studied, particularly through extensions of epidemic dynamics models. Originally designed to quantify the spread of infectious diseases, epidemic models offer powerful frameworks for analyzing various aspects of transmission, including human interaction, environmental factors, and growth patterns. These models have proven effective in simulating how diseases spread, enabling researchers to identify key causes and influencing factors.

Recognizing the similarities between disease transmission and information dissemination, researchers have adapted these models to better understand the flow of information in social networks. One notable contribution in this area is the SEIRD model proposed by Sang et al. [17], which extends the traditional Susceptible–Infectious–Recovered (SIR) model. By incorporating users' cognitive levels, the SEIRD model introduces five distinct states for individuals, allowing for a more nuanced analysis of how users' conscious behaviors and knowledge levels impact information spread in mobile social networks. This extension highlights the importance of individual cognition in information dissemination and provides a more detailed framework for simulating user behavior in digital environments. Building on the SEIRD framework, Liang et al. [31] focused on the challenges posed by uncertainty, incomplete information, and the disconnection of trust relationships in social networks. To address these issues, they developed an epidemic-based model that categorizes users into three groups, reflecting their trust levels within the network. By proposing improved methods for trust propagation and aggregation, Liang and colleagues successfully created a fully connected social network model, contributing to a more comprehensive understanding of trust dynamics in information dissemination.

In another extension of epidemic dynamics, Chen et al. [32] integrated the concept of group polarization into the Susceptible–Infectious–Recovered–Susceptible (SIRS) model. By incorporating relationship strength factors, their model was able to simulate the formation and evolution of group polarization behavior, providing valuable insights into how strong social ties among users enhance information spread and interaction. This contribution is particularly relevant for understanding the dynamics of opinion formation and polarization in online communities, where the strength of interpersonal relationships plays a crucial role in shaping collective behavior. To address the challenge of rumor propagation in social media, especially in the aftermath of emergencies, Wang et al. [18] proposed the 3SI3R model. This model extends the classical 2SI2R framework by introducing additional states and mutual reinforcement mechanisms. The inclusion of time lags and forgetting rates further refines the model, allowing it to simulate the complex dynamics of rumor spread. Through numerical simulations, Wang and colleagues demonstrated how these factors influence the propagation process.

In a related vein, Kumar et al. [33] developed a model that categorizes individuals into the following four distinct groups: spreaders, mitigators, ignorants, and stiflers. By proposing five mathematical models to simulate interactions among these groups in both homogeneous and heterogeneous environments, the authors provided a flexible and robust framework for analyzing rumor diffusion. Their work highlights the importance of considering diverse user behaviors and environments when modeling information dissemination, offering valuable tools for both researchers and practitioners. Lastly, Pan et al. [34] contributed to the study of information propagation in multilayer networks by introducing a strategy for connecting isolated networks through interconnecting edges. Their dynamic propagation model based on the susceptible–infectious–susceptible (SIS) model sheds light on how information can be promoted across different layers of a network, providing insights into cross-network propagation strategies. This work underscores the significance of network structure in enhancing the spread of information, particularly in complex, multilayered environments.

2.3. Social Factor-Based Information Dissemination Research

Research on social factor-based information dissemination models focuses on how real-world scenarios and user behaviors shape the spread of information across social networks. The simultaneous dissemination of multiple pieces of information is a critical aspect of understanding information propagation in the digital age. In practice, various pieces of information are often disseminated together, interacting in either cooperative or competitive ways. For instance, one piece of information may promote the spread of another, or conversely, it may suppress competing information. This complexity makes it essential to investigate the intricate dynamics of multi-information dissemination in social networks.

Yin et al. [19] made a valuable contribution by proposing the Cross-Transmission Susceptible Forwarding Immune (CT-SFI) model, which emphasizes cross-dissemination effects. Their model accounts for how multiple pieces of information spread simultaneously and interact with one another, offering insights into the cooperative and competitive relationships between different information streams. This approach enhances our understanding of how public events influence the overall flow of information and provides a framework for analyzing complex, real-world information dissemination scenarios. Building on this, Xiao et al. [20] introduced the MM-SIS (Multiple Information and Multiplex Network-SIS) model, which explores the varied dissemination paths and interactions among multiple pieces of information in multiplex networks. By investigating the multi-route nature of information dissemination, their model provides a detailed analysis of how different pieces of information influence each other through diverse paths in a networked environment. This work is particularly important for capturing the dynamic processes of information diffusion in systems where users are exposed to multiple streams of information, such as social media platforms, where competition between topics is a frequent occurrence.

In addition to focusing on the interactions between pieces of information, some models incorporate the social factors of individual users. Xiong et al. [35] proposed an emotion-independent cascade model to analyze the spread of emotional contagion on social media. Their model illustrates how individual emotions can influence the emotions of connected friends, contributing to the overall emotional climate within a network. By studying the contagion effect of emotions, this research highlights the role of user emotions in amplifying or dampening the spread of information. Further extending the analysis of user behavior, Li et al. [36] developed a model that integrates social factors such as user influence, topic popularity, and topic interest to study information dissemination patterns on social media. By incorporating the opinion evolution HK [37] and SEIR models, they proposed an opinion evolution model that analyzes how opinions converge or diverge among users over time. This model provides a comprehensive framework for understanding how users' opinions evolve and interact with the broader flow of information, particularly in scenarios where public opinion plays a critical role in shaping information dissemination.

2.4. User Influence-Based Information Dissemination Research

The role of influential users or opinion leaders in the spread of popular topics across social networks is a well-established area of research. Opinion leaders play a critical role by either supporting or opposing newly released information, significantly impacting the dissemination process. Their stance can directly affect the actions and decisions of their followers, thereby amplifying the reach and influence of certain information. Consequently, identifying these key nodes quickly and accurately is vital to controlling information flow, preventing potential network incidents, and optimizing the spread of critical content. The importance of identifying influential users has been widely acknowledged, as these individuals facilitate decision making and play a central role in the overall dynamics of information dissemination in complex networks [38].

In the pursuit to identify these key influencers, various methodologies have been proposed. Gupta et al. [39] focused on using the network structure to identify the top N influential nodes, addressing the issue of information dissemination in undirected and unweighted networks. Their work provides a straightforward yet effective approach to recognizing the most impactful users based on the inherent structural properties of the network, demonstrating that even in unweighted networks, key influencers can significantly affect information flow. Building on this concept, Li et al. [40] introduced a gravity model inspired by the laws of physics to measure node importance. Their model incorporates both neighborhood and path information, offering a more comprehensive understanding of a node's influence within the network. By integrating these additional factors, Li's gravity model reflects the true dynamics of information dissemination more accurately, highlighting how proximity and connectivity between users shape their influence. In contrast, Yu et al. [41] took a profitability-based approach to rank node influence. By treating each node as a person and evaluating the "profit" they provide to others, their method emphasizes the benefits that one node brings to the rest of the network. The more benefits a node offers, the more influential it is deemed, providing an innovative way to assess a node's importance by focusing on the value it generates for the network rather than purely structural features.

In a broader approach, Zhao et al. [42] identified influential nodes from a global perspective, offering insights into the network-wide dynamics of influence. Their approach underscores the necessity of considering the entire network structure rather than focusing solely on local or individual factors when identifying key influencers. Further refining the identification of key individuals, Luo et al. [43] integrated community structure into their method, considering the relationships between opinion leaders and structural hole users—individuals who bridge gaps between communities. Their work reflects a deeper understanding of how community dynamics and network structures shape information dissemination. Additionally, Luo and colleagues proposed several advancements, including an improved harmonic modularity algorithm based on Q-modularity gain (Q-HAM) and a network embedding model regulated by social hierarchy and community structure (RaComNE). These innovations offer refined tools for detecting critical individuals in social networks by balancing community cohesion with social hierarchy, improving accuracy in identifying influential nodes.

2.5. Structural Hole

The identification of structural hole users in social networks plays a crucial role in preventing potential network incidents and controlling the flow of information. Structural hole users act as bridges between different communities or subgroups, enabling the transfer of information across otherwise disconnected sections of the network. The accurate and timely identification of these key users is essential for managing the spread of information and ensuring network stability. Xu et al. [44] made significant progress in this area by proposing a top-k structural hole identification algorithm. This algorithm effectively identifies and removes the k most critical nodes in a network, thereby halting the spread of the largest portions of information. Their approach provides a practical solution for controlling information dissemination by targeting the most influential nodes that connect various parts of the network.

Expanding on the concept of structural holes, Zhang et al. [45] provided a more intuitive and detailed definition of structural hole users. They introduced a structural hole identification algorithm based on the community forest model and diminishing marginal utility. This approach highlights the relationship between community detection and structural hole identification. By integrating the diminishing marginal utility concept, Zhang et al. improved the efficiency of identifying key users, offering a refined methodology for pinpointing the most impactful structural hole users. Recognizing the strong topological correlation between communities and structural hole users, Luo et al. [43] advanced the field further by designing a deep learning model that simultaneously addresses both community detection and structural hole identification. Their unified framework leverages the inherent connection between these two tasks, enabling more accurate and efficient detection of structural hole users. By integrating these tasks within a single model, Luo et al. provided a more comprehensive understanding of how community structure influences the overall dissemination process, paving the way for more sophisticated methods of managing information flow in social networks.

Despite the substantial progress made in understanding structural hole users, there remain significant gaps in research, particularly regarding how the position of a node within the network influences the extent of information dissemination. Most existing studies rely on static models that fail to capture the dynamic nature of information propagation over time. As the dissemination of information is inherently a temporal process, it is essential to develop models that account for these time-dependent factors to gain a more complete understanding of how structural hole users influence information spread across dynamic networks.

3. Methodology

This section delineates the theoretical and computational frameworks pertinent to our research, encompassing the assessment of node influence within social network theory, network clustering coefficients, structural hole metrics, and established information dissemination models, while also introducing the novel dynamic information dissemination model predicated on user influence developed in this study.

3.1. Social Network

3.1.1. Node Importance Metrics

In a social network, the influence of a user is closely tied to their connectivity within the network. Users who can connect with a larger number of others tend to have greater influence. To quantify this, various centrality metrics are employed to evaluate the importance of individual nodes. Each of these metrics captures different aspects of a node's role in the network, from direct connections to the ability to act as a bridge in information dissemination. However, in real-world social networks, the influence exerted by nodes is not always deterministic; it is often influenced by probabilistic factors such as user behavior, time-dependent dynamics, and random events. One commonly used metric is degree centrality, which measures the importance of a node based on the number of direct connections it has to other nodes.

Specifically, the metrics are continuously updated as the network evolves, ensuring that the influence of nodes is assessed in real time. This approach captures both the structural changes in the network and the probabilistic nature of user interactions, which is vital for accurately reflecting influence in a constantly shifting environment. The evaluation metrics introduced in this section are continuously updated to ensure a real-time assessment of node influence Drawing from studies on user information in social networks, the proposed approach considers the dynamic nature of networks through the use of network snapshots at different points in time. This approach allows researchers to capture both the structural evolution of the network and the probabilistic characteristics of user interactions, which is crucial for accurately reflecting influence in an ever-changing environment. If the focus is on understanding the continuous temporal dynamics, graph neural networks or similar structures could be considered for further analysis and prediction. The underlying assumption here is that the more connections a node has, the more likely it is to influence others within the network. Thus, the degree centrality of a node reflects its potential to propagate information. Mathematically, for a given node (i) in an undirected network with n total nodes, the degree centrality is defined as the number of edges connecting node i to the remaining $n-1$ nodes. The formal expression for degree centrality is given by Equation (1).

$$C_D(N_i) = \sum_{j=1}^{n} x_{ij} \quad (i \neq j) \tag{1}$$

where N_i represents the set of nodes excluding node i and $C_D(N_i)$ denotes the degree centrality of node i. The term $\sum_{j=1}^{n} x_{ij}$ represents the number of edges connecting node i to the other $n-1$ nodes. If there is an edge between nodes, it is denoted as 1, and the condition $i \neq j$ excludes self-loops from the calculation.

Closeness centrality reflects the proximity between nodes. The shorter the average distance between a node and all other nodes, the higher its closeness centrality. However, in dynamic networks, these distances can vary due to changing network topology; thus, centrality itself may be a probabilistic value rather than a fixed one. The model updates closeness centrality dynamically as nodes and edges are added or removed, thereby ensuring that the current network structure is always accurately represented. The formula for closeness centrality is given by Equation (2).

$$C_C(N_i) = \left[\sum_{j=1}^{n} d(i,j)\right]^{-1} \quad (i \neq j) \tag{2}$$

In a graph with n nodes, $C_C(N_i)$ represents the closeness centrality of node i, where $d(i,j)$ is the distance between nodes i and j. This expression calculates the reciprocal of the sum of the average distances between node i and the remaining $n-1$ nodes. The final value lies within the range of $[0,1]$.

Betweenness centrality measures the role of a node as a bridge in the process of information dissemination. A node with a high betweenness centrality is considered an important node. The formula for betweenness centrality is shown in Equation (3).

$$C_B(N_i) = \sum_{u \neq i \neq v \in m} \frac{\sigma_{uv}(i)}{\sigma_{uv}} \tag{3}$$

where $C_B(N_i)$ represents the betweenness centrality, where $\sigma_{uv}(i)$ is the number of shortest paths from node u to node v that pass through node i and σ_{uv} is the total number of shortest paths from node u to node v. In classic structural hole theory, betweenness centrality is considered a key indicator of whether a node acts as a bridge. Nodes with high betweenness centrality are critical for information dissemination, as they facilitate the flow of information across otherwise disconnected parts of the network. However, the influence of these nodes is not static; instead, it is influenced by the dynamic context of network activity and user behavior. By updating betweenness centrality as the network evolves, the model captures the temporal aspects of node influence through Equations (1)–(3).

3.1.2. Clustering Coefficient

In graph theory, the clustering coefficient (also known as the clustering ratio or clustering index) represents the degree to which nodes in a graph tend to cluster together. Evidence shows that in real-world networks, nodes tend to form densely connected clusters. In other words, compared to randomly connected networks, real-world networks typically exhibit a higher clustering coefficient. The clustering coefficient can be divided into the global clustering coefficient for the entire network and the local clustering coefficient for each node. The global clustering coefficient reflects the overall clustering tendency of the graph, while the local clustering coefficient measures the clustering around each node.

The local clustering coefficient is an important metric in social networks, describing the degree of aggregation around a particular node. It captures the "cohesion" of a node in terms of triadic closure [46,47]. A higher clustering coefficient for a node indicates that more relationships are tightly clustered around that node. In social networks, the local clustering coefficient of a node is the probability that two neighbors of the node are also neighbors of each other. A user's friends also being friends with one another indicates a high degree of clustering among those users.

As shown in Figure 1, the clustering coefficient of node A is calculated as the ratio of the number of edges connecting node A's neighbors to the maximum number of possible edges among those neighbors. Specifically, the neighbors of node A in the figure are nodes B, C, D, and E. Based on the network topology, the maximum number of edges that could form between these neighboring nodes includes B-C, B-D, B-E, C-D, C-E, and D-E, for a total of 6 possible edges. However, in reality, only one edge (C-D) exists between these neighboring nodes. Therefore, the clustering coefficient of node A is $C_A = \frac{1}{6}$.

Additionally, the more formal definition of the local clustering coefficient can be expressed as

$$C_i = \frac{t_i}{\frac{1}{2}d_i(d_i - 1)} = \frac{2t_i}{d_i(d_i - 1)} \tag{4}$$

where d_i represents the degree of node i and t_i is the number of triangles formed by the neighbors of node i. Take Figure 2 as an example; from left to right, each of the four diagrams shows node i with three connected edges. The maximum number of triangles that could form is three. Based on the actual number of triangles formed with node i, the clustering coefficients of node i, from left to right, are $C_i = 0$, $C_i = \frac{1}{3}$, $C_i = \frac{2}{3}$, and $C_i = 1$.

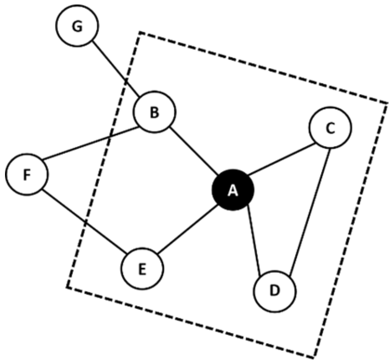


Figure 1. Schematic diagram of local clustering coefficient.

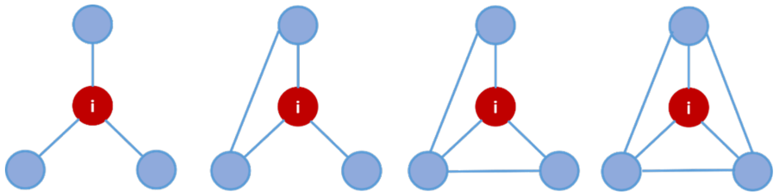


Figure 2. Clustering coefficient. The red node labeled i represents the central node in the clustering coefficient calculation, while the blue nodes represent its neighboring nodes.

3.1.3. Structure Hole

In social networks, individuals occupying structural hole positions hold a significant advantage by acting as bridges between otherwise disconnected communities. These nodes facilitate the flow of information across different groups, granting them access to more resources and opportunities. The concept of structural holes provides a valuable framework for evaluating the significance of a node's position within the network structure. In networks with multiple communities, nodes occupying structural holes play a critical role in enabling the efficient circulation of information across the entire network.

As shown in Figure 3, node B occupies a structurally pivotal position. If node B were removed, the network would fragment into three disconnected components. This highlights the importance of structural hole nodes like B, which serve as the only connection between different parts of the network. Node B not only has access to information from multiple communities but also acts as a shortcut for information flow, enhancing its influence. The strategic position of structural hole users can even grant them a certain level of "power" over neighboring nodes, as they control the flow of information between otherwise isolated groups.

To quantify the importance of structural hole positions in social networks, we used various metrics, including effective network size [40], network constraint, and betweenness centrality, to measure the presence of structural holes. The effective size of the network for node i is calculated using Equation (5).

$$ES_i = \sum_{j \in \tau(i)} \left(1 - \sum_q p_{iq} p_{jq}\right) = n - \frac{1}{n} \sum_q P_{iq} P_{jq} \tag{5}$$

where n represents the degree of node i, $\tau(i)$ is the set of all neighbors of node i, and node q is a common neighbor of nodes i and j. The terms p_{iq} and p_{jq} represent the proportion of node q among the neighbors of nodes i and j, respectively.

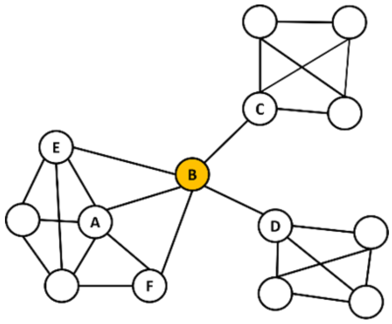


Figure 3. Structural hole diagram. A–F are different nodes, with node B highlighted in yellow to represent the key node involved in bridging structural holes between different subgroups of nodes.

The network constraint coefficient refers to the degree of dependency between nodes. The larger the value, the stronger the constraint effect, meaning the node's capacity is weaker and the likelihood of forming structural holes is smaller. The constraint is measured based on dependency relationships, calculating the strength of ties across structural holes. The formula for constraint is shown in Equation (6).

$$C_i = \sum_{j \in \tau(i)} \left(p_{ij} + \sum_{q \neq i \neq j} p_{iq} p_{qj} \right)^2 \tag{6}$$

where $\tau(i)$ is the set of neighboring nodes of node i and p_{ij} represents the proportion of the neighbor relationship between nodes i and j relative to all neighbors of node i. The terms p_{iq} and p_{qj} are calculated similarly to p_{ij}. The relationships among nodes i, j, and q are illustrated in Figure 4.

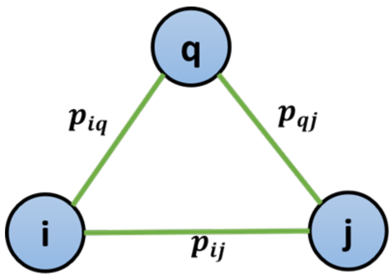


Figure 4. Schematic diagram of network constraint coefficient calculation.

Betweenness centrality refers to the extent to which a node occupies a structural hole position within the network. The larger the value, the more central the node is within the structural hole. The calculation for the betweenness centrality of node i is shown in Equation (7).

$$C_B(i) = \sum_u \sum_v \frac{g_{uv}(i)}{g_{uv}} \tag{7}$$

In this formula, g_{uv} represents the total number of shortest paths between nodes u and v, and $g_{uv}(i)$ represents the number of shortest paths between nodes u and v that pass through node i. The more frequently node i appears in these shortest paths, the more central it is within the network, particularly in bridging otherwise disconnected communities.

3.2. Information Dissemination Models

3.2.1. Independent Cascade Mode

Influence propagation models describe how nodes in a social network transition from an inactive state to an active state based on the influence of their neighbors' states. As nodes that were activated in the previous time step go on to activate other nodes in the next time step, the number of activated nodes in the network increases, eventually leading to the activation of all nodes that can be influenced. This process is known as influence diffusion within the network. Many researchers have investigated this process and proposed various models. Among them, the independent cascade (IC) model and linear threshold (LT) model are two of the most classic influence propagation models.

The IC model is a classic model for discretizing the propagation process. Extensive research has been conducted on this model by scholars worldwide. The model formalizes the social network as a directed graph $G = (V, E, W)$, where V represents the set of nodes and E represents the set of directed edges between nodes. For each directed edge ($(u \to v) \in E$), $p_{uv} \in [0, 1] \in W$ denotes the probability of the activation of node v by its incoming neighbor node (u). This probability indicates the likelihood that node v will be successfully activated by node u.

The IC model follows a process whereby activated nodes attempt to activate their unactivated neighbors at discrete time steps. Algorithm 1 details the process, which begins by selecting an initial set of activated nodes, known as the seed set (S_0), and proceeds iteratively. At each time step (t), nodes activated in the previous time step attempt to activate their neighboring nodes, with each attempt succeeding based on a predefined probability (p_{uv}). The process continues until no further activations occur in the network.

Algorithm 1 Independent Cascade Model

Input: Social network graph $G = (V, E)$, activation probability p_{uv}, seed set S_0, user activation threshold θ_v, current time step t_c, end time step t_e
Output: Set of activated nodes S_t
Initialize: $AS \leftarrow S_0$, $S_t \leftarrow S_0$
while $t_c \leq t_e$ **do**
  for each user $u \in S_{t-1} \setminus S_{t-2}$ **do**
    for each neighbor $v \in N_{\text{out}}(u) \setminus S_{t-1}$ **do**
      if $p_{uv} > \theta_v$ **then**
        Activate node v
        $v \in S_t \setminus S_{t-1}$
        $AS \leftarrow AS \cup \{v\}$, $S_t \leftarrow S_t \cup \{v\}$
      else if $p_{uv} < \theta_v$ **then**
        $v \in V \setminus S_t$, node v remains inactive
      end if
    end for
  end for
  Increment time step $t_c \leftarrow t_c + 1$
end while
return S_t

3.2.2. The Linear Threshold Model

The LT model is another classical influence propagation model. The basic concept of the LT model is that the activation state of an inactive node is determined by the collective influence of its active neighbors at a certain time. A user can only be activated if enough of its neighboring active users have sufficient influence. The main difference between the IC model and the LT model lies in how neighboring nodes exert influence. Like the IC model, the LT model defines the social network as a weighted graph represented as $G = (V, E, W)$, where each edge has a weight of $p_{uv} \in W$.

In network G, each user $v \in V$ is assigned a random threshold (θ_v, where θ_v ranges between 0 and 1). This value represents how easily user v can be activated by information; the larger the value of θ_v, the more difficult it is for the node to be influenced by its neighboring nodes. Let $u \in N(v)$ represent the set of neighboring nodes of user v, and if any neighboring node ($u \in N(v)$) is activated, the influence of node u on v is denoted by l_{uv}.

The specific propagation process of the LT model is described as follows:

- Step 1: In the network ($G = (V, E, W)$), a seed set (S) is defined. At the initial time step, only the nodes in set S are in the activated state, meaning they have already been influenced by the information.
- Step 2: Check whether the cumulative influence (l_{uv}) from neighbors is greater than or equal to the threshold (θ_v). If this condition is met, node v is successfully activated; otherwise, the activation fails.
- Step 3: This process continues until no further nodes can be activated in the network.

At any time step (t_c), nodes activated at the previous time step (t_{c-1}) remain in the activated state. Any inactive nodes at step t_{c-1} are activated at the next time step if the influence from their neighbors meets the required conditions. The overall process of information dissemination for the LT model is illustrated in Algorithm 2.

Algorithm 2 Linear Threshold Model

Input: Social network graph $G = (V, E)$, activation probability p_{uv}, seed set S_0, user activation threshold θ_v, current time step t_c, end time step t_e
Output: Set of activated nodes S_t
Initialize: $AS \leftarrow S_0$, $S_t \leftarrow S_0$
while $t_c \leq t_e$ **do**
  for each user $v \in V \setminus S_t$ **do**
    Compute the cumulative influence l_{uv} from each neighbor $u \in N(v)$
    if $\sum_{u \in N(v)} l_{uv} \geq \theta_v$ **then**
      Activate node v
      $v \in S_t \setminus S_{t-1}$
      $AS \leftarrow AS \cup \{v\}$, $S_t \leftarrow S_t \cup \{v\}$
    else
      $v \in V \setminus S_t$, node v remains inactive
    end if
  end for
  Increment time step $t_c \leftarrow t_c + 1$
end while
return S_t

To better illustrate the algorithmic details of the propagation process for this model, Figure 5 depicts the complete process of propagation in the linear threshold (LT) model. In the initial state, the network topology consists of four nodes and six edges, as shown in Figure 5a. Figure 5b shows the state at time t_1, where nodes a and b (represented as black nodes) are in the activated state, meaning they have the ability to spread information. At this point, nodes a and b begin attempting to activate their neighboring nodes (c and d).

At time t_1, the cumulative influence from nodes a and b on node c (with a weighted influence of 0.3) exceeds node c's activation threshold (0.25). As a result, node c successfully transitions from an inactive state to an active state at time t_2, as shown in Figure 5c. Similarly, at time t_1, the cumulative influence from nodes a and b on node d (0.3) is less than node d's activation threshold (0.5), so node d remains inactive.

The propagation process continues at time t_2, and now node c also gains the ability to activate other nodes. At this point, the cumulative influence from nodes a, b, and c on node d (0.6) exceeds node d's activation threshold (0.5). Node d, which was not activated at time t_1, is successfully activated under the combined influence of nodes a, b, and c. Finally, Figure 5d shows the final state of the network, where the set of activated nodes is ActiveNode $= \{a, b, c, d\}$. No further nodes can be activated in the network, marking the end of the information propagation process.

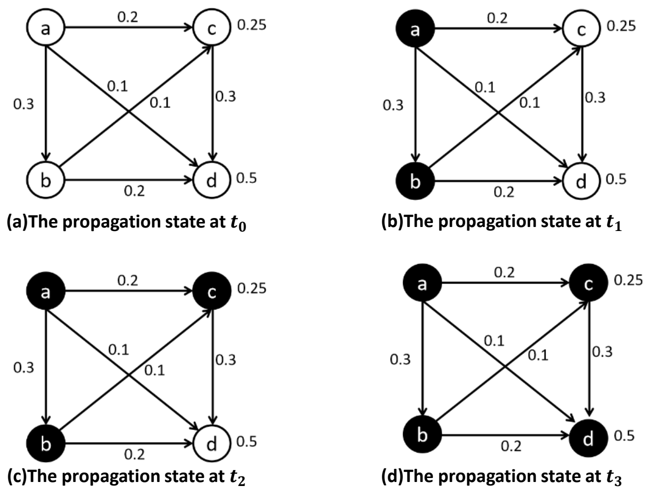


Figure 5. Propagation process of linear threshold model.

3.3. Theoretical Framework

This paper introduces a novel User Propagation Influence Linear Threshold (UPI-LT) model, which evaluates the influence of neighboring nodes on the dissemination of information based on their positions in the network. In this context, the propagation influence of a user is defined as the capability of that user to spread information, which is heavily influenced by the structural and topological attributes of its neighboring nodes. This propagation influence plays a critical role in determining the effectiveness and reach of information dissemination within the network.

To capture this propagation influence, the UPI-LT model integrates several key factors that contribute to the spread of information. First, the influence of each neighboring node is assessed based on its PageRank score, network constraint (related to structural holes), and clustering coefficient, which represent different dimensions of a node's role within the network. The PageRank metric captures a node's global importance, while the constraint coefficient measures its bridging power in structural holes, and the clustering coefficient reflects the cohesion of the node's local neighborhood. In addition to the propagation influence, the model introduces a homophily coefficient to account for the similarity between users and their neighbors, which can significantly affect the likelihood of information adoption. This homophily effect is incorporated into the activation threshold of each node, modifying the threshold based on the degree of similarity between connected nodes. The more similar the users, the lower the activation threshold, leading to a higher probability of information propagation.

Furthermore, the UPI-LT model incorporates a time decay factor to address the dynamic nature of information dissemination. Over time, the influence of information weakens as it propagates through the network. The time decay factor ensures that the cumulative influence of previously activated nodes diminishes, reflecting the natural decrease in information relevance and user engagement over time. This factor mitigates the impact of multiple activations and models the realistic attenuation of influence as the information spreads. The combined effect of these factors—the PageRank propagation influence of neighboring nodes (local clustering coefficient), the structural hole constraint, the homophily coefficient, and the time decay factor—provides a comprehensive framework for modeling information dissemination in social networks. The framework is depicted in Figure 6, illustrating the iterative process of calculating node activation based on these interconnected influences. This approach allows for a more nuanced and realistic simulation of how information flows through networks, accounting for both user attributes and temporal effects.

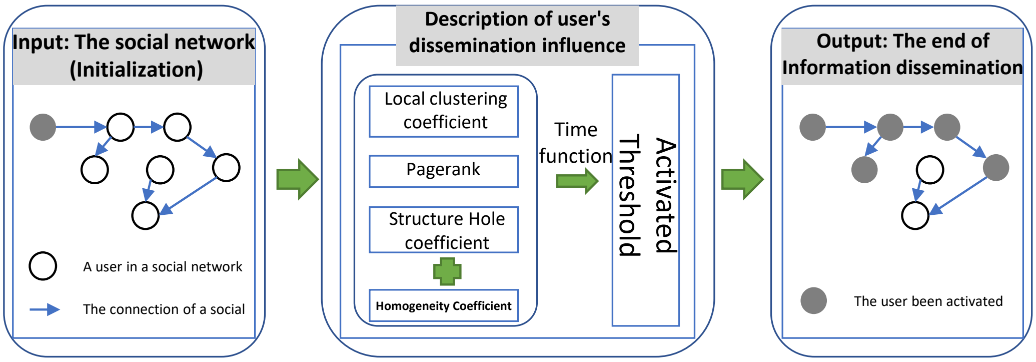


Figure 6. Frame diagram based on the user influence information dissemination model.

3.3.1. Node Influence Calculation

In social research, based on the understanding if strong and weak relationships between users, nodes in a social network play various roles. The positions of different nodes in the network topology vary, and their functions differ accordingly. Currently, many methods are available to describe the positional attributes of nodes in the network.

PageRank is a method that measures the importance of a node based on the directionality of edges, treating the directed edges in a graph as a form of endorsement. This allows for the calculation of the value or importance of each node in a directed graph. PageRank effectively illustrates a real-world scenario; when users have more connections, they tend to exert greater influence within their community. By leveraging the characteristics of the PageRank algorithm, the impact of users within a network can be measured. Therefore, this study primarily uses the PageRank value to measure a user's influence at the individual level. The PageRank calculation is shown in Equation (8):

$$PR(i) = \frac{1-\alpha}{N} + \alpha \sum_{v_j \in N_i} \frac{PR(j)}{\mathrm{Num}(N_i)} \tag{8}$$

where $\mathrm{Num}(N_i)$ represents the number of users in set N_i and N represents the total number of users in the network. The parameter α, which is usually set to 0.85 [48], is a damping factor in the PageRank algorithm.

A PageRank calculation example is shown in Figure 7 to calculate the PageRank value of node C. The iterative process for this calculation is depicted in Figure 8.

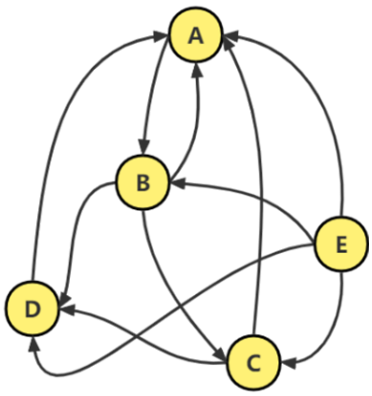


Figure 7. User link diagram.

- Initialization: Assign an initial value of 1 to all nodes in the network.
- Identify Influence: Determine which nodes influence node C. From the directed connections between the nodes, it can be seen that node C receives influence from nodes B and E, making node C a follower of B and E. During information propagation, node C is influenced by nodes B and E. As shown in Figure 7, node B influences nodes

A, C, and D, meaning node C receives an influence of $\frac{PR(B)}{3}$ from B. Similarly, nodes A and D also receive $\frac{PR(B)}{3}$ from B. Likewise, node C receives an influence of $\frac{PR(E)}{4}$ from E.

- Calculate PageRank for node $\boldsymbol{C}$: Based on the directed relationships in the network, node $\boldsymbol{C}$ is influenced by nodes $\boldsymbol{B}$ and $\boldsymbol{E}$, meaning $\boldsymbol{PR(C) = \frac{PR(B)}{3} + \frac{PR(E)}{4}}$. However, there is a certain probability that node $\boldsymbol{C}$ does not receive influence from $\boldsymbol{B}$ and $\boldsymbol{E}$ but, instead, randomly receives information from the entire network. The probability of receiving information from $\boldsymbol{B}$ and $\boldsymbol{E}$ is defined by $\boldsymbol{\alpha}$, and the PageRank formula for node $\boldsymbol{C}$ is expressed as

$$\boldsymbol{PR(C) = \frac{1-\alpha}{7} + \alpha\left(\frac{PR(B)}{3} + \frac{PR(E)}{4}\right)} \tag{9}$$

- Iterate: Repeat the above steps iteratively until the results converge.

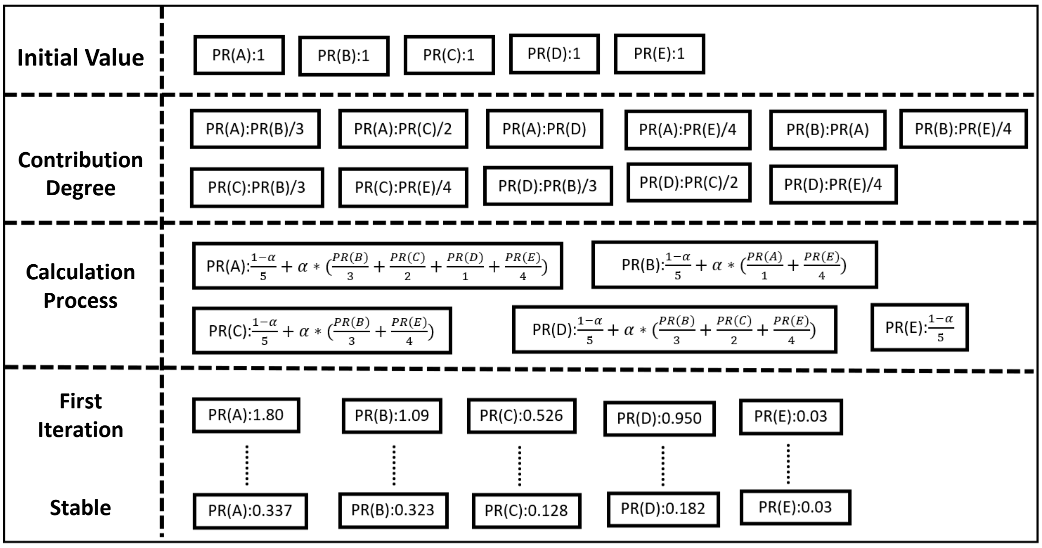


Figure 8. Iterative PageRank process.

When considering information dissemination, structural hole nodes are often overlooked. As shown in Figure 9, both nodes A and B occupy critical positions in the network. Node A is located at the center of a closely connected community, while node B is situated at the intersection of multiple communities, representing a structural hole. Node A exhibits a high degree of connectivity with its neighbors, and it has high centrality in the network topology. According to triadic closure and sociological theories, nodes with high connectivity often hold some degree of authority.

In information dissemination, nodes with high centrality are typically regarded as key nodes, whereas structural hole nodes on the periphery are often ignored. In Figure 9, although node B is less connected to its neighbors, it benefits from its cross-community position. Within the same network topology, considering two time steps, node B, situated in the structural hole, has a broader reach and influences more users at $t = 2$ compared to highly central node A. Therefore, the influence of node B in information propagation cannot be overlooked. In this study, constraint is used as a metric to quantify the structural hole's influence. When assessing a node's propagation influence in social networks, both the intrinsic properties of the node and the external environment must be considered.

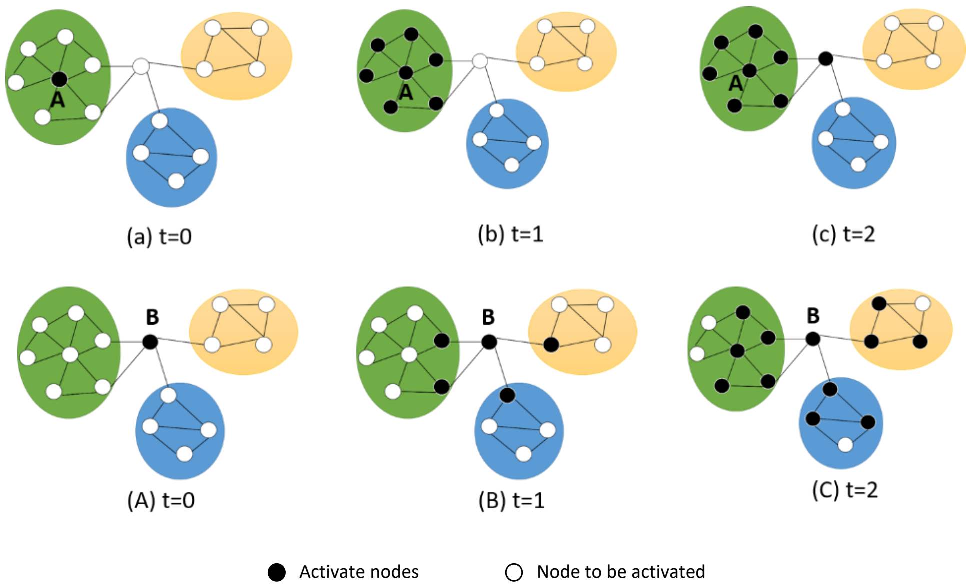


Figure 9. Analysis of user influence of a structural hole.

In this study, the network constraint coefficient ($C_s(v)$) is defined, which is calculated using Equation (10).

$$C_s(v) = \sum_{w \in \tau(v)} \left(p_{vw} + \sum_{q \neq v,w} p_{vq} p_{qw} \right)^2 \tag{10}$$

$$p_{vw} = \frac{z_{vw}}{\sum_{w \in \tau(v)} z_{vw}} \tag{11}$$

$$z_{vw} = \begin{cases} 1, & \text{if there is a direct connection between nodes } v \text{ and } w \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

In this case, set $\tau(v)$ represents the neighbors of node v, and p_{vw} indicates the proportion of the relationship between nodes v and w relative to all of node v's neighbor relationships[]. The calculations for p_{vq} and p_{qw} follow the same logic, representing the proportion of neighbor relationships between nodes v, q, and w.

If the value of $C_s(v)$ is smaller, the node experiences less constraint in occupying a structural hole and is more likely to influence information dissemination. Therefore, a node's influence should be inversely proportional to its constraint ($C_s(v)$). The local clustering coefficient is another important factor used to measure the local influence of a node. The clustering coefficient reflects the degree of cohesion around a node, describing the tightness within a local range. It represents how closely knit a group is. A group with high cohesion often exhibits greater internal trust, making it easier for members to reach a consensus. In this paper, a node with a high clustering coefficient is considered a key authority in the local network topology and an important influential user in the information dissemination process. The calculation method is expressed as follows:

$$C_i = \frac{t_i}{\frac{1}{2} \cdot d_i \cdot (d_i - 1)} = \frac{2 \cdot t_i}{d_i \cdot (d_i - 1)} \tag{13}$$

where t_i represents the number of triangles involving node i and d_i is the degree of node i. A higher clustering coefficient indicates a stronger local network presence, meaning the node plays a more influential role in its neighborhood.

The positional features of nodes from various perspectives is comprehensively considered, selecting appropriate metrics and integration methods to assess the influence of nodes in a social network. The goal is to capture a node's propagation influence from multiple perspectives, resulting in a robust calculation of node influence across different levels. The propagation influence of user v, denoted as $Inf_e(v)$, is computed using the following formula:

$$Inf_e(v) = \alpha PR(v) + \frac{1}{\beta C_s(v)} + \gamma C_v \tag{14}$$

where $PR(v)$ is the user's PageRank score, $C_s(v)$ is the constraint coefficient representing the structural hole, and C_v is the clustering coefficient. The parameters α, β, and γ are weights that sum to 1, meaning $\alpha + \beta + \gamma = 1$, and these values reflect the relative importance of each component.

Finally, to normalize the propagation influence, the following normalization is applied:

$$Inf'_e(v) = \frac{Inf_e(v) - \min(Inf_e(v))}{\max(Inf_e(v)) - \min(Inf_e(v))} \tag{15}$$

As shown in Equation (14), the propagation influence ($Inf_e(v)$) of user v during the information dissemination process is primarily determined by the following three components: the user's PageRank ($PR(v)$), the local clustering coefficient (C_v), and the constraint coefficient ($C_s(v)$), which measures the structural hole. By considering the influence at the individual level, local characteristics, and global network scope, this approach offers a comprehensive evaluation of a user's role in information propagation. Equation (15) shows the normalization of the user's propagation influence $Inf'_e(v)$, providing a standardized measure of their impact in the dissemination process.

3.3.2. The Impact of Homophily on User Thresholds

In real-world social settings, the decision as to whether a user accepts or propagates information is influenced by a variety of factors. One crucial factor is homophily, which refers to the tendency for individuals with similar characteristics, such as behaviors, interests, or beliefs, to form connections. Homophily significantly shapes group behavior and influences how information spreads within social networks. People are naturally inclined to connect with others who resemble them, and as a result, the more similar individuals are, the more likely they are to influence each other's decisions. To account for this phenomenon, a homophily coefficient is introduced into the model to adjust the activation threshold of each node. The activation threshold reflects the minimum influence required for a user to adopt and share a piece of information. By incorporating homophily, the threshold is lowered when there is a high degree of similarity between a node and its neighbors, making it easier for the node to be influenced and activated.

To better represent the dynamics of user behavior in the network, nodes are divided into the following two categories: activated (those who have adopted the information) and non-activated (those who have not). A higher proportion of a node's neighbors sharing the same activated or non-activated state signals stronger similarity or homophily within that local network. In this context, the similarity between nodes is used as a proxy for homophily, and this similarity is factored into the calculation of each node's activation threshold. Nodes with higher similarity to their neighbors are more likely to be activated, as the threshold is adjusted to reflect the greater likelihood of influence in such scenarios. The degree of homophily is quantified by examining the proportion of edges in the network that connect nodes sharing the same activation state. Let p represent the probability that a node is activated and q represent the probability that it is not, where $q = 1 - p$. The probability that two connected nodes share the same state is $p^2 + q^2$, which measures the similarity in the network.

To calculate homophily more formally, Let S represent the number of edges that connect two nodes in the same state (both activated or both non-activated) and E represent the total number of edges in the network. An S/E ratio exceeding $p^2 + q^2$ indicates the presence of homophily in the network, meaning that users are more likely to be connected to others who are in the same state, whether activated or non-activated.

3.3.3. The Impact of Time Decay on Information Diffusion

The dissemination of information in social networks is often subject to diminishing influence over time, much like the cooling of an object as described by Isaac Newton's law of cooling. In the context of information propagation, as time passes, the relevance and impact of a given piece of information gradually decreases until it is eventually ignored or forgotten by the recipients. To model this phenomenon, this study adapts Newton's law of cooling to capture the decay in information influence over time. The law of cooling states that the rate at which an object's temperature decreases is proportional to the difference between the object's temperature and the ambient temperature of its surroundings. The mathematical expression for this law is given by Equation (16).

$$T'(t) = -\alpha(T(t) - H) \tag{16}$$

where $T(t)$ represents the temperature of the object at time t, $T'(t)$ is the rate of temperature decrease, H is the ambient temperature, and α is a constant that regulates the rate of cooling. The difference $(T(t) - H)$ reflects how much hotter the object is compared to its environment. This differential equation can be solved to model the temperature $(T(t))$ at any given time as follows:

$$\frac{T'(t)}{T(t) - H} = -\alpha \tag{17}$$

$$\int \frac{T'(t)}{T(t) - H} dt = \int (-\alpha) dt \tag{18}$$

$$\ln(T(t) - H) = -\alpha t + c \tag{19}$$

$$T(t) - H = e^{-\alpha t + c} \tag{20}$$

$$T(t) = H + Ce^{-\alpha t} \tag{21}$$

Assuming that at time t_0, the object's temperature is $T(t_0)$ (abbreviated as T_0), substitution into the above equation yields the following:

$$T(t_0) = H + Ce^{-\alpha t_0} \tag{22}$$

$$C = (T(t_0) - H)e^{\alpha t_0} \tag{23}$$

Substituting Equation (23) into Equation (21) yields

$$T(t) = H + (T(t_0) - H)e^{-\alpha(t - t_0)} \tag{24}$$

Assuming that all objects cool to the ambient temperature of zero (i.e., $H = 0$), this formula is simplified as follows:

$$T(t) = T(t_0)e^{-\alpha(t - t_0)} \tag{25}$$

This paper hypothesizes that the influence of information on users decays over time in a similar manner to the cooling of an object. A "heat" value for each piece of information (e.g., a tweet) is defined, which decreases over time for each recipient. The specific formula is shown in Equation (26).

$$P(u', v') = P(u, v)e^{-\alpha(t - t_u)} \tag{26}$$

where $P(u', v')$ represents the reduced influence of information on user v' from user u' at time t and t_u is the time at which user u first received the information. The heat of the information decays exponentially over time, as described by the adapted cooling law shown in Figure 10.

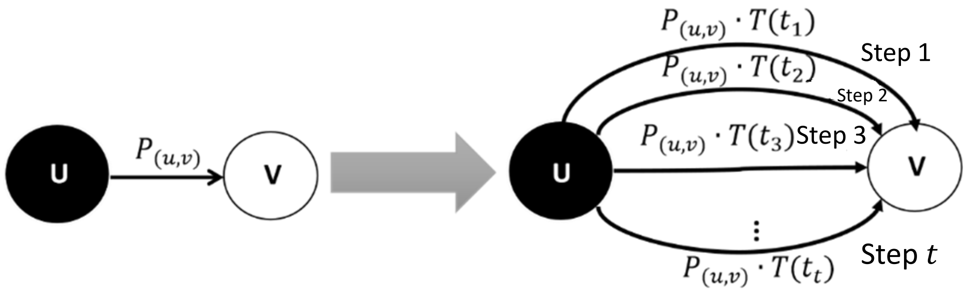


Figure 10. The information dissemination process considers the time attenuation factor.

3.3.4. The Role of Time Decay in Social Media Information Dissemination

On social media platforms, users are constantly exposed to an overwhelming amount of information, which is continually updated. Over time, older information is quickly replaced by new content. During the information dissemination process, users typically react most strongly to a piece of information when they first receive it. For example, assume that user A follows three other users, namely B, C, and D. Over a period, users B, C, and D each share the same message (I) but at different times. Their actions have varying impacts on user A's decision making.

Assume that user B shares message I first. User A sees the message from B but decides not to share it. Next, user C also shares the same message (I), and again, user A sees it but chooses not to forward it. Finally, user D shares the message (I), and at this moment, user A decides to share it. According to the linear threshold model, the order in which user A receives the information does not matter; only the cumulative influence of B, C, and D is considered. However, in real social networks, the most influential factor for user A's decision to share is typically the most recent source of information.

Hence, while sharing information, the influence of a user who recently interacted carries more weight than that of users who interacted earlier. To accommodate this, a method is suggested to measure user influence that changes over time.

The total influence on user v at time t from all active neighbors is computed as follows:

$$Sum_{P(v)} = \sum_{u_i \in neighbor^{in}} p(u,v) e^{-\alpha(t-t_0)} \tag{27}$$

where $Sum_{P(v)}$ represents the total influence from all active neighbors at time t, t_u is the time when user u attempted to activate user v, and α is the time decay factor. At time t, if the total influence ($Sum_{P(v)}$) exceeds the activation threshold ($\theta(v)$), then user v transitions from an inactive state to an active state, i.e., $Sum_{P(v)} > \theta(v)$.

3.3.5. User Propagation Influence-Based Linear Threshold Model (UPI-LT)

In the UPI-LT model, a user's activation threshold is primarily influenced by two factors, namely the information propagation capability of neighboring nodes and the level of homophily among nodes in the surrounding environment. The detailed methods for calculating these factors are provided in previous sections. The formula for calculating a user's activation threshold in the UPI-LT model is expressed as follows:

$$\theta_i = \alpha_1 \frac{\sum_{v_i \in neighbor^{in}} Inf_e(v_i)}{|v_i|} + (1-\alpha_1)\frac{S}{E} \tag{28}$$

where θ_i is the activation threshold of node i; $Inf_e(v_i)$ is the propagation influence of neighboring node v_i; and $\frac{S}{E}$ represents the homophily coefficient, where S is the number of edges between nodes of the same color and E is the total number of edges.

After calculating the θ_i value, it is necessary to normalize θ_i as follows:

$$\theta_i' = \frac{\theta_i - \min(\theta_i)}{\max(\theta_i) - \min(\theta_i)} \tag{29}$$

The final activation threshold is given by

$$\theta(i) = 1 - \theta_i' \tag{30}$$

The framework of the proposed UPI-LT model is illustrated in Figure 11. As shown in the figure, the improvements made to the LT in the UPI-LT model lie in the consideration of the propagation influence of neighboring nodes on the user's ability to receive information. On this basis, the model introduces the user's propagation influence and time decay factor, dynamically accumulating their effects over time.

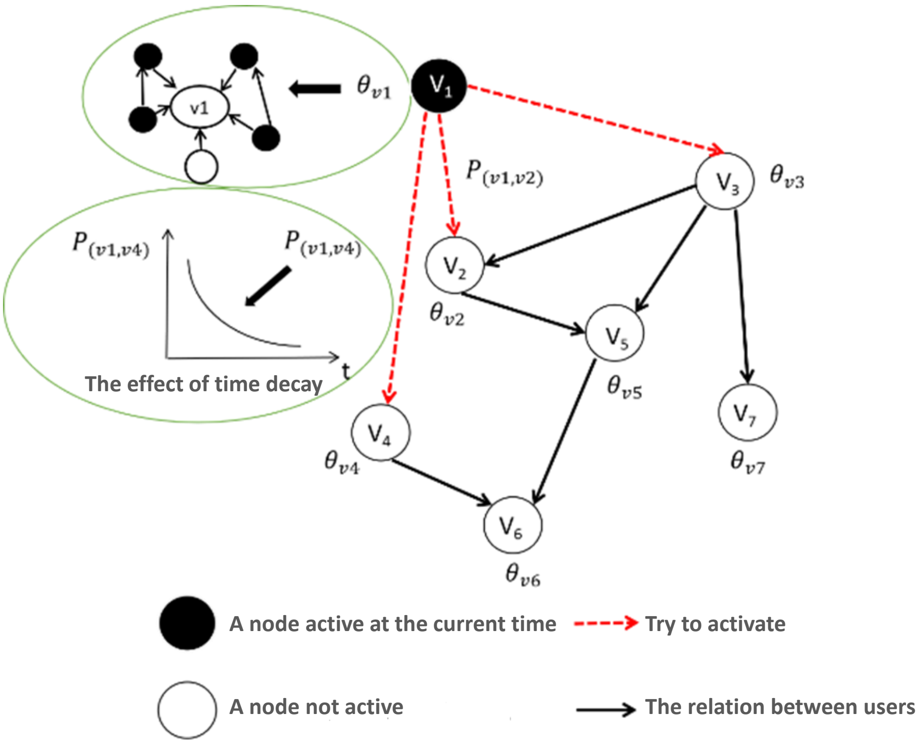


Figure 11. Model block diagram of the UPI-LT.

Figure 12 provides an example to explain the dynamic changes in propagation probability during information dissemination. In the figure, black nodes represent users currently in an activated state, while white nodes are inactive. The red dashed lines indicate the attempts of activated nodes to influence inactive nodes. Additionally, each node in Figure 12 is assigned an activation threshold, and the edges between nodes are assigned random weights to represent the influence probabilities between users. A time decay factor of $\alpha = 0.2$ is set.

The process of information dissemination with time decay characteristics is explained as follows:

The initial state of the network at time T_0 and the connections between nodes are shown in Figure 12a. The network consists of 7 nodes and 9 directed edges. The black nodes represent activated nodes, and the red dashed lines represent activation attempts between nodes.

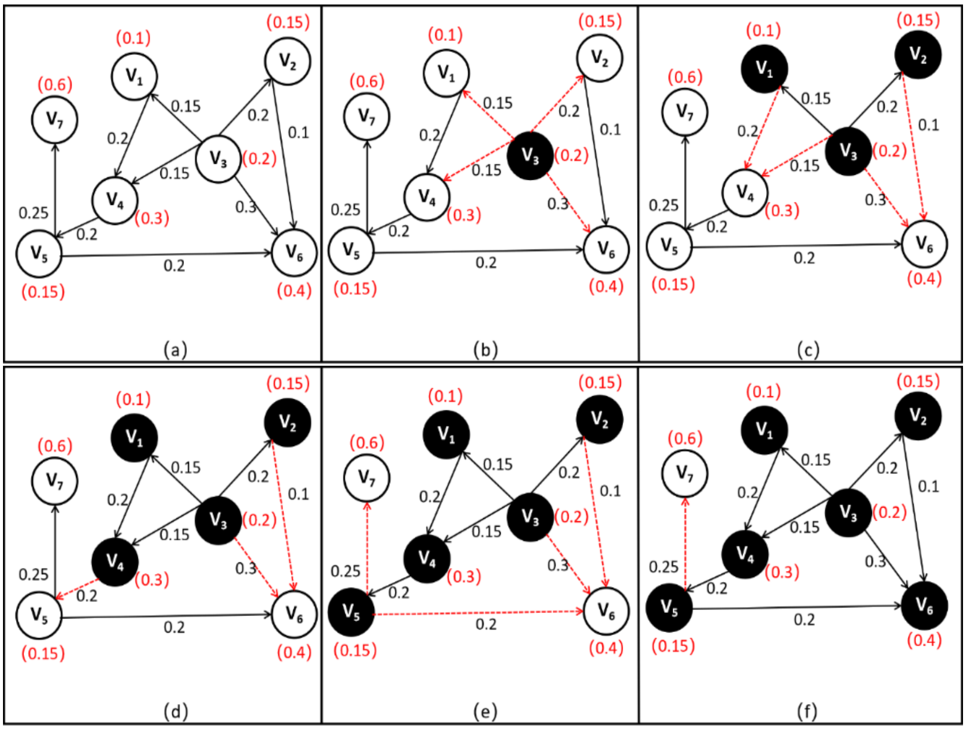


Figure 12. Diagram of the dynamic information dissemination process.

At time T_1, node v_3 is the seed node. As shown in Figure 12b, node v_3 attempts to activate its followers, which include v_1, v_2, v_4, and v_6, as indicated by the red dashed arrows. The cumulative influence on each node is calculated as follows:

- $Sum_{P(v_1)} = P(v_3, v_1) = 0.15$, and $\theta(v_1) = 0.1$, so $P(v_3, v_1) > \theta(v_1)$, resulting in node v_1 being successfully activated.
- $Sum_{P(v_2)} = P(v_3, v_2) = 0.2$, and $\theta(v_2) = 0.15$, so $P(v_3, v_2) > \theta(v_2)$, resulting in node v_2 being successfully activated.
- $Sum_{P(v_4)} = P(v_3, v_4) = 0.15$, and $\theta(v_4) = 0.3$, so $P(v_3, v_4) < \theta(v_4)$, meaning node v_4 is not activated at this moment.
- $Sum_{P(v_6)} = P(v_3, v_6) = 0.3$, and $\theta(v_6) = 0.4$, so $P(v_3, v_6) < \theta(v_6)$, meaning node v_6 is not activated.

Thus, at time T_1, the set of activated nodes is $A_{T_1} = \{v_1, v_2, v_3\}$.

At time T_2, node v_1 attempts to activate its follower, v_4, while node v_2 attempts to activate its follower, v_6. As shown in Figure 12c, the influence values are expressed as follows:

- For node v_4, the total influence is $Sum_{P(v_4)} = P(v_3, v_4) \times e^{-0.2(t_2 - t_1)} + P(v_1, v_4) \approx 0.322$, and $Sum_{P(v_4)} > 0.3$, so node v_4 is successfully activated.
- For node v_6, the total influence is $Sum_{P(v_6)} = P(v_3, v_6) \times e^{-0.2(t_2 - t_1)} + P(v_2, v_6) \approx 0.346$, and $Sum_{P(v_6)} < \theta(v_6) = 0.4$, so node v_6 is not yet activated.

Thus, at time T_2, the set of activated nodes is $A_{T_2} = \{v_1, v_2, v_3, v_4\}$.

At time T_3, node v_4 attempts to activate its follower, v_5. As shown in Figure 12d, $Sum_{P(v_5)} = P(v_4, v_5) = 0.2$, and $\theta(v_5) = 0.15$, so $P(v_4, v_5) > \theta(v_5)$, and node v_5 is successfully activated. Thus, at time T_3, the set of activated nodes is $A_{T_3} = \{v_1, v_2, v_3, v_4, v_5\}$.

At time T_4, the newly activated node (v_5) attempts to activate its followers, v_6 and v_7. As shown in Figure 12e,

- For node v_6, the total influence is $Sum_{P(v_6)} = P(v_3, v_6) \times e^{-0.2(t_3 - t_1)} + P(v_2, v_6) \times e^{-0.2(t_3 - t_2)} + P(v_5, v_6) \approx 0.483$, and since $Sum_{P(v_6)} > \theta(v_6)$, node v_6 is successfully activated.

- For node v_7, the total influence is $Sum_{P(v_7)} = P(v_5, v_7) = 0.25$, and $\theta(v_7) = 0.6$, so $P(v_5, v_7) < \theta(v_7)$, and node v_7 is not activated.

Thus, at time T_4, the set of activated nodes is $A_{T_4} = \{v_1, v_2, v_3, v_4, v_5, v_6\}$.

At time T_5, the activated node (v_5) attempts, again, to activate its follower, v_7. As shown in Figure 12f, the total influence on node v_7 is $Sum_{P(v_7)} = P(v_5, v_7) \times e^{-0.2(t_5 - t_4)} = 0.204$, and since $\theta(v_7) = 0.6$, $P(v_5, v_7) < \theta(v_7)$, and node v_7 remains inactive.

Given the decay of information over time, node v_7 is not activated. At this point, no further nodes can be activated, and the information propagation process concludes. The final set of activated nodes in the network is $A_{node} = \{v_1, v_2, v_3, v_4, v_5, v_6\}$.

The UPI-LT model introduces the following three key innovations that distinguish it from existing models: the homophily ratio, time decay factor, and dynamic thresholds. These innovations specifically enhance the model's ability to accurately capture the propagation of influence in dynamic social networks.

The homophily ratio measures the degree of similarity between users and directly influences user activation thresholds. Traditional models often assume uniform influence between users, neglecting the fact that in real social networks, individuals with similar characteristics are more likely to influence each other. The UPI-LT model incorporates homophily as a critical factor, allowing nodes with higher similarity to have a stronger likelihood of activating each other. This approach captures real-world social dynamics more effectively, making influence propagation predictions more accurate.

The time decay factor is designed to reflect the temporal aspect of influence within a social network. In many existing models, the influence of a node is considered static, remaining constant over time. However, in reality, the relevance of information diminishes as time passes. The UPI-LT model introduces a time decay factor that gradually reduces a node's influence as time progresses, accurately reflecting the diminishing impact of outdated information. By incorporating the time decay factor, our model better captures the temporal nature of information dissemination, leading to more realistic modeling of influence spread in dynamic environments.

Unlike traditional models that employ fixed activation thresholds, the UPI-LT model features dynamic thresholds that adapt to changing network conditions. Activation thresholds are recalibrated continuously based on the evolving network topology and user similarity. This dynamic adjustment allows the model to more effectively capture the fluctuating influence of nodes, particularly in response to changes in network structure or user behavior. The use of dynamic thresholds ensures that the influence propagation process remains responsive to ongoing changes, resulting in more robust and realistic predictions.

4. Results and Discussion

In this section, the experimental environment of this paper is introduced, as well as the real social network dataset. Then, the effectiveness of the proposed UPI-LT model is verified by experiments.

4.1. Environment and Datasets

The environment of this experiment and the statistics of the datasets are shown in Table 1 and Table 2, respectively.

In the conducted experiments, two real-world social network datasets published by Stanford were utilized. Among these, the Wiki-Vote network stands out as a voting dataset. In this network, the nodes symbolize Wikipedia users, and a directed edge from node *i* to node *j* indicates the voting behavior of user *i* towards user *j*. The Facebook dataset discussed in the study represents a social friendship network. It comprises users and their friendship relations extracted from Facebook, the largest social networking platform globally.

Table 1. The environment of the experiment.

	Name	Parameter
Hardware	CPU	AMD Ryzen 5 3600 6-Core Processor
	HardDisk	SSD500G
	RAM	16.0 GB
Software	Language	Python 3.10
	IDE	PyCharm 2023.1
	OS	Windows 10
	Data Format	txt, csv

Table 2. Statistical characteristics of the Wiki-Vote and Facebook datasets, along with their distributions, are shown in Figures 13 and 14.

Dataset	Node	Edge	Average Clustering Coefficient	The Longest Shortest Path
Wiki-Vote	7115	103,689	0.1409	7
Facebook	4039	88,234	0.6055	8

Figure 13. Wiki-Vote dataset. The different colors represent different communities within the Wiki-Vote dataset, with each color indicating a distinct cluster of nodes.

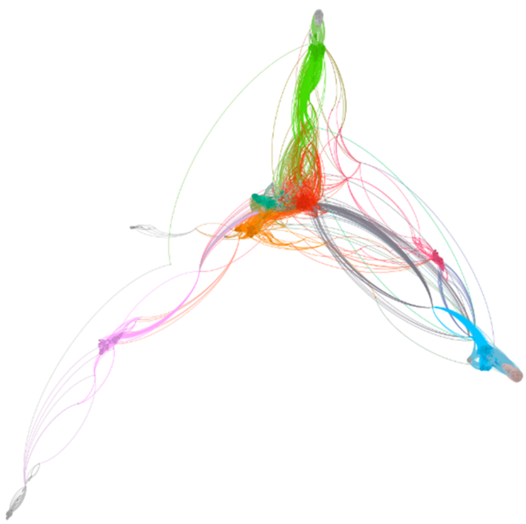


Figure 14. Facebook dataset. The different colors represent different communities within the Facebook dataset, with each color indicating a distinct cluster of nodes.

4.2. Evaluation

The number of activated users in the process of information diffusion in social networks is used as the evaluation standard for the UPI-LT model described in this paper to conduct the simulation experiment of information diffusion so as to verify that the model proposed in this paper can promote the impact of information on users and effectively expand the scope of information diffusion.

4.3. Experimental Settings

Considering that the proposed UPI-LT model considers multiple attribute characteristics of nodes, comparative methods are designed that consider only specific factors to evaluate the effectiveness of each factor. The proposed information dissemination models are then compared against these methods using two real-world datasets. To evaluate the impact of individual factors on information propagation, different factors are set as the activation thresholds in the LT model and assigned names accordingly, as shown in Table 3.

- The LT model that only considers the node's importance (PageRank) is named PR-LT.
- The LT model that only considers the local clustering coefficient is named CC-LT.
- The LT model that only considers the structural hole factor is named SH-LT.
- The classic LT model with random threshold assignments and the IC model are used as baseline models.

The number of activated users is utilized as the metric to assess the effectiveness of the proposed UPI-LT model. Influence diffusion was compared across two real-world networks, namely Wiki-Vote (http://snap.stanford.edu/data/wiki-vote.html (accessed on 7 January 2024)) and Facebook (http://snap.stanford.edu/data/ego-Facebook.html (accessed on 7 January 2024)), from publicly available Stanford datasets. During the experiment, seed nodes were randomly selected, with the number of seed nodes incrementally increasing from 5 to 35.

Table 3. Descriptions of five models of information dissemination. Notation '✓' is used to indicate that a model includes the corresponding feature, and '✕' to indicate that the feature is not included in the model.

Models	PageRank	Clustering Coefficient	Constraint	Time Decay	Homogeneity
IC	✕	✕	✕	✓	✓
LT	✕	✕	✕	✓	✓
PR-LT	✓	✕	✕	✓	✓
SH-LT	✕	✕	✓	✓	✓
CC-LT	✕	✓	✕	✓	✓
UPI-LT	✓	✓	✓	✓	✓

4.4. Experimental Result

The experimental results demonstrate the effectiveness of the proposed UPI-LT model compared to several baseline models in terms of information propagation. As illustrated in Figure 15, an increase in the number of seed nodes leads to a corresponding rise in the number of activated (infected) nodes across all models. However, the UPI-LT model consistently outperforms others, showing a broader spread of information. This highlights the significance of considering multiple node attributes in enhancing the diffusion process.

When comparing different models, it is evident that the classic LT and IC models activate the fewest nodes. This is primarily because these models rely on randomly assigned activation thresholds, leading to suboptimal propagation results. In contrast, models that incorporate specific node attributes, such as the SH-LT and PR-LT models, demonstrate significantly better performance. The SH-LT model, which emphasizes the role of structural hole nodes, activates more nodes than the CC-LT model, which only focuses on local clustering coefficients. This indicates the critical role that structural hole nodes play in expanding information dissemination, as their unique positions across multiple communities allow them to act as key conduits for information flow.

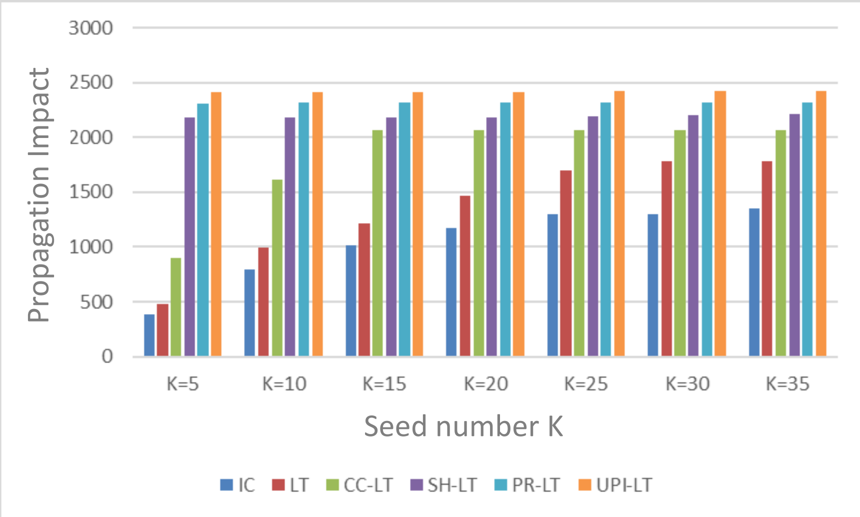


Figure 15. Propagation impact on the Wiki-Vote dataset.

Moreover, the PR-LT model, which prioritizes nodes based on their PageRank scores, achieves a performance level comparable to that of the UPI-LT model in terms of the number of activated nodes. However, the UPI-LT model exhibits a clear advantage in computational efficiency. As shown in Figure 16, when the number of seed nodes is set to $k = 30$, the UPI-LT model requires significantly less runtime than the PR-LT model while still activating a comparable number of nodes. This demonstrates that the UPI-LT model not only excels in terms of the number of nodes it activates but also in maintaining a lower computational cost, making it particularly suitable for large-scale networks. The experimental results also highlight the differences in time complexity across models. The LT model takes the longest time to run, likely due to its inefficiency in handling the random assignment of activation thresholds. In contrast, the UPI-LT model achieves the shortest runtime, followed closely by the CC-LT model. This demonstrates that the UPI-LT model provides an ideal balance between computational efficiency and activation effectiveness, making it a highly efficient model for large networks. The experimental results also highlight the differences in time complexity across models. The LT model takes the longest time to run, likely due to its inefficiency in handling the random assignment of activation thresholds. In contrast, the UPI-LT model achieves the shortest runtime, followed closely by the CC-LT model. This demonstrates that the UPI-LT model provides an ideal balance between computational efficiency and activation effectiveness, making it a highly efficient model for large networks.

Figure 17 visualizes the activation patterns of different models on the Wiki-Vote dataset. Light-green nodes represent inactive nodes, while pink nodes represent activated nodes. As seen in Figure 17a, the IC model activates the fewest nodes, resulting in a sparse spread of information. The LT model performs slightly better, as shown in Figure 17b, but it still falls short compared to models that consider node attributes. The CC-LT model, as depicted in Figure 17c, activates more nodes by focusing on local clustering, leading to more effective diffusion within localized regions. In contrast, the SH-LT model, as seen in Figure 17d, activates a larger number of nodes by leveraging the influence of structural hole users, which allows information to bridge disconnected communities.

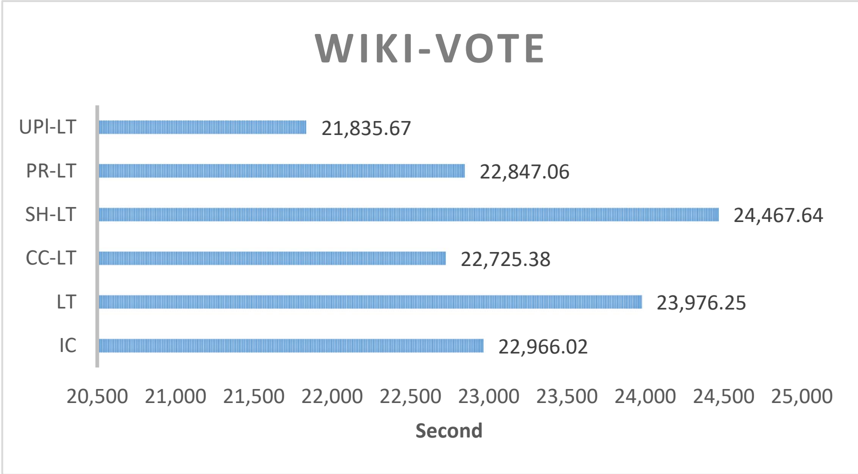


Figure 16. Runtime on the Wiki-Vote dataset.

As shown in Figure 17e,f, the UPI-LT model consistently activates more nodes than the other models, with performance closely matching that of the PR-LT model but with a slight advantage in terms of the number of activated nodes. This demonstrates that by integrating multiple factors, the UPI-LT model can more effectively spread information across the network. When examining the spatial distribution of activated nodes, it becomes evident that models focusing on single factors, such as the local clustering coefficient (CC-LT) or PageRank (PR-LT), offer a limited view of the diffusion dynamics. The Wiki-Vote dataset, which lacks distinct community structures and exhibits weak clustering, challenges these models to produce meaningful diffusion patterns. This dataset's structural properties make it difficult to visualize clear diffusion trends. However, the UPI-LT model, which combines PageRank, structural holes, and local clustering, stands out as a more comprehensive approach for the modeling of information dissemination. It shows a more balanced distribution of activated nodes across the network, offering a deeper understanding of how different node attributes contribute to information spread.

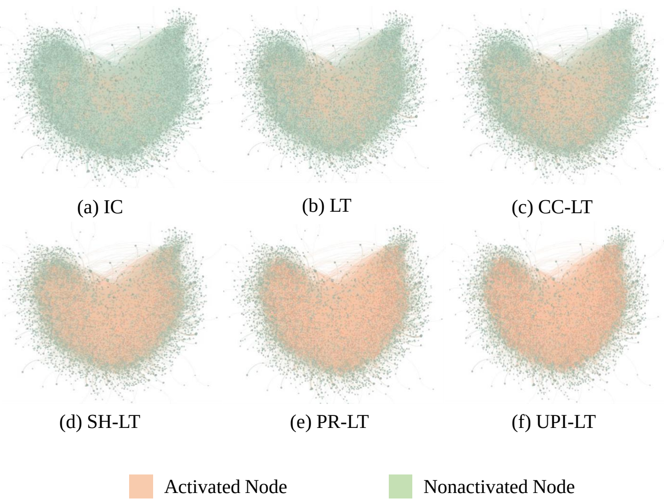


Figure 17. Visualization of propagation models on the Wiki-Vote dataset.

As seen in Figure 18, the UPI-LT model consistently outperforms other models on the Facebook dataset as well, activating more nodes across various seed node counts, ranging from $k = 5$ to $k = 35$. The number of activated nodes increases proportionally with the number of seed nodes for all models, but the UPI-LT model shows superior performance across the board. While the classic LT model demonstrates a stronger propagation effect on the Facebook dataset compared to Wiki-Vote, models that incorporate node-specific attributes, such as SH-LT and CC-LT, perform even better. The SH-LT model highlights the importance of structural hole users, who bridge different communities and facilitate the spread of information. It confirms that nodes located at structural holes play a pivotal role in ensuring the flow of information between otherwise disconnected parts of the network.

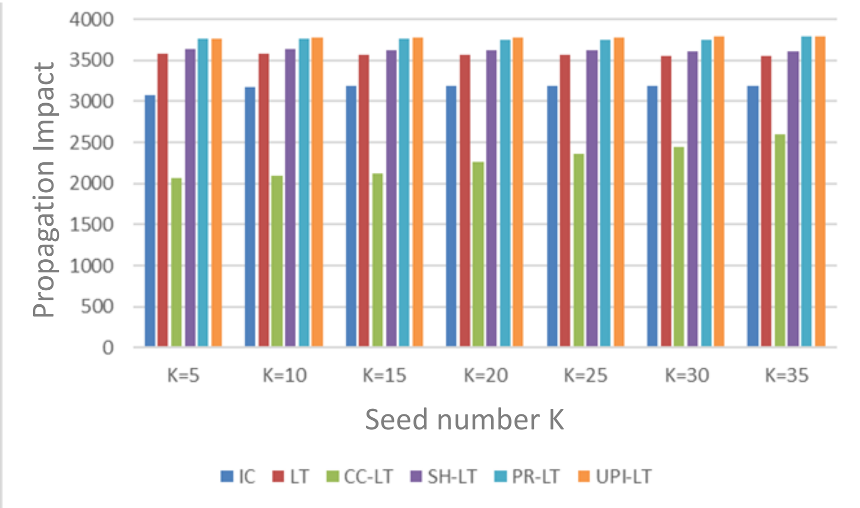


Figure 18. Propagation impact on the Facebook dataset.

The PR-LT model also performs well in activating a large number of nodes, but as indicated by the runtime comparison shown in Figure 19, it requires more computational time than other models. In contrast, the UPI-LT model strikes the optimal balance between node activation and computational efficiency. Figure 19 shows that the UPI-LT model achieves the shortest runtime after the IC model, which does not account for node attributes. This suggests that the UPI-LT model is not only effective but also computationally efficient, making it a superior choice for large-scale networks. The CC-LT model also performs well in terms of runtime but fails to match the node activation level of the UPI-LT model. The SH-LT model, although effective in spreading information through structural holes, takes the most time to execute due to the computational complexity involved in identifying these nodes. This highlights the trade-off between activation effectiveness and runtime efficiency for models focusing on structural hole nodes.

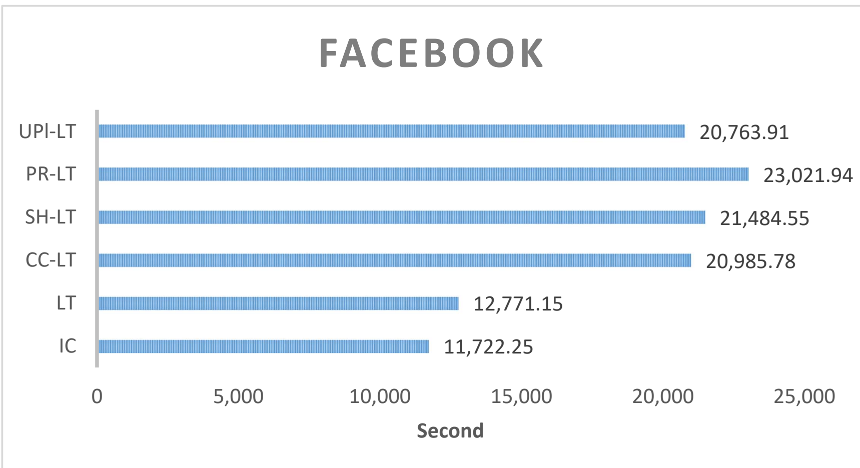
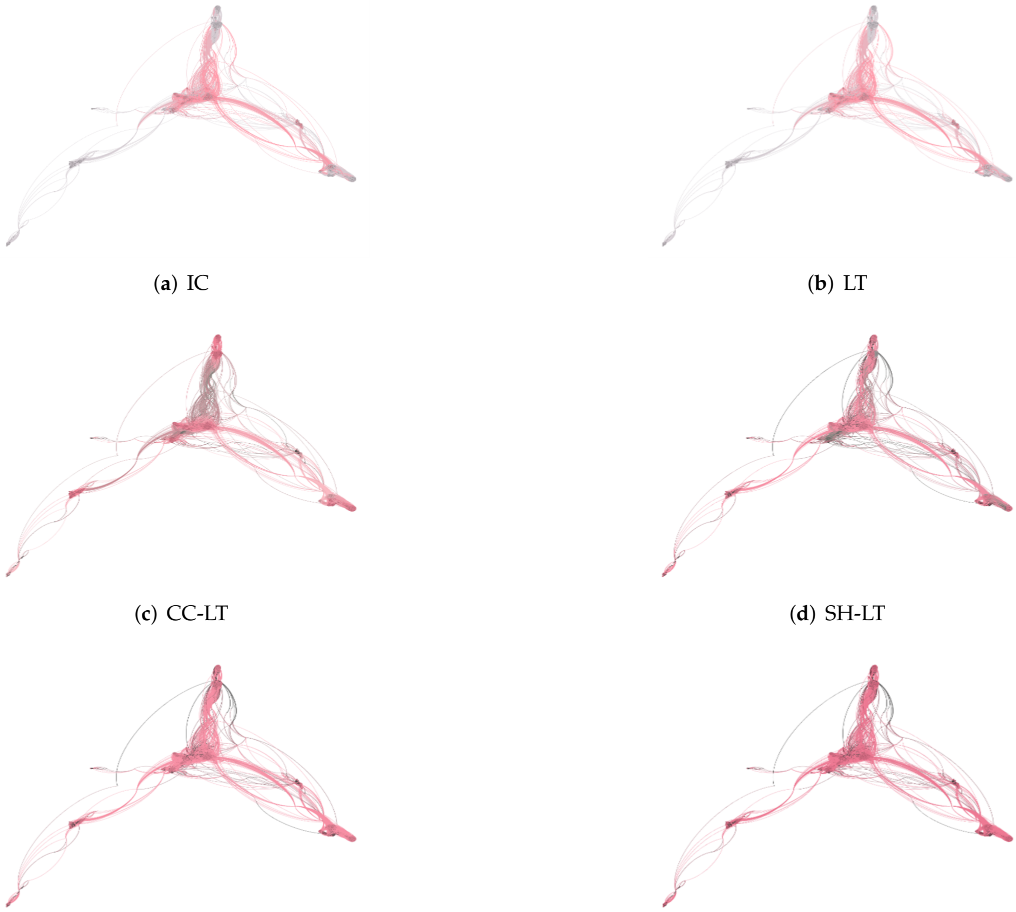


Figure 19. Runtime on the Facebook dataset.

Figure 20 provides a visualization of the infection status of nodes in the Facebook dataset across different models. Inactive users are represented by gray nodes, while active (infected) users are shown in pink. The IC and LT models activate fewer nodes, with most of the network remaining gray. The CC-LT model activates more nodes, primarily within tightly connected groups, as shown in Figure 20c. However, it struggles to activate nodes across different communities. Inversely, the SH-LT model (Figure 20d) demonstrates a more distributed diffusion pattern, with information spreading across multiple communities due to the role of structural hole nodes. The UPI-LT model, as shown in Figure 20f, achieves the most widespread activation, effectively reaching nodes across the entire network. This indicates that integrating multiple factors, such as PageRank, structural holes, and clustering, provides a more robust and comprehensive approach to information diffusion.



(**a**) IC (**b**) LT

(**c**) CC-LT (**d**) SH-LT

(**e**) PR-LT (**f**) UPL-LT

Figure 20. Visualization of propagation models on the Facebook dataset.

Analysis of the experimental results on the Wiki-Vote dataset reveals that certain characteristics limit the full potential of the UPI-LT model. The high homogeneity among nodes reduces the effectiveness of the homophily ratio, as there is limited variability to exploit for influence propagation. Additionally, the static nature of the Wiki-Vote network minimizes the benefits of dynamic thresholds, which are intended to adapt to changing conditions but have limited impact in this stable setting.

In contrast, the Facebook dataset presents a heterogeneous and dynamic network environment, which highlights the strengths of the UPI-LT model more effectively. The homophily ratio plays a larger role here due to the diversity of user attributes, allowing the model to better target influential nodes. The dynamic thresholds also shine in this context, enabling the model to adjust activation criteria in response to the frequent changes in user relationships, which is crucial for maintaining effective influence propagation. Moreover, the combination of PageRank scores, structural holes, and local clustering coefficients in the UPI-LT model allows it to comprehensively capture both global and local dynamics of influence. This approach facilitates broader and more efficient information spread, especially in the dynamic and community-diverse Facebook network. The UPI-LT model's adaptability and multifaceted strategy make it particularly effective for complex and evolving social networks, while its advantages are less pronounced in more homogeneous and stable environments like Wiki-Vote.

The experimental results from both the Wiki-Vote and Facebook datasets clearly show that the UPI-LT model excels in both node activation and runtime efficiency. By combining multiple node attributes, the UPI-LT model achieves broader and more efficient information dissemination compared to models that focus on single factors. These results validate the effectiveness of integrating the PageRank, clustering coefficient, and structural hole metrics into a single model.

5. Conclusions

This paper introduces a novel information propagation model, the User Propagation Influence Linear Threshold (UPI-LT) model, which advances the understanding of information dissemination by incorporating multiple node attributes and social factors. In contrast to traditional models, which typically rely on a single metric to assess user influence, the UPI-LT model integrates a comprehensive set of factors, including global, local, and individual node characteristics. The UPI-LT model employs metrics such as PageRank, the local clustering coefficient, and the network constraint coefficient to provide a more nuanced and accurate representation of the influence of network topology on information spread. A significant methodological contribution of the UPI-LT model is its capacity to account for both topological and social factors affecting information propagation in a dynamic manner. The model incorporates the homophily ratio to adjust a node's activation threshold, thereby acknowledging the importance of similarity between connected users. Furthermore, the time decay of influence is factored into the model, reflecting the real-world phenomenon whereby a user's impact on others diminishes over time. This dynamic consideration transforms the traditionally static approach of influence accumulation into a more realistic, temporally aware process.

Experimental results on real-world datasets validate the effectiveness of the UPI-LT model. The findings demonstrate that the model consistently activates more nodes and achieves broader information dissemination than baseline models. The superior performance of UPI-LT confirms that incorporating both network structure and temporal dynamics is critical for accurately simulating and enhancing information spread.

In addressing the research questions posed at the outset, this study has demonstrated that the UPI-LT model effectively identifies and quantifies the influence of key users within evolving social networks. By integrating homophily, time decay, and dynamic thresholds, the model provides a more nuanced understanding of how influence propagates, especially in heterogeneous and dynamic environments. Furthermore, the combination of multiple influence factors, namely PageRank, structural holes, and local clustering, results in significant improvements in the breadth and efficiency of information dissemination, answering the research questions regarding the benefits of a multifaceted approach.

The implications of this research extend beyond the scope of influence modeling, suggesting potential applications in areas such as rumor control, targeted marketing, and emergency information dissemination. However, there are limitations to the current study. One major limitation is that the model's effectiveness is highly dependent on the accuracy of the input data, particularly regarding user attributes and network dynamics. Additionally, while the model accounts for temporal dynamics, it does not explicitly address the impact of external factors, such as sudden changes in user behavior or platform-specific interventions, which could significantly influence propagation outcomes.

Future research should focus on several areas to address these limitations. First, enhancing the model to incorporate external events or shocks, such as news events or policy changes, could provide a more comprehensive understanding of influence dynamics. Additionally, extending the model to support multi-layered networks, where users participate in multiple overlapping social contexts, could further improve the accuracy of influence propagation predictions. Finally, exploring adaptive learning mechanisms that allow the model to continuously refine its parameters based on real-time data could make the UPI-LT model more robust in rapidly changing environments.

Author Contributions: Conceptualization, Z.H. and X.G.; methodology, Z.H.; software, X.G.; validation, Z.H. and X.G.; formal analysis, Z.H.; investigation, X.G.; resources, J.H.; data curation, X.G.; writing—original draft preparation, Z.H.; writing—review and editing, X.C.; visualization, X.G.; supervision, J.H.; project administration, J.H. and X.C.; funding acquisition, J.H. and X.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Talent Project of the Chengdu Technological University (No. 2024RC021), the Science and Technology Program of Sichuan Province (No. 2023YFS0424), and the National Natural Science Foundation (Nos. 62402395, 61902324).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in [Stanford SNAP] at [http://snap.stanford.edu/data/wiki-Vote.html and https:/snap.stanford.edu/data/ego-Facebook.html (accessed on 7 January 2024)].

Acknowledgments: We would like to extend our sincere appreciation to Duoqian Miao and Hongyun Zhang for their valuable discussions.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

COVID-19	Coronavirus disease 2019
SI	Susceptible–infected model
SIS	Susceptible–infected–susceptible model
SIR	Susceptible–infected–recovered model
HK	Hegselmann–Krause model
IC	Independent cascade model
LT	Linear threshold model

References

1. Bastopcu, M.; Etesami, S.R.; Başar, T. The Role of Gossiping in Information Dissemination over a Network of Agents. *Entropy* **2024**, *26*, 9
2. Li, Y.; Sheng, D. Determinants of public emergency information dissemination on social networks: A meta-analysis. *Comput. Hum. Behav.* **2024**, *152*, 108055. [CrossRef]
3. Kunachowicz, D.; Kłosowska, K.; Sobczak, N.; Kepinska, M. Applicability of Quantum Dots in Breast Cancer Diagnostic and Therapeutic Modalities. *Nanomaterials* **2024**, *14*, 1424. [CrossRef]
4. Safyari, Y.; Mahdianpari, M.; Shiri, H. A Review of Vision-Based Pothole Detection Methods Using Computer Vision and Machine Learning. *Sensors* **2024**, *24*, 5652. [CrossRef]
5. Mukil, A.; Kakade, K.S.; Alexander, C.H.C.; Susiapan, Y.S. Wavelet-based Spectral Clustering Algorithm for Network Communication Big Data. In *AIP Conference Proceedings*; American Institute of Physics: College Park, MD, USA, 2024.
6. Gu, X.; Chen, X.; Lu, P.; Li, Z.; Du, Y.; Li, X. AGCVT-prompt for sentiment classification: Automatically generating chain of thought and verbalizer in prompt learning. *Eng. Appl. Artif. Intell.* **2024**, *132*, 107907. [CrossRef]
7. Jin, M.; Liu, F.; Ning, Y.; Gao, Y.; Li, D. A Mathematical Optimization Model Designed to Determine the Optimal Timing of Online Rumor Intervention Based on Uncertainty Theory. *Mathematics* **2024**, *12*, 2457. [CrossRef]
8. Feng, Y.; Li, K.; Wang, Z. Controlling COVID-19 Rumors Through Administrative Punishment: Empirical Evidence from Zhejiang, China. *Asian J. Criminol.* **2024**, 1–20. [CrossRef]
9. Xia, Y.; Jiang, H. A Stochastic Rumor Spreading Model with Event-triggered Discontinuous Feedback Control in Multilingual Online Networks. *J. Appl. Math. Comput.* **2024**, 1–28. [CrossRef]
10. Chinebu, T.; Okafor, K.; Anoh, K.; Uzoeto, H. Smart Waterborne Disease Control for a Scalable Population Using Biodynamic Model in IoT Network. *Comput. Biol. Med.* **2024**, *181*, 109034. [CrossRef]
11. Hoang, M.T.; Pham, H.T. Global Dynamics and Numerical Simulation of a Modified Epidemiological Model for Viral Marketing on Social Networks. *Math. Comput. Simul.* **2024**, *228*, 225–244. [CrossRef]

12. Wu, J.T.; Peak, C.M. 1 Case Study: Modeling Fractional-Dose Emergency Vaccination Campaigns for Yellow Fever. In *Principles and Practice of Emergency Research*; Springer: Berlin/Heidelberg, Germany, 2024. [CrossRef]
13. Zhang, X.B.; Chang, S.; Shi, Q.; Huo, H.F. Qualitative study of a stochastic SIS epidemic model with vertical transmission. *Phys. A: Stat. Mech. Its Appl.* **2018**, *505*, 805–817. [CrossRef]
14. Kuniya, T.; Wang, J. Global dynamics of an SIR epidemic model with nonlocal diffusion. *Nonlinear Anal. Real World Appl.* **2018**, *43*, 262–282. [CrossRef]
15. Aron, J.L. Acquired immunity dependent upon exposure in an SIRS epidemic model. *Math. Biosci.* **1988**, *88*, 37–47. [CrossRef]
16. Kermack, W.O.; McKendrick, A.G. Contributions to the mathematical theory of epidemics. II.—The problem of endemicity. *Proc. R. Soc. London. Ser. Contain. Pap. Math. Phys. Character* **1932**, *138*, 55–83. [CrossRef]
17. Sang, C.; Liao, S. Modeling and simulation of information dissemination model considering user's awareness behavior in mobile social networks. *Phys. A Stat. Mech. Its Appl.* **2020**, *537*, 122639. [CrossRef]
18. Wang, Z.; Liang, J.; Nie, H.; Zhao, H. A 3SI3R model for the propagation of two rumors with mutual promotion. *Adv. Differ. Equations* **2020**, *2020*, 109. [CrossRef]
19. Yin, F.; Shao, X.; Tang, B.; Xia, X.; Wu, J. Modeling and analyzing cross-transmission dynamics of related information co-propagation. *Sci. Rep.* **2021**, *11*, 268. [CrossRef]
20. Xiao, Y.; Zhang, L.; Li, Q.; Liu, L. MM-SIS: Model for multiple information spreading in multiplex network. *Phys. A Stat. Mech. Its Appl.* **2019**, *513*, 135–146. [CrossRef]
21. Page, L.; Brin, S.; Motwani, R.; Winograd, T. *The PageRank Citation Ranking: Bringing Order to the Web*; Technical Report; Stanford InfoLab: Stanford, CA, USA, 1999. [CrossRef]
22. McPherson, M.; Smith-Lovin, L.; Cook, J.M. Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **2001**, *27*, 415–444. [CrossRef]
23. Goldenberg, J.; Libai, B.; Muller, E. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Mark. Lett.* **2001**, *12*, 211–223. [CrossRef]
24. Rogers, E.M. *Diffusion of Innovations*; Simon and Schuster: New York, NY, USA, 2010; ISBN 0-02-874074-2 (cloth); ISBN 0-02-926671-8 (paper).
25. Kempe, D.; Kleinberg, J.M.; Tardos, E. 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; pp. 137–146. [CrossRef]
26. Shakarian, P.; Bhatnagar, A.; Aleali, A.; Shaabani, E.; Guo, R. The independent cascade and linear threshold models. *Diffus. Soc. Netw.* **2015**, 35–48. [CrossRef]
27. 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, 24–27 August 2014; pp. 701–710. [CrossRef]
28. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of the 1st International Conference on Learning Representations—Workshop Track, Scottsdale, Scottsdale, AZ, USA, 2–4 May 2013.
29. Grover, A.; Leskovec, J.; Assoc Comp, M. 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; pp. 855–864. [CrossRef]
30. Tang, J.; Qu, M.; Wang, M.; Zhang, M.; Yan, J.; Mei, Q. LINE: Large-scale Information Network Embedding. In Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 18–22 May 2015; pp. 1067–1077. [CrossRef]
31. Liang, D.; Yi, B.; Xu, Z. Opinion dynamics based on infectious disease transmission model in the non-connected context of Pythagorean fuzzy trust relationship. *J. Oper. Res. Soc.* **2021**, *72*, 2783–2803. [CrossRef]
32. Chen, T.; Shi, J.; Yang, J.; Cong, G.; Li, G. Modeling public opinion polarization in group behavior by integrating SIRS-based information diffusion process. *Complexity* **2020**, *2020*, 4791527. [CrossRef]
33. Kumar, A.; Swarnakar, P.; Jaiswal, K.; Kurele, R. SMIR model for controlling the spread of information in social networking sites. *Phys. A: Stat. Mech. Its Appl.* **2020**, *540*, 122978. [CrossRef]
34. Pan, L.; Wang, W.; Cai, S.; Zhou, T. Optimal interlayer structure for promoting spreading of SIS model in two-layer networks. *Phys. Rev. E* **2019**, *100*, 022316. [CrossRef]
35. Xiong, X.; Li, Y.; Qiao, S.; Han, N.; Wu, Y.; Peng, J.; Li, B. An emotional contagion model for heterogeneous social media with multiple behaviors. *Phys. A Stat. Mech. Its Appl.* **2018**, *490*, 185–202. [CrossRef]
36. Li, Q.; Du, Y.; Li, Z.; Hu, J.; Hu, R.; Lv, B.; Jia, P. HK–SEIR model of public opinion evolution based on communication factors. *Eng. Appl. Artif. Intell.* **2021**, *100*, 104192. [CrossRef]
37. Hegselmann, R.; Krause, U. Opinion dynamics and bounded confidence models, analysis, and simulation. *J. Artif. Soc. Soc. Simul.* **2002**, *5*, 1–33.
38. Mahyar, H.; Hasheminezhad, R.; Ghalebi K., E.; Nazemian, A.; Grosu, R.; Movaghar, A.; Rabiee, H.R. Compressive sensing of high betweenness centrality nodes in networks. *Phys. A Stat. Mech. Its Appl.* **2018**, *497*, 166–184. [CrossRef]
39. Gupta, M.; Mishra, R. Spreading the information in complex networks: Identifying a set of top-N influential nodes using network structure. *Decis. Support Syst.* **2021**, *149*, 113608. [CrossRef]

40. Li, Z.; Ren, T.; Ma, X.; Liu, S.; Zhang, Y.; Zhou, T. Identifying influential spreaders by gravity model. *Sci. Rep.* **2019**, *9*, 8387. [CrossRef]
41. Yu, Z.; Shao, J.; Yang, Q.; Sun, Z. Profitleader: Identifying leaders in networks with profit capacity. *World Wide Web* **2019**, *22*, 533–553. [CrossRef]
42. Zhao, J.; Wang, Y.; Deng, Y. Identifying influential nodes in complex networks from global perspective. *Chaos Solitons Fractals* **2020**, *133*, 109637. [CrossRef]
43. Luo, J.; Du, Y. Detecting community structure and structural hole spanner simultaneously by using graph convolutional network based Auto-Encoder. *Neurocomputing* **2020**, *410*, 138–150. [CrossRef]
44. Xu, W.; Li, T.; Liang, W.; Yu, J.X.; Yang, N.; Gao, S. Identifying structural hole spanners to maximally block information propagation. *Inf. Sci.* **2019**, *505*, 100–126. [CrossRef]
45. Zhang, Y.; Xu, H.; Xu, Y.; Deng, J.; Gu, J.; Ma, R.; Lai, J.; Hu, J.; Yu, X.; Hou, L.; et al. Finding structural hole spanners based on community forest model and diminishing marginal utility in large scale social networks. *Knowl.-Based Syst.* **2020**, *199*, 105916. [CrossRef]
46. Yang, M.; Chen, X.; Chen, B.; Lu, P.; Du, Y. DNETC: Dynamic network embedding preserving both triadic closure evolution and community structures. *Knowl. Inf. Syst.* **2023**, *65*, 1129–1157. [CrossRef]
47. Gu, X.; Chen, X.; Yang, M.; Du, Y.; Tang, M. Bi-DNE: Bilayer evolutionary pattern preserved embedding for dynamic networks. *Complex Intell. Syst.* **2024**, *10*, 3763–3788. [CrossRef]
48. Langville, A.N.; Meyer, C.D. A Survey of Eigenvector Methods for Web Information Retrieval. *SIAM Rev.* **2005**, *47*, 135–161. [CrossRef]