



REVIEW

# Information diffusion analysis: process, model, deployment, and application

Shashank Sheshar Singh<sup>1</sup> , Divya Srivastava<sup>2</sup> , Madhushi Verma<sup>2</sup> and Samya Muhuri<sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering, Thapar Institute of Engineering & Technology, Patiala, Punjab, India

<sup>2</sup>School of Computer Science Engineering and Technology, Bennett University, Greater Noida, India

**Corresponding authors:** Shashank Sheshar Singh; Email: [shashank.sheshar@gmail.com](mailto:shashank.sheshar@gmail.com), Divya Srivastava; Email: [divyalknw@gmail.com](mailto:divyalknw@gmail.com)

**Received:** 30 October 2023; **Revised:** 3 October 2024; **Accepted:** 7 October 2024

**Keywords:** Information Diffusion; Link Prediction; Social Networks; Community Detection

## Abstract

The information deployment on social networks through word-of-mouth spreading by online users has contributed well to forming opinions, social groups, and connections. This process of information deployment is known as information diffusion. Its process and models play a significant role in social network analysis. Seeing this importance, the present paper focuses on the process, model, deployment, and applications of information diffusion analysis. First, this article discusses the background of the diffusion process, such as process, components, and models. Next, information deployment in social networks and their application have been discussed. A comparative analysis of literature corresponding to applications like influence maximization, link prediction, and community detection is presented. A brief description of performative evaluation metrics is illustrated. Current research challenges and the future direction of information diffusion analysis regarding social network applications have been discussed. In addition, some open problems of information diffusion for social network analysis are also presented.

## 1. Introduction

In the digital era, social networks are gaining popularity and contributing to the information diffusion process. Therefore, analyzing the importance of social networks in information diffusion (Granovetter, 1973) is one of the focused areas of research among researchers. Users are not the only receivers of information these days; instead, they contribute largely to the evolution and propagation of information. The behavior of the individuals plays an essential role in the diffusion process as they participate in the exchange of information among peers and neighbors, defining their social relations and creating networks. This social network gradually grows as the topological relations of the users are covered, eventually leading to a large and complex network that can act as a medium for information diffusion. Social influence can be explained as a sort of behavioral change that is induced in an individual by another called an influencer. This change can be either knowingly/directly or unknowingly/indirectly (Chen *et al.*, 2014). Factors like network distance, the strength of relationships among the individuals, characteristics of the individuals, timing, characteristics of the network, etc, are the factors that affect the level of social influence. The rapid growth and emergence of technology have increased the pace of information diffusion. Social media websites have contributed largely to the diffusion process as they are handy tools for spreading information faster. This leads to the need to analyze the influence and diffusion

---

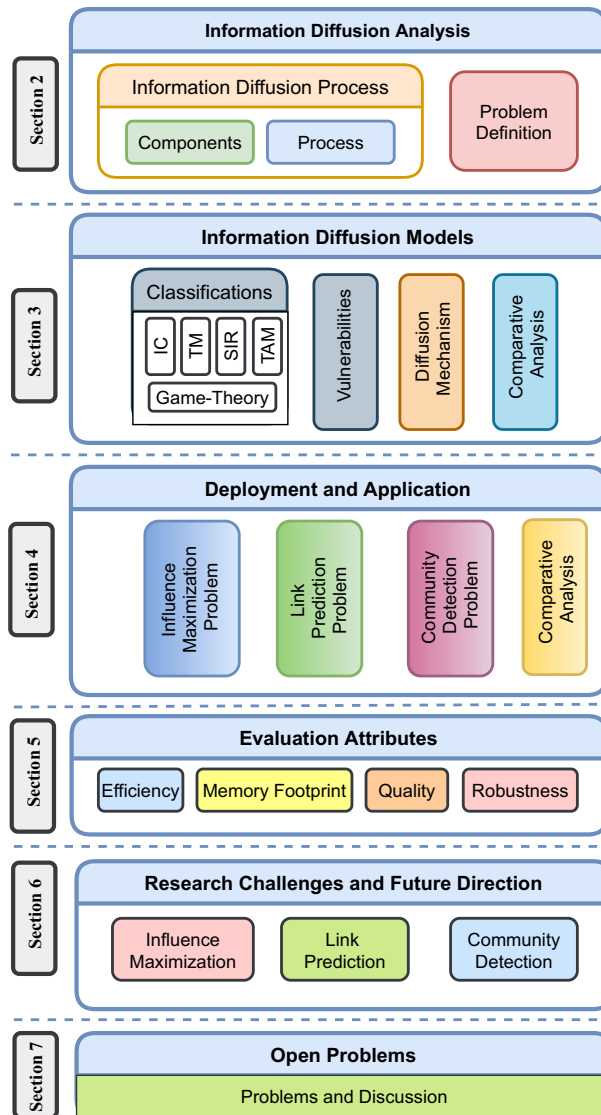
**Cite this article:** S. S. Singh, D. Srivastava, M. Verma and S. Muhuri. Information diffusion analysis: process, model, deployment, and application. *The Knowledge Engineering Review* 39(e11): 1–45. <https://doi.org/10.1017/S0269888924000109>

process through the available frameworks as sometimes it creates public issues, especially in cases like marketing and rumors related to emergencies (Razaque *et al.*, 2019). The diffusion of information in social platforms is like the spread of an infectious disease (Abdullah & Wu, 2011). The situation of diffusion can be related to the rule of one percent, which highlights that mostly 1% of the users contribute to the spread of the information largely, and the rest 99% of the users have a nominal contribution in the propagation of the information (Hargittai & Walejko, 2008). Therefore, it is important to bring forth the existing theories and analyze the advantages and disadvantages of the existing models. Along with this, identify an apt social network information interaction design that can solve the problems induced by the evolution and diffusion of information. The four stages that help in analyzing information diffusion are: information diffusion process modeling (Varshney *et al.*, 2014), influence evaluation (Varshney *et al.*, 2017), algorithm design of SNA problems, and identification of influential users (Sheshar *et al.*, 2021).

Digitalization has led to the generation of a tremendous amount of data through social networks because of user participation. This data contains a variety of information. People share details about the events happening in their daily lives, posting their opinions about breaking news, companies promoting their products by hiring influential users, etc. This way, the information propagates through the network and diffuses across the users (Chang *et al.*, 2018). The connections of such users can further re-share the posts, and the process of diffusion continues. Users who propagate the information and users who adopt the information are called influencers and influenced, respectively. However, the information diffusion mechanism is unknown but analyzing the process of diffusion is important from the perspective of real-life applications like social recommendation (Elsweiler *et al.*, 2007; Backstrom & Leskovec, 2011; Xu *et al.*, 2014), viral marketing (Richardson & Domingos, 2002; Kempe *et al.*, 2003; Leskovec, 2007), social recommendation (Elsweiler *et al.*, 2007; Backstrom & Leskovec, 2011; Xu *et al.*, 2014), social behavior prediction (Xu *et al.*, 2016; Zhao *et al.*, 2016; Ma *et al.*, 2017), and community detection (Fortunato 2007, 2010). Therefore, studying information diffusion analysis for SNA is vital for enhancing scholarly discourse in the field. A potential research gap could be explored, focusing on developing a comprehensive framework that brings together different elements of information diffusion. This framework would encompass the processes, models, deployment strategies, and practical applications of information diffusion. Despite the extensive research on information diffusion, there seems to be a lack of a comprehensive approach that effectively addresses all the different components in a coherent manner. This manuscript aims to bridge the gap by thoroughly exploring and synthesizing the elements mentioned, providing valuable insights for both theoretical understanding and practical applications in various contexts. This has attracted several researchers from sociology, computer science, etc., to study the mechanism and propose models to simulate and describe the process. Some of these models include the linear threshold (LT) model (Granovetter, 1978; Girvan & Newman, 2002; Richardson & Domingos, 2002; Kempe *et al.*, 2003; Leskovec, 2007; Fortunato, 2007, 2010; Xu *et al.*, 2016; Zhao *et al.*, 2016; Ma *et al.*, 2017), independent cascade (IC) model (GoldenbergJ, 2002), and epidemic models (Kermack & McKendrick, 1932). Most of the models assume that the diffusion of information starts from a set of source nodes called seed nodes, and the other nodes can access the same through their neighbors. The influencers control the spread of information across the network, leading to influence maximization.

This article presents a review that studies about information diffusion analysis and its applications in various fields. The organization and contribution of the paper are illustrated in Figure 1. The main contribution of the study is given as follows.

1. A brief discussion of information diffusion, component, and deployment is presented.
2. The comparative analysis of classical diffusion models, along with their vulnerabilities, are discussed.
3. Discussion of information diffusion analysis methods corresponding to different applications like community detection, influence maximization, and link prediction is presented.



**Figure 1.** The information diffusion analysis survey overview

- The performance evaluation measures, along with research challenges and future prospects regarding different applications, are detailed.
- The open problems are discussed, corresponding to information diffusion analysis.

### 1.1. Difference from existing surveys

Several surveys exist in the literature that deal with influence maximization, diffusion models, link prediction approaches, etc. Each such survey focuses on some key points.

- The authors in Das and Biswas (2021b) mainly focus on how information diffusion can be utilized for community detection. Different aspects have been covered in this survey like how the social facets and network properties are affected by the information flow, where the information

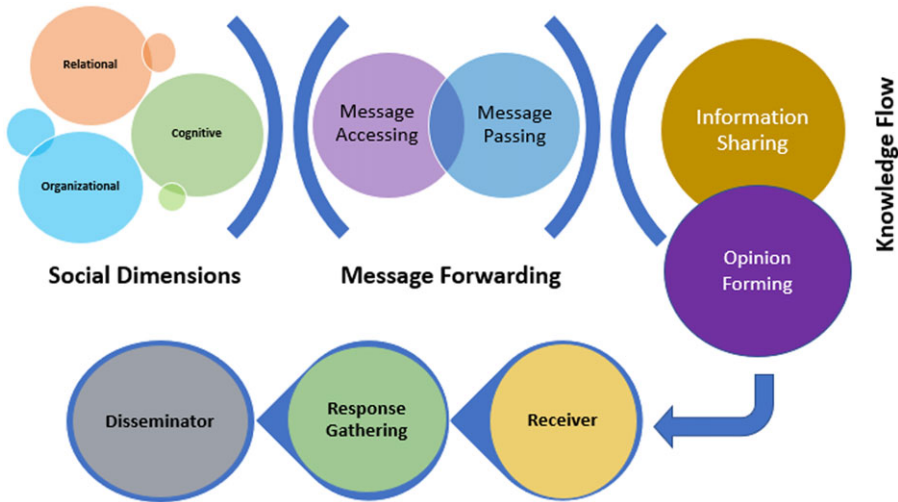
is generated and how it is propagated in the network has been explored, how the different community detection algorithms are impacted by the network parameters and the social facets has been studied and also the evaluation metrics for quality and accuracy measurement of communities has been discussed.

- In Kumar *et al.* (2020), a survey has been presented that explores the link prediction aspect of complex social networks. Several models have been discussed, including fuzzy models, network embedding methods, probabilistic methods, deep learning models, clustering-based models, etc. In this work, the aspect of different types of networks has also been considered like heterogeneous, temporal, and bipartite networks, and experimental results on real-world datasets have been discussed.
- The authors of Azaoui *et al.* (2021), presented a survey on influence maximization and their contributions include the exploration of influence maximization models based on node topology technique, the impact of privacy preservation and security on these models have been studied, a survey has been done for the group-based models, and finally the open challenges and future directions have been depicted to lay down the foundation for further research.
- Information diffusion models have been extensively studied and classified based on characteristics, and pertinent vulnerabilities, as well as threats, have been elaborated in Razaque *et al.* (2019). Applications and limitations of the models have been discussed, and recommendations have been provided for future improvement.
- In Sheshar *et al.* (2021), authors have presented a survey of information diffusion models as well as influence maximization algorithms for single/multiple networks. In present survey covers the theoretical analysis of influence maximization techniques and their respective frameworks. It also covers the context-aware influence maximization approaches and performance metrics related to the influence maximization algorithms; their comparison and challenges have also been explored and depicted.
- The authors of Singh *et al.* (2024, 2023) present surveys on social network analysis with respect to language information analysis, privacy, tools, SNA process, and application. They have also discussed the advantages and limitations of SNA techniques.

In this survey, Section 2 focuses on the information diffusion process, how it propagates through the network, affects the various parameters like network topology, social facets, etc., and impacts the various applications and users in the social networks. Section 3 discusses different types and aspects of the information diffusion models. Section 4 presents various research verticals such as influence maximization, community detection, and link prediction of SNA with respect to information diffusion. Section 5 discusses the evaluation metrics of SNA corresponding to various research verticals. Section 6 discusses research challenges and future directions, while Section 7 discusses open problems. Finally, Section 8 presents concluding remarks.

## 2. Information diffusion process

Gaining a deep understanding of how information spreads is crucial in our modern, interconnected world. The rapid dissemination of information has a profound impact on shaping opinions, influencing behaviors, and driving societal trends. Nevertheless, despite its importance, there is a noticeable lack of research in fully examining the complexities of the information diffusion process. The manuscript seeks to explore the complex aspects of information diffusion. This manuscript aims to explain the fundamental processes that facilitate the spread of information across different platforms and networks. This research aims to provide valuable insights for scholars, practitioners, and policymakers by identifying key factors that influence the diffusion process, including network structure, content characteristics, and user behavior. In addition, the manuscript aims to provide practical applications and deployment strategies to bridge the gap between theoretical understanding and real-world implementation. This will help facilitate more effective information dissemination strategies in various contexts. The information diffusion



**Figure 2.** The information diffusion process components (Razaque *et al.*, 2019)

process on social networks is affected by several factors. These factors can be analyzed considering three broad dimensions, namely cognitive dimension, structural dimension, and relational dimension:

- *Cognitive Dimension.* The emphasis of this dimension is on common vision and understanding among the receivers of information, similar preferences of the users, etc. These commonalities affect the thought process of the entire circle of connections and lead to the spread of information as well as the establishment of interpersonal relations. This gives a chance to the user to judge and assess the value of the information and accordingly consume it (Bhattacharjee & Sanford, 2006).
- *Organizational Dimension.* It consists of the network connections of an individual, flow of information, and resource pipeline access. It allows the users to access the information of their choice (Kim & Galliers, 2004).
- *Relational Dimension.* It considers the details of the stable and long-lived connections that can lead to shorter paths for effective acquisition of information and further assure quality diffusion of information (Rajamma *et al.*, 2011).

Further, concerning each of these dimensions, there are several components that affect the information diffusion and should be considered in the analysis of the process (Razaque *et al.*, 2019) as shown in Figure 2.

- *Message Access.* It consists of a platform that allows the users to access diffused messages, and this platform further forwards the messages that are received to the neighbors. Later, the message is received by all the nodes in the network, which increases the overhead of some needless messages but leads to a very vigorous distribution of the information (Khelil *et al.*, 2002).
- *Trust, Reciprocity and Cooperation.* In a social network, both the receiver and the disseminator are connected based on trust; sharing and reciprocation of information happens based on the status of personal relations, and sometimes, the users choose to cooperate among themselves to gain access to more messages and gather more information (Manapat & Rand, 2012).
- *Value.* In a network, the existence of multiple connections leads to excellent quality and a large amount of knowledge dissemination through information circulation. The recipients can further

evaluate the information based on the value and usage of information that solely depends on the needs, choices, and experiences of the users (Guille *et al.*, 2013).

- *Network Link.* The links in the networks allow the users to share their views, make the level of knowledge and values more balanced so that the information can be popularized, and the breadth of diffusion can be expanded so that more and more receivers can get access to it.
- *User Information.* It can be obtained from YouTube, social sites, blogs, etc. This information highlights the role of social connections in information gathering and how information diffusion impacts the choice of one user based on the preferences of other connections (Susarla *et al.*, 2012).
- *Individual Behavior.* Once the receiver gets access to the messages, they can filter out the useful information that matches their requirements, preferences, and values.
- *Feedback and Resource Sharing.* The information diffusion process gives a fair chance to the receiver to exchange views and feedback about the information with the disseminator and also share other resources (Greenhalgh *et al.*, 2008).

## 2.1. Information deployment

In social networks, communities' formation and detection depend on several factors like network attributes, temporal characteristics, and social attributes. The data stream in the organization influences every one of these variables contrastingly at various levels. In this manner, it is fascinating to see what the data stream means for local area recognition and informal communities. The elements influencing the informal communities can be extensively delegated (Das & Biswas, 2021b):

### 1. **Network Properties.** Topological and structural characteristics of the network.

- *Edge Strength Measures.* These measures, like connection strength, value strength, and tie strength, are used to assess the relationship between two vertices that share a common edge. The strength, however, can be measured by computing the number and the type of vertices impacted by the diffusion process.
- *Belongingness Measures.* For any vertex  $v$ , the belongingness in a specific local area  $c$  is controlled by having a place degree or having a place factor. Thinking about the viewpoint of data dissemination, the belongingness strength can be estimated by deciding the kind of vertices impacted during dispersion.
- *Centrality Measures.* The increasing popularity of social networks has led to an increase in the amount of information that is generated and made available to users. However, the ideas absorbed by the individuals highly depend on certain influential entities that can activate the inactive vertices through their influence. Such entities control the propagation of information in the network to a great extent and enhance the rate of diffusion. Some of the measures for identifying influential entities include betweenness centrality, degree centrality, and closeness centrality.
- *Clustering Coefficient.* During the time spent on data dissemination, the data is proliferated across the organization through the innate vertices that are available in the base conceivable length. Simultaneously, some topological properties of the organization are taken advantage of during the dissemination interaction. One such property is the grouping coefficient, which decides the degree to which the vertices bunch together and contribute fundamentally to local area identification.
- *Structural Equivalence.* Part of data is accessible in interpersonal organizations nowadays, yet while getting the data, the dependability of the data assumes a significant part for the client. Here impact makes an effect, that is, the data spread directed by the impact is more dependable and arrives at a bigger arrangement of vertices. Such an impact can be addressed by data not really set in stone from client credits, client demands, and other pertinent



elements. The features that can measure the information diffusion probability by utilizing appropriate similarity/dissimilarity evaluation techniques. Thus, these methods are used to evaluate the structural equivalence. Structural equivalence measures include the Jaccard coefficient, cosine similarity, and Euclidean distance.

- *Boundary Vertex.* Users in social networks connect with other users who are either similar or known to them. Hence, the information propagation in the network is either controlled by the similarity or by the topology. In terms of topology, the extent to which the neighboring nodes tend to connect. Concerning similarity, it is the user's attribute, interest, or characteristic that controls the connections. The diffusion of information in the network sometimes affects a pair of vertices differently as they may be exposed to two types of information that are different from each other. Such neighboring vertices of a vertex are referred to as boundary vertex.

## 2. **Social Facets.** It quantifies the various aspects of social interaction.

- *Contagion.* In social networks, following a particular trend, is a usual practice. Hence, any activity of a user is affected by the neighbors' actions. The inactive vertices get activated by the neighboring vertices during the information diffusion process. This influence of neighbors on the inactive vertex is called contagion. A higher degree of the vertex indicates higher chances of getting affected by the information propagated by the neighbors and vice versa.
- *Common Neighbors.* Formation of groups based on some similarity is a common characteristic of social networks. These groups exchange information during the process of diffusion. It is observed that there is a strong relationship among those users who have a huge number of common neighbors. Therefore, a dense sub-graph can depict the diffusion outcomes in such cases, and the community detection algorithms can utilize this facet to recognize inherent communities.
- *Topic.* Interest in the group and the dynamics of the diffusion network highly affect the speed at which the diffusion of a topic happens in a social network. For example, a politically inclined group will be affected more by a political campaign and will not be bothered about the advertisement of trekking equipment. A high diffusion speed indicates a higher affinity for that topic in the underlying network.
- *Homophily.* The users in social networks tend to connect with like-minded or similar users. This similarity is termed homophily which is exploited during the information diffusion process. The interaction pattern between similar users can be modeled through the independent cascade model. It is assumed that the diffusion initiated based on homophily affects many homophilic connections.
- *Social Interaction Channels.* Nowadays, users spend quite some time on social networks to interact with their peers and friends. Much information exchange happens on social networks, which opens up a kind of interaction channel through which people staying at physically different locations can also interact. This enhances and broadens the information diffusion process and increases the efficiency of the interaction channels.
- *Influential Spreaders.* Certain entities are popular enough and have one too many relations in the network. Such users are referred to as influential spreaders as they can initiate information diffusion in social networks by activating inactive users. However, the strength of information propagation depends upon the data (whether it meets the user's requirement) and the popularity of the spreader in the network. The higher the impact of the influential spreader, the higher the probability of diffusion.

### 3. Information diffusion models

This manuscript highlights the significance of information diffusion models in understanding SNA analysis. These models are extremely useful for researchers, allowing them to analyze how information spreads across different platforms and communities. By understanding the inner workings and patterns of information spread, these models help in creating better strategies for communication, marketing, and policy-making. In addition, they enable individuals and organizations to navigate the intricacies of the digital era, promoting well-informed decision-making and strengthening societal resilience. Therefore, the investigation and improvement of information diffusion models are crucial pursuits that contribute to a deeper comprehension of our information-driven world. Information diffusion models have an essential role in several real-life applications, including sociology, ethnography, and epidemiology. The main components of the information diffusion model are disseminator, resource sharing, receiver, and opinion formation (Sheshar *et al.*, 2021). We have discussed the classical diffusion models in the following section and summarized the characteristics in Table 1.

- **Threshold Model (TM).** In the TM, or then again set of edge esteems, is utilized to separate scopes of qualities for the conduct anticipated by the model (Schelling, 2006). Straight limit model (LTM) is the most famous edge dispersion The linear threshold model (LTM) is one of the most popular threshold diffusion models where a node becomes active if the influence of the neighbors is greater than a defined threshold. In LTM, the threshold always follows uniform distribution over  $[0,1]$ . One of the modified TMs is a linear threshold with color that considers the client's involvement in an item and catches item reception instead of affecting (Bhagat *et al.*, 2012). Direct TM is additionally altered to deal with the assessment change of clients and permits them to change among dynamic and latent states (Pathak *et al.*, 2010).
- **Cascade Model (CM).** In CM, if a node becomes active at one point in time, it has the same probability of activating its inactive neighbors at the next timestamp (Goldenberg *et al.*, 2001). As the cascading process continues, an already active node will never become inactive in the future. The other variations of independent CM (ICM) are ICM with negative opinion (Chen *et al.*, 2011) or ICM with positive and negative opinion (Nazemian & Taghiyareh, 2012).
- **Time-aware Model (TAM).** In some engendering models, the spread of social impact has been amplified explicitly to a decent time frame. These are known as TAM, where the intermingling of proliferation relies upon the period rather than the number of cycles (Chen *et al.*, 2012). The model is acquainted with accomplishing time-basic interest. TM can be isolated into two sub-classes which are a discrete-time model and a ceaseless time model (Rodriguez *et al.*, 2011).
- **Triggering Model (TRM).** TRM is the blend of TM and CM. In the TRM, every hub is related to an edge esteem and an appropriation work. A selection of its neighbors is guided by the characterized work with a similar likelihood (Kempe *et al.*, 2003). This model freely chooses an irregular subset of its neighbors in each occurrence of the dissemination interaction (Kempe *et al.*, 2005).
- **Epidemic Model (EM).** The decent individuals can be divided into three classes in an epidemic model: helpless (S), irresistible (I), and recuperated(R). Three plague models are generally utilized which incorporate defenseless irresistible recuperated (SIR) (Schütz *et al.*, 2008), powerless irresistible helpless (SIS) (Parshani *et al.*, 2010), and vulnerable irresistible recuperated defenseless (SIRS) (Magal & Ruan, 2014). This scourge model proliferates the data dependent on the strength of the neighbors.
- **Game Theory Model (GTM).** In GTM, individual conduct systems are depicted in harmony by examining the practices and advantages of every member (Easley *et al.*, 2012; Jiang *et al.*, 2014). In this model, the players make a choice and pay off to every player dependent on the choice made by all of the playing players (Muhuri *et al.*, 2018).



**Table 1.** The comparison of the diffusion models characteristics (Singh et al., 2019; Singh et al., 2021), where Ma=Multiple Activation, Ta=Time-specific Activation, Di=Diminishing Returns, Mo=Monotone

lightgray Model	Activation condition	Activation		Submodular		Application
		Ma	Ta	Di	Mo	
Cascade Model (CM)						
Independent CM (Kempe <i>et al.</i> , 2003)	$\prod_{i=1}^k (1 - P_y(v_i S \cup M_i))$ $= \prod_{i=1}^k (1 - P_y(v_i S \cup M_i))$	✗	✗	✓	✓	Viral Marketing, Collective behavior
Opinion CM (Zhang <i>et al.</i> , 2013)	$\sum_{x \in N^a(y)} W_{xy} \geq T_y$	✓	✓	✗	✗	Political campaign
ICM-NO (Chen <i>et al.</i> , 2011)	$\prod_{i=1}^k (1 - P_y(v_i S \cup M_i))$ $= \prod_{i=1}^k (1 - P_y(v_i S \cup M_i))$	✗	✗	✓	✓	Political campaign
Threshold Model (TM)						
Linear TM (Kermack & McKendrick, 1927)	$\sum_{x \in N^a(y)} W_{xy} \geq T_y$	✗	✗	✓	✓	Rumors & diseases control
Majority TM (Richardson & Domingos, 2002)	$T_y = \frac{1}{2}D(y)$	✗	✓	✓	✓	Voting system, Distributed computing
Small TM (Eiselt & Laporte, 1989)	$\sum_{x \in N^a(y)} W_{xy} \geq T_y$	✗	✗	✓	✓	Network security & vulnerability
Unanimous TM (Richardson & Domingos, 2002)	$\sum_{x \in N^a(y)} W_{xy} \geq T_y$	✗	✗	✗	✗	
Separated TM (Kermack & McKendrick, 1927)	$\sum_{x \in N^a(y) \cap \varphi_A^{t-1}} w_{x,y}^A \geq T_y^A$	✗	✗	✗	✗	Competitive marketing
Weighted-proportional TM (Kermack & McKendrick, 1927)	$P[y \in \varphi_A^t   y \in \varphi^t \setminus \varphi^{t-1}] = \frac{\sum_{x \in \varphi_A^t} w_{x,y}}{\sum_{x \in \varphi^t} w_{x,y}}$	✗	✗	✗	✗	Competitive marketing
Time-aware Model (TAM)						
DTAM (Chen <i>et al.</i> , 2011; Chen <i>et al.</i> , 2012; Lee <i>et al.</i> , 2012)	–	✓	✓	✗	✓	Viral marketing
CTAM (Gomez-Rodriguez <i>et al.</i> , 2011; Lee <i>et al.</i> , 2012)	–	✓	✓	✗	✓	Viral marketing
Triggering model						
Decreasing CM (Kempe <i>et al.</i> , 2005)	$P_y(x S) \leq P_y(x M)$	✗	✗	✓	✓	Collective behavior
Epidemic model						
SIR (Kermack & McKendrick, 1927, 1991)	$\frac{dS}{dt} = -\beta SI, \frac{dI}{dt} = \beta SI - \gamma I, \frac{dR}{dt} = \gamma I$	✗	✗	✗	✓	Epidemiology
SIS (Kermack & McKendrick 1927, 1991)	$\frac{dS}{dt} = -\beta SI + \gamma I, \frac{dI}{dt} = \beta SI - \gamma I$	✓	✗	✗	✓	Epidemiology
SIRS (Kermack & McKendrick 1927, 1991)	$\frac{dS}{dt} = -\beta SI + fI,$ $\frac{dI}{dt} = \beta SI - \gamma I, \frac{dR}{dt} = \gamma I - fR$	✓	✗	✗	✓	Epidemiology

### 3.1. Information diffusion vulnerabilities

Information diffusion models can face many vulnerabilities if the common weaknesses are uncontrolled. As with the consistent advancement of networks over organizations, social media, etc., online protection has become a significant subject. Vulnerabilities of Information Diffusion mainly depend upon the model applied for link prediction. Some of the common vulnerabilities of Information Diffusion models are discussed below:

- **Node Paralysis.** A lot of exploration is directed at node paralysis of motion in various models proposed in state-of-the-art approaches (Razaque *et al.*, 2019). A one-dimensional model begins with a random initial state. The intermittent limit conditions are dependent on different models. The likelihood of the resistance is 0.5, a big part of the endorsement and half resistance. Under various boundaries, the likelihood of developing the last attractors of ferromagnetism is likewise unique. The likelihood that the attractor gives up at a similar starting thickness fluctuates with the collection file. Under a similar collection file, the more noteworthy the likelihood of a similar beginning thickness keeps an eye on the attractor when the association likelihood becomes bigger; the expansion of the association likelihood will make the framework bound to look steady. When the accumulation record expands, something similar starting thickness will generally be more noteworthy in the likelihood of the attractor (i.e., the framework is bound to look reliable because of the expansion in the number of short-range associations). Such different circumstances prompt Vulnerabilities.
- **Continuously Changing.:** The organization is in an implicit, persistent way. The method involved with adding additional associations proceeds until the number of huge distance affiliations and the number of short distance affiliations (i.e., affiliation probability) shows up at a particular worth. Like this, the typical neighbor number of each center point in the little world organization is  $4(1 + x)$ . At the point when the amassing record ( $T$ ) reaches 0, the theory of the little world organization has transformed into a little world association. Regardless, with  $T_1$ , there are incredibly close center points. Such extraordinary minimal world associations are relative to standard associations. At the point when  $0 \leq T \leq 3$ , because of the characteristics of extra-critical distance affiliations, this reach has rich characteristics. In any case, it creates a shortcoming that can be the justification for the attack (Razaque *et al.*, 2019).
- **Instability.** Instability in the model significantly reduces the efficiency and performance of the model. Given the Instability of network geography, deciding the impact of joining and leaving the host is essential, particularly in a P2P organization. To more readily comprehend the effect of entering and leaving on the transmission of dynamic worms, the proportion of the pestilence and irresistible infection boundaries ought to be utilized. The aftereffects of several models show that the expansion and flight of hosts significantly affect the notoriety and spread execution of the topology-aware dynamic worms (Razaque *et al.*, 2019).
- **Harder Node Evaluation.** To control the diffusion of viruses, certain models combine the minimum traffic with the time Markov chain features. One such model is the SIR model, wherein the infection and its removal can be described through this vulnerability. The traffic flow is affected in the congestion and free flow phases by the viral transmission, and to assess its dynamics, models like SIR may be used. The goal is to identify the effect on the traffic conditions and dynamics of the system due to virus diffusion, that is, in the congestion and free flow stages (Razaque *et al.*, 2019).
- **Limited Size.** In social networks, limited size is one of the vulnerabilities. It can be studied considering the network construction. The complex process of the experimental methods leads to large size. Also, the data sets that are compiled are of varying and considerable sizes. The data that is obtained from multiple sources have sufficient subjective biases which is not compatible with existing social network size. However, this issue can be tackled by using the affiliation network as performed in the SIR model to study the state of the nodes in the network.

- **Data Reduction.** In Serban *et al.* (2016), the authors have suggested that, in general, the distance between the source of every estimator and its closest tree network is constant. The number of sources can be estimated using the proposed algorithm, even if the real and existing sources are unknown. Collisions between the wireless intermediate nodes and transmission signals lead to a reduction in the throughput of end-to-end data transmission. Therefore, in certain models like SIR, there are increased chances of inviting vulnerabilities due to this reduction.
- **Abhorrent Types.** It is difficult to assess the authenticity and truth value of the information flowing through social networks. When the information arrives at a node, it is the choice of the netizen whether he/she wants to forward it or not. Similarly, sending real or distorted information to the disseminator is also the choice of the netizen. The multi-agent and game theory concepts can well describe both situations, and it is utilized in the background to study the information dissemination in the networks. The model of Anshuman *et al.* (Chhabra *et al.*, 2017) can be explored to identify the communication behavior of online users connected through social networks.
- **Independence of Values.** Based on the user's awareness of security, their behavior is affected. The quantifying factors for these behavioral changes have been analyzed by studying the propagation mechanism of the worms in the social networks (Seddiki & Frikha, 2012). The information transmission for multi-layer online networks follows the mechanism of information dissemination in social networks. This model is based on the idea of the evolutionary game model that is used in the information diffusion process established between government bodies and netizens. For example, the strategies executed by the government to control the harmful network group events that negatively impact the social order have been proven to be very effective.

#### 4. Information diffusion deployment and application

Online engagement and interaction of users over social networks have attracted researchers to solve real-world problems such as rumor control, viral marketing, recruitment, and social recommendation by utilizing information flow analysis. The topological information and social facets have been used to analyze the information flow over the network. Therefore, researchers in recent years have shifted focus to information diffusion analysis for finding solutions for influence maximization, community detection, and link prediction. This section is devoted to analyzing the role of information dissemination for link prediction, influence maximization, and community detection problem-solving.

##### 4.1. Influence maximization

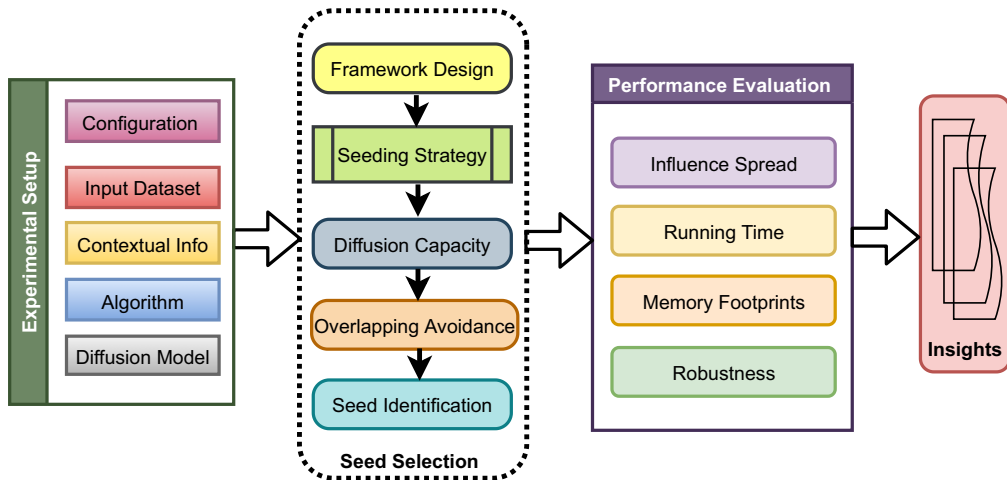
Identification of influential word-of-mouth promoters is a challenge in influence maximization (IM), which aims to increase product uptake. Kempe *et al.* (2003) was the first to present an optimization model for the IM problem under the linear threshold and independent cascade models of classical diffusion. They have proposed a simulation-based greedy solution to identify seed users with an approximation guarantee  $1 - 1/e - \epsilon$ . The greedy algorithm utilizes the information diffusion capacity of each individual in the network to evaluate the user's influence spread. The algorithm finds influential seeds iteratively to avoid overlapping avoidance. The main limitation of the greedy solution is its time efficiency due to the time-consuming Monte Carlo (MC) simulation. It is because of the stochastic nature of the diffusion process that a new set of users is activated at each iteration. The influence maximization algorithms have been categorized based on the context of the product, and network (Singh *et al.*, 2021) as shown in Table 2. The basic framework of an influence maximization algorithm is depicted in Figure 3 (Arora *et al.*, 2017; Singh *et al.*, 2021). The first component experimental setup of the IM framework configures the diffusion process with influence and activation probabilities. The seed selection process identifies the seed users corresponding to the seeding strategy and diffusion model.

**Table 2.** Illustration of information diffusion analysis over different applications (Kumar et al., 2020; Singh et al., 2021; Das & Biswas, 2021b)

	Application	Categories	Problem solving perspective	Representative
Information diffusion analysis	Influence maximization	Classical	Simulation-based	Leskovec <i>et al.</i> (2007), Goyal <i>et al.</i> (2011a)
		IM2	Heuristic-based	Freeman (1978), Page <i>et al.</i> (1999), Liu <i>et al.</i> (2014)
			Mixed approaches	Chen <i>et al.</i> (2009), Cheng <i>et al.</i> (2013)
			Simulation-based	Saito <i>et al.</i> (2012), Zhang (2015), Zhang <i>et al.</i> (2016)
			Heuristic-based	Erlandsson <i>et al.</i> (2017), Wang <i>et al.</i> (2016), Zhang (2015)
		MIM	Mixed Approaches	Wang <i>et al.</i> (2016)
			Simulation-based	Sun <i>et al.</i> (2016)
			Heuristic-based	Singh <i>et al.</i> (2019b )
		Context-aware	Spatial	Li <i>et al.</i> (2014), Zhou <i>et al.</i> (2015)
			Topical	Singh <i>et al.</i> (2019a)
			Time	Chen <i>et al.</i> (2012), Lee <i>et al.</i> (2012), Liu <i>et al.</i> (2014)
			Competitive	Carnes <i>et al.</i> (2007), Bharathi <i>et al.</i> (2007), Borodin <i>et al.</i> (2010)
			Dynamic	Aggarwal <i>et al.</i> (2012), Zhuang <i>et al.</i> (2013), Ohsaka <i>et al.</i> (2016), Meng <i>et al.</i> (2019)
		Profit maximization	OSN Providers	Tang <i>et al.</i> (2018), Lu & Lakshmanan (2012), Tang <i>et al.</i> (2017)
	Revenue maximization	Discounting	Han <i>et al.</i> (2018), Khan <i>et al.</i> (2016)	
Link prediction	Similarity based	Local similarity indices	Kumar <i>et al.</i> (2020), Li <i>et al.</i> (2018), Yao <i>et al.</i> (2018), Divakaran & Mohan (2020)	

Table 2. (Continued)

Application	Categories	Problem solving perspective	Representative
Community detection	Probabilistic and Maximum likelihood	Global similarity indices	Chen <i>et al.</i> (2018), Bliss <i>et al.</i> (2014)
		Hierarchical structure model	Clauset <i>et al.</i> (2008)
		Stochastic block model	Wang <i>et al.</i> (2015), Stanley <i>et al.</i> (2019)
	Dimensionality Reduction	Parametric model	Kashima & Abe (2006)
		Non-parametric model	Williamson (2016)
		Factor graph model	Loeliger <i>et al.</i> (2007)
		Embedding-based link prediction	Cao <i>et al.</i> (2018)
		Matrix factorization/decomposition based	Koren <i>et al.</i> (2009)
	Information diffusion	Independent Cascade	Saito <i>et al.</i> (2008)
		Threshold Model Time-Aware Model	Braun (1995) Daneshmand <i>et al.</i> (2015)
	Classical community detection	Network structure	Shen <i>et al.</i> (2010), Das and Biswas (2021a)
		Similarity among nodes Decentralized	Liu and Wang (2016) Chen (2011), Yang <i>et al.</i> (2013), Bhattacharya & Sarkar (2021)
	Network centrality	Centrality measures Network decomposition	Wang <i>et al.</i> (2015), Bo <i>et al.</i> (2014), Wu <i>et al.</i> (2015) Li <i>et al.</i> (2015), Devi & Tripathi (2020)
	Overlapping Community	Cascade Topological	He <i>et al.</i> (2021), Sun <i>et al.</i> (2018), Wang <i>et al.</i> (2021) Sun <i>et al.</i> (2020), Ramezani <i>et al.</i> (2018), Alviri <i>et al.</i> (2014)
	Stochastic Optimization	Neighbourhood selection Community optimization	Hajibagheri <i>et al.</i> (2012); Hajibagheri <i>et al.</i> (2013) Obregon <i>et al.</i> (2019), Mozafari & Hamzeh (2015), Kalantari <i>et al.</i> (2020)



**Figure 3.** The influence maximization framework under information diffusion model (Singh et al., 2021)

Finally, the algorithm's performance is evaluated regarding influence spread and efficiency. The characteristic summary of the influence maximization method is present in Table 2.

#### 4.1.1. Classical influence maximization

The authors of Sviridenko (2004) modified the IM problem by introducing individual node price constraints, which was unitary in Kempe et al. (2003). The greedy solution is not scalable and computationally efficient for large-scale networks. Therefore, much research is devoted to computing effective and efficient solutions for the IM problem. Three types of solutions exist in literature to tackle classical IM problems; simulation-based, heuristics, and mixed approaches. The heuristic-based methods do not guarantee approximation while simulation and mixed approaches have. By utilizing the submodular property to reduce the amount of MC simulations on the greedy algorithm, Leskovec et al. (2007) offer a simulation-based method known as cost-effective lazy forward (CELf). The influence evaluation function under IC and LT models is submodular as proved by Kempe et al. (2003). Therefore, Leskovec et al. applied submodularity to the greedy algorithm for iterative computation of seed nodes based on individual influence, significantly decreasing the number of MC simulations. The CELf is almost 700 times more efficient than the greedy method. The further optimization of the CELf method is presented by Goyal et al. (2011a) and is known as CELf++. CELf++ utilizes the marginal influence gain of each individual in previous and current iterations to avoid unnecessary MC simulations. This method also uses traditional diffusion models LT and IC to estimate the marginal influence gain of an individual. CELf++ is 30–50% faster than CELf.

Some works focused on improving the time complexity of MC simulations rather than reducing MC simulations. The authors of Wang et al. (2010) present a community-inspired greedy solution known as CGA. This algorithm divides the network into subnetworks to identify seed users without overlapping influence. Therefore, it significantly reduces the search space, which improves the time complexity of MC simulations. However, because of the complexity associated with community identification, CGA is not ideal for large-scale networks. Because there are more MC simulations, it has been shown that simulation-based methods are not practical for large-scale networks. By giving each person an approximate score, certain heuristic approaches have been developed to increase scalability. This scoring or ranking is done by features like information diffusion, centrality, topological information, etc., to identify seed nodes. There are some efforts (Page et al., 1999; Jung et al., 2012; Liu et al., 2014) have been made in this direction.



**Table 3.** The comparison of influence maximization algorithm based on information dissemination model (Singh et al., 2021)

Algorithm	Diffusion models				Approximation		Solving perspective		
	LT	IC	TRM	TAM	Ratio	Time complexity	Simulation	Heuristic	Mixed
Greedy (Kempe et al. 2003)	✓	✓	✓	✓	$1 - 1/e - \epsilon$	$O(kNMI)$	✓	✗	✗
Knapsack Greedy (Sviridenko 2004)	✓	✓	✓	✓	$1 - 1/e - \epsilon$	$O(N^5)$	✓	✗	✗
Diffusion Degree (Kundu et al. 2011)	✗	✓	✗	✗	NA	$O(N + M)$	✗	✓	✗
CELF (Leskovec et al. 2007)	✓	✓	✓	✓	$1 - 1/e - \epsilon$	$O(kNMI)$	✓	✗	✗
NewGreedy (Chen et al. 2009)	✗	✓	✗	✗	$1 - 1/e - \epsilon(r)$	$O(kIM)$	✗	✗	✓
MIA/PMIA Chen et al. (2010)	✓	✓	✓	✓	$1 - 1/e$	$O(Nt_{i\theta} + kn_{o\theta}n_{i\theta}(n_{i\theta} + \log N))$	✗	✓	✗
LDAG (Chen et al. 2010)	✓	✗	✗	✗	NA	$O(Nt_{i\theta} + kn_{o\theta}m_{i\theta}(m_{i\theta} + \log N))$	✗	✓	✗
CELF++ (Goyal et al. 2011a)	✓	✓	✓	✓	$1 - 1/e - \epsilon$	$O(kNMI)$	✓	✗	✗
SA (Jiang et al. 2011)	✗	✓	✗	✗	NA	$O(TIM)$	✓	✗	✗
SIMPAT (Goyal et al. 2011b)	✓	✗	✗	✗	NA	$O(kINP_{\theta})$	✗	✓	✗
IPA (Kim et al. 2013)	✓	✗	✗	✗	NA	$O(\frac{NO_vn_{uu}}{c} + k^2(\frac{O_vn_{uu}}{c} + (c - 1)))$	✓	✗	✗
StaticGreedy (Cheng et al. 2013)	✗	✓	✗	✗	$1 - 1/e - \epsilon$	$O(\frac{kMN^2 \log(\frac{N}{k})}{\epsilon^2})$	✗	✗	✓
PRUNEDMC (Ohsaka et al. 2014)	✗	✓	✗	✗	$1 - 1/e - \epsilon$	$O(\frac{kMN^2 \log(\frac{N}{k})}{\epsilon^2})$	✗	✗	✓
GROUP-PR (Liu et al. 2014)	✗	✓	✗	✗	NA	$O(kMN)$	✗	✓	✗
RIS (Borgs et al. 2014)	✓	✓	✓	✓	$1 - 1/e - \epsilon$	$O(\frac{k(N+M) \log^2 N}{\epsilon^2})$	✗	✗	✓
IMRANK Cheng et al. (2014)	✗	✓	✗	✗	NA	$O(NTd_{max} \log d_{max})$	✗	✓	✗
SKIM(Cohen et al. 2014)	✗	✓	✗	✗	$1 - 1/e - \epsilon$	$O(\frac{kN^2 M \log(\frac{N}{k})}{\epsilon^2})$	✗	✗	✓
IMM (Tang et al. 2015)	✓	✓	✓	✓	$1 - 1/e - \epsilon$	$O(\frac{(k+l)(N+M) \log N}{\epsilon^2})$	✗	✗	✓
UBLF (Zhou et al. 2015)	✓	✓	✗	✗	$1 - 1/e - \epsilon(I)$	$O(kINM)$	✓	✗	✗
TIM (Tang et al. 2014)	✓	✓	✓	✗	$1 - 1/e - \epsilon$	$O(\frac{k(M+N) \log N}{\epsilon^2})$	✗	✗	✓
Degree Discount (Chen et al. 2009)	✓	✓	✓	✓	NA	$O(k \log N + M)$	✗	✓	✗

Table 3. (Continued)

Algorithm	Diffusion models				Approximation		Solving perspective		
	LT	IC	TRM	TAM	Ratio	Time complexity	Simulation	Heuristic	Mixed
EASYIM (Galhotra <i>et al.</i> 2016)	✓	✓	✗	✗	NA	$O(kD(N + M))$	✗	✓	✗
SSA/D-SSA (Nguyen <i>et al.</i> 2016)	✓	✓	✓	✓	NA	–	✗	✗	✓
BKRIS (Wang <i>et al.</i> 2017)	✓	✓	✗	✗	$1 - \frac{1}{e} - \epsilon - \epsilon'$	$O(\frac{NM(\log N + \log \binom{N}{k})}{\epsilon^2})$	✗	✗	✓
SP1M (Kimura & Saito 2006)	✗	✓	✗	✗	$1 - 1/e$	$O(kNM)$	✗	✓	✗
DPSO (Gong <i>et al.</i> 2016; Wang <i>et al.</i> 2017)	✓	✓	✓	✓	NA	$O(k^2 \log kn\bar{D}^2)$	✗	✗	✓
IRIE (Jung <i>et al.</i> 2012)	✗	✓	✗	✗	NA	$O(k(n_{\theta}k + M))$	✗	✓	✗
LAIM (Ge <i>et al.</i> 2017)	✓	✓	✓	✓	$1 - 1/e - \epsilon$	–	✓	✗	✗
TW Greedy (Wang & Feng 2009)	✓	✓	✓	✓	$1 - 1/e - \epsilon$	$O(kNMI)$	✓	✗	✗
Cost-Degree (Yang <i>et al.</i> 2018)	✗	✓	✗	✗	NA	$O(M)$	✗	✓	✗
LAPSO-IM (Singh <i>et al.</i> 2019b)	✓	✓	✓	✓	NA	$O(Ink(\log k + \bar{D}^2) + nkN)$	✗	✗	✓
CGA (Wang <i>et al.</i> 2010)	✗	✓	✗	✗	$1 - e^{-1/(1+\delta_C)}$	$O(M + IM_C(N(Z - C) + k(C + N_C)))$	✓	✗	✗
IM-SSO (Singh <i>et al.</i> 2020)	✓	✓	✓	✓	NA	$O(n(k\bar{D}^2 + n) + n \log n)$	✗	✗	✓
ACO-IM (Singh <i>et al.</i> 2019a)	✓	✓	✓	✓	NA	$O(I_{\max} R_G  V  +  V  \log  V )$	✗	✗	✓
LCI (Zhang <i>et al.</i> 2016)	✓	✓	✗	✗	NA	$O((N + M)N.d)$	✓	✗	✗
MPMN-SIMPATH (Zhang 2015)	✓	✗	✗	✗	NA	$O(k \ln P_{\theta})$	✗	✓	✗
ASMTc (Wang <i>et al.</i> 2016)	✗	✓	✗	✗	NA	$O( V^s ^2 +  V^s )$	✗	✗	✓
MIM-Greedy (Sun <i>et al.</i> 2016)	✓	✓	✗	✗	$1 - 1/e$	$O(kmNMI)$	✓	✗	✗
MIM2 (Singh <i>et al.</i> 2019b)	✓	✓	✓	✓	NA	$O((l + m)(N + M) + (k + m)(N \log N + M))$	✗	✓	✗
C2IM Singh <i>et al.</i> (2019a)	✓	✓	✗	✗	NA	–	✗	✗	✓

#### 4.1.2. Influence maximization across multiple networks (IM2)

The classical IM methods consider a single type of relationship and networks to estimate users' influence while avoiding a scenario where users engage in multiple networks simultaneously. Considering simultaneous engagement of users across networks leads to a new framework of IM known as influence maximization across multiple networks (IM2). The authors of Zhang (2015) introduce a multi-phase multilayer network-based framework to identify influential users by utilizing SIMPATH (Goyal *et al.*, 2011b) and CELF++ (Goyal *et al.*, 2011a) algorithms. Zhang *et al.* adopt traditional diffusion models IC and LT to propagate influence across layers with different influence probability assignments. A new modified greedy algorithm is presented for the IM2 framework by utilizing lossless and lossy coupling strategies (Nguyen *et al.* 2013; Zhang *et al.*, 2016). The authors of Wang *et al.* (2016) introduced an IM algorithm known as agent selection problem (ASP) under IM2 and distributed settings. ASP considers a multilayer framework for seed selection in mobile ad-hoc networks.

#### 4.1.3. Multiple Influence Maximization (MIM)

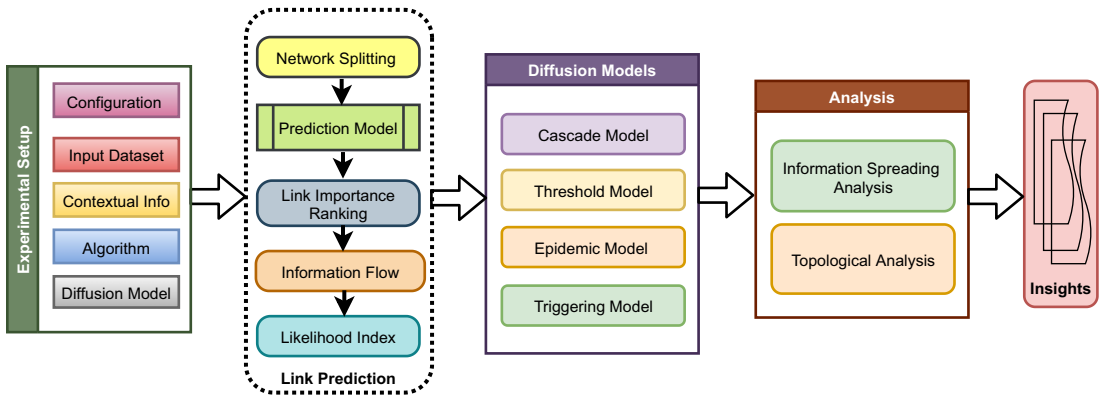
Most of the literature focuses on the IM scenario with only one product for marketing. However, some work considers competitive product marketing but not for non-competitive distinct product adoption. Based on the presumption that people have the purchasing power to accept many products at once, Sun *et al.* (2016) have presented a MIM greedy algorithm. They improved the time efficiency of greedy under the IC diffusion model by using the influence evaluation function's submodular characteristic. MIM framework first constructs different product diffusion graphs by the IC model. Then it identifies influential users for each product corresponding to product diffusion graphs. Finally, select seed nodes across diffusion graphs iteratively. The MIM greedy ensures  $(1 - \frac{1}{e})$  approximation guarantee.

#### 4.1.4. Multiple Influence Maximization across multiple networks (MIM2)

The MIM2 framework considers multiple non-competitive products and multiple relationship networks simultaneously to maximize product adoption. Therefore, the product diffusion graph construction and network coupling will be the key functionality under this framework. Singh *et al.* (2019b) were the first to introduce a heuristic-based MIM2 algorithm to identify seed nodes by back-propagation. MIM2 first performs network coupling based on overlapping users and topological structures. Then it creates product diffusion graphs corresponding to each product from the coupled multiplex network. The following influential users are identified for each diffusion graph. Finally, seed nodes are selected across different products to maximize product adoption under traditional diffusion models.

#### 4.1.5. Context-aware influence maximization

Most of the conventional IM approaches focus on the topological information of the network and avoid some critical factors like product information, target audience, contextual and semantic information, and geographical information. It was observed that conventional IM approaches are ineffective in real-world scenarios. Therefore, some researchers started focusing on context-aware IM solutions in recent years to improve the effectiveness of seed users. Some of the topic-relevant approaches such as LGA/ELGA (Guo *et al.*, 2013), IMIP/IMAX (Lee & Chung, 2015), CTVM (Nguyen *et al.*, 2016), MFIP (Li *et al.*, 2020) focus on the users interest to a different topic for the diffusion process. This will help to identify seed users from product-specific audiences. The authors of Singh *et al.* (2019a) proposed a context-aware community-based influence maximization (C2IM) algorithm that utilizes the topic-dependent contextual diffusion models to propagate influence. C2IM considers community and topical information to improve the efficiency and effectiveness of the seeding strategy. The seed strategy under C2IM estimates the diffusion degree by back-propagation of influence from non-desirable nodes. The diffusion degree works as a ranking method for seed selection.



**Figure 4.** The link prediction framework under information diffusion model (Kumar et al., 2020)

#### 4.1.6. Profit maximization

Profit maximization (PM) extends the classical IM problem, which considers actual product adoption rather than social influence. Profit maximization integrates the costs of advertisers and network service providers. There are some efforts (Lu & Lakshmanan, 2012; Tang et al., 2017; Li et al. 2017; Tang et al., 2018; Weersink & Fulton, 2020) have been made in this direction. The authors of Lu and Lakshmanan (2012) utilize the extension of traditional diffusion models to estimate actual product adoption over social influence. The traditional diffusion models IC and LT are modified by adopting new states and active and inactive states. These states help to compute actual product adoption, which drives profit and revenue computation. The authors also proved that the profit function is submodular under these diffusion models.

#### 4.2. Link prediction

Link prediction (LP) (Singh et al., 2022; Mishra et al., 2022; Singh et al., 2024) licenses to gather absent or future associations in an organization are termed Link Prediction. The organization association characterizes how data spreads through the hubs. Thus, the spreading might actuate changes in the associations and accelerate the organization's development. The development of web-based media has drawn little consideration from specialists and organizations. New stages are ceaselessly arising, for example, Facebook and Flickr (2004), YouTube (2005), Twitter (2006), and Miniature blog (2009), among others. Given the significance of various spaces and regions, research points, for example, Connection Prediction (LP). Furthermore, data diffusion has gotten considerable consideration in perplexing social network regions during the last few years. Notwithstanding, they are points, for the most part, examined in isolation, even though their outcomes are applied in comparative areas, such as viral promoting, political missions, and business process displaying.

The issue of suggesting joins has a few applications, such as recommending absent and likely associations in cloud data or, on the other hand, robust hub identification. Specifically, the expectation of future connections is helpful in comprehending the organization and correspondence evolution. For instance, in web-based media stages, promising associations that do not exist yet can advance commitment and collaboration among users, influencing the organization structure. Thus, the organizational structure impacts the correspondence or the spread of data. The link prediction working framework is shown in Figure 4.

LP techniques gauge the new edges as indicated by some association systems, similar to the distance and briefest ways among hubs, the triangles or Triassic conclusion, and the likeness with shared neighbors, among others. These primary components are fundamental in deciphering network development. For instance, renowned and persuasive clients will generally acquire associations, making traffic-based

**Table 4.** The comparison of link prediction algorithm based on information dissemination model (Kumar et al., 2020)

Algorithm	Diffusion models				Time complexity	Solving perspective	Social facets		
	LT	IC	TRM	TAM			Local	Global	Quasi
Similarity based link prediction									
Common neighbors (Yao <i>et al.</i> , 2016)	–	–	–	–	$O(N(k)^5)$	Time-varied weight	✓	✗	✗
Adamic adar (Aleta <i>et al.</i> , 2020)	–	–	–	–	$O(N(k)^5)$	Single & multiplex	✓	✗	✗
Jaccard coefficient (Ayoub <i>et al.</i> , 2020)	–	–	–	–	$O(N(k)^5)$	Path depth	✓	✗	✗
Random selection (Nasiri <i>et al.</i> , 2021)	–	–	–	–	$O(1)$	Biased random-walk	✓	✗	✗
SimRank (Dey <i>et al.</i> , 2020)	–	–	–	–	$O(N^3)$	Scale-free simrank	✓	✗	✗
Rooted pagerank (Scholz <i>et al.</i> , 2013)	–	–	–	–	$O(N^3)$	Resource allocation	✓	✗	✗
Probabilistic and maximum likelihood based link prediction									
Stochastic Block Models (SBM) (Stanley <i>et al.</i> , 2019)	✗	✓	✗	✗	Exponential	Continuous attributes	✗	✗	✓
Dimensionality reduction based link prediction									
PCA (Singh <i>et al.</i> , 2019)	–	–	–	–	Exponential	Supervised	✗	✓	✗
Forward feature selection based (Pecli <i>et al.</i> , 2015)	–	–	–	–	Exponential	Supervised	✗	✓	✗
Information diffusion based link prediction									
CLP-ID (Singh <i>et al.</i> , 2020)	✗	✓	✗	✗	$O(D_{avg}(M + \tau N + l^2 C_{avg}))$	Community-based	✓	✗	✗
Hierarchical Structure Model (HSM) (Singh <i>et al.</i> , 2020)	✗	✓	✗	✗	Exponential	Hierarchical structure	✗	✓	✗
Structure-based personalized link prediction model (Yin <i>et al.</i> , 2011)	✗	✓	✗	✗	Exponential	Ego-centric framework	✓	✗	✗
Proximity topological and aggregated features (Al Hasan <i>et al.</i> , 2006)	✓	✓	✗	✗	Exponential	Classification problem	✗	✗	✓
Local probabilistic model (Li <i>et al.</i> , 2016)	✗	✗	✗	✓	Exponential	Markov random field	✓	✗	✗
Logistic regression model (Leskovec <i>et al.</i> , 2010)	✗	✗	✓	✗	Exponential	Signed link prediction	✗	✓	✗

alternate ways and working on the productivity of data spreading on the network. Thus, investigating the dispersion interaction can assist with understanding the effect of clients' cooperation, for example, what re-posting a message means for the spread of images, recordings, or fake news (bits of hearsay) on the networks. Clients in web-based interpersonal organizations make new companions and look for and share data. When a client shares a message, his/her contacts can be affected to re-post that data, driven by the homophily property that creates a diffusion cycle. Link Prediction Approaches are broadly categorized into:

- Similarity-based Approaches
- Probabilistic and Maximum Likelihood-based Approaches
- Dimensionality Reduction-based Approaches
- Information Diffusion-based Link Prediction

#### 4.2.1. Similarity-based approaches

Interface expectation relies on similarity-based measurements, in which a comparability score  $S(x,y)$  is calculated for each pair  $x$  and  $y$ . The score  $S(x,y)$  can be calculated by considering the underlying or hub properties of the thought-about pair. The joins that are not noticed (i.e.,  $U - ET$ ) are assigned scores based on their similarity. Centers with a higher score are more likely to have the expected association. The resemblance measures between each pair are still up in the air using a couple of properties of the association, one of which is an essential property. Scores subject to this property can be assembled in a couple of classes like area and around the world, center ward and way dependent, limit ward and limit free, and so forth.

1. **Local Similarity Indices.** Nearby records are, for the most part, determined utilizing data about normal neighbors and hub degrees. These files think about quick neighbors of a hub. Instances of such files contain normal neighbors, special connections, Adamic/Adar, asset distribution, etc (Lü *et al.*, 2009). A few of the popular local similarity approaches are discussed in the following section:

- *Common Neighbors (CN).* In a given association or graph, the size of common neighbors for a given pair of center points  $x$  and still up in the pair as the size of the combination of the two center regions.
- *Preferential Attachment (PA).* The possibility of a particular connection is applied to create a developing sans-scale network. The term developing addresses the steady idea of hubs over the long run in the organization. The probability augmenting new association related with a hub  $x$  corresponds to  $k_x$ , the level of the hub.
- *CAR based Common Neighbor Index (CAR).* Car setup documents are presented based on the assumption that the interface presence between two center points is additionally plausible in the event that their ordinary neighbors are people from a close by (neighborhood (LCP) speculation).
- *Local Naive Bayes-based Common Neighbors (LNBCN).* This strategy depends on the Guileless Bayes's hypothesis and contentions that diverse ordinary neighbors assume a distinctive part in the organization and contribute contrastingly to the score work processed for non-noticed hub sets.
- *Node Clustering Coefficient (CCLP).* This record is also based on the organization's grouping coefficient property, which registers and adds the bunching coefficients of a seed hub pair's relative abundance of normal neighbors to determine the pair's most recent closeness score.

2. **Global Similarity Indices.** Global files are processed utilizing the whole topological data of an organization. The computational intricacies of such techniques are termed global



similarity indices. These have higher time complexity and appear to be infeasible for enormous organizations. Some of the common Global Similarity Indices approaches are discussed below

- *Katz Index*. It can be considered as a variation of the briefest way metric. It straightforwardly totals over each of the ways among  $x$  and  $y$  and dumps dramatically for longer ways to punish them (Katz, 1953).
- *Leicht–Holme–Newman Global Index (LHNG)*. This overall record relies upon the standard that two centers are similar if the two of them have a speedy neighbor, which resembles the other center point. This is a recursive importance of closeness where an end condition is required. The end condition is introduced similarly to self-closeness, that is, a center is such as itself (Leicht *et al.*, 2006).
- *Matrix Forest Index (MF)*. This matrix depends on the idea of spreading over a tree which is characterized as the sub-graph that traverses complete hubs without shaping any cycle. The crossing tree might contain an out or less number of connections when contrasted with the first chart. The network forests hypothesis expresses that the quantity of traversing the tree in a chart is equivalent to the co-factor of any section of the Laplacian lattice of the diagram. Here, the term forest addresses the association of all established disjoints spreading over trees.

#### 4.2.2. Probabilistic and maximum Likelihood-based approaches

For a given organization  $G(V, E)$ , the probabilistic model upgrades a genuine capacity to set up a model that is made out of a few boundaries. Noticed information of the given organization can be assessed by this model pleasantly. By then, the probability of the presence of a non-existing connection  $(i, j)$  is assessed utilizing restrictive likelihood  $P(A_{ij} = 1|\theta)$ . A few probabilistic models and the greatest probability models have been proposed in writing to gather missing joins in the organizations. The probabilistic models require more data like a hub or edge quality information, notwithstanding primary data. Removing these trait data is difficult; in addition, boundary tuning is likewise no joking matter in such models that limit their materialness. The greatest probability strategies are mind-boggling and tedious, so these models are not reasonable for the genuinely enormous organization (Clauset *et al.*, 2008).

#### 4.2.3. Dimensionality reduction based approaches

Dimensionality Reduction depends on network implanting and lattice decay. A portion of the procedures utilized for dimensionality decrease are Embedded connection expectation, Matrix factorization/disintegration-based connection forecast, and so on. Connection forecast is utilized when any association establishment is considered as a dimensionality decline system in which higher  $D$  layered centers (vertices) are intended to a lower ( $d \ll D$ ) dimensional portrayal (implanting) space by protecting the hub area structures. All in all, find the implanting of hubs to a lower  $d$ -aspects with the end goal that comparable hubs (in the first organization) have comparative implanting (in the portrayal space) (Fukumizu *et al.*, 2004).

Matrix factorization has been used extensively in a number of publications based on connect expectation and recommendation frameworks over the past ten years. Typically, the inactive elements are extracted, and each vertex in the inactive space is addressed using these elements. In a controlled or solitary organization, these depictions are utilized to connect expectations. To improve the anticipated findings, additional hub/connect or characteristic data may be used. Non-negative grid factorization and single-value decomposition have also been applied in the majority of the works.

#### 4.2.4. Information diffusion based link prediction

In web-based interpersonal organizations, when a client sees that his neighbors share or re-post a snippet of data, the client will be affected to think about whether to share or re-post the data, which prompts data diffusion. Data dissemination permits clients to get or notice data that is past the extent of their social cycles. Besides, this peculiarity will impact the production of new connections. Prediction of this network information diffusion-based link prediction. Link Prediction licenses to derive absent or future associations in an organization. The organization association characterizes how data spreads through the hubs. Thus, the spreading might initiate changes in the associations and accelerate the organization's advancement. LP calculations depend on various dispersion processes—Epidemics, Information, Rumor models, and so forth. Every one of these rare occasions of data dissemination-based connection link prediction techniques (Mack, 1985).

Organizations likewise powerfully develop designs in which associations might show up or vanish every once in a while. In this unique situation, interface forecast (LP) targets expecting future affiliations and different applications straightforwardly advantage from such forecasts, for example, kinship investigation in friendly organizations, affiliations and observing of suspects in psychological oppressor organizations, protein affiliations, suggestion frameworks in internet business, and applicable future joint efforts in participating organizations. Given an organization, we approach the two nearby and worldwide constructions, like the neighborhood of a hub, the distance among hubs, and the hubs' local area structure. LP strategies target finding likely connections with primary impact dependent on a neighborhood or worldwide view, which can prompt distinctive expectation results. Normally, the nearby measurements are conventional and clear, while worldwide strategies need generous handling time and can't deal with huge and thick organizations. Because of the networks' intricate design and commotion nature, it is hard to anticipate future connections. Displaying data dispersion in web-based informal communities is a difficult issue, and different analysts contributed to the cycle to this end. In the LP research region, a few systems think about an irregular stroll in an organization, for example, DeepWalk (Perozzi *et al.*, 2014), which has been proposed to look through logical edges utilizing uniform arbitrary strolls, and Node2vec (Grover & Leskovec, 2016), which investigates network areas through lopsided arbitrary strolls.

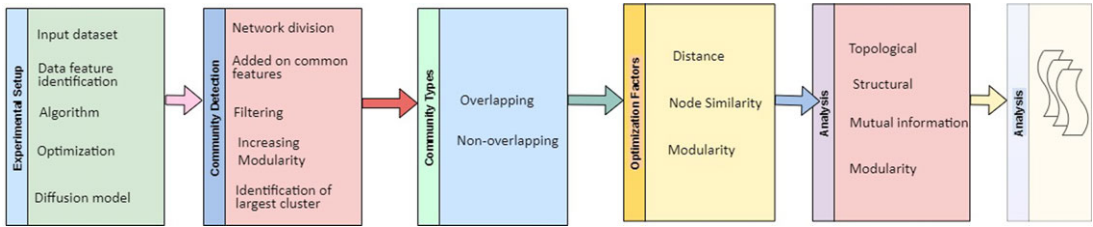
In the dissemination processes, Ally *et al.* (Ally & Zhang, 2018) proposed two overhauling models and thought about the impacts on data spreading in scale-free and little-world organizations. Be that as it may, the creators didn't think about adding new edges or the impacts on the organizations' design. In Li *et al.* (2016), the authors have tended to Sina Weibo and identified a significant component from the data dissemination process, which advanced LP execution. Wu *et al.* (2019) proposed a system called compelling hubs ID LP (INILP) to evaluate the significance of a hub in an organization by appointing every hub a positioning score. The impact of a hub addresses its capacity to spread data to different hubs. Be that as it may, the proposed measurements are primary hub rankings of centrality measures, not an appropriate persuasive spreading model, like pandemics, bits of gossip, or data proliferation. As of late, Wang *et al.* (2020) proposed a neighborhood antagonistic LP technique. However, they ignored the nearby impact of the dispersion elements on the connection development.

### 4.3. Community detection

Several community detection (Kumar *et al.* 2024; Mishra *et al.* 2024) algorithms already exist concerning the information diffusion procedure. We can broadly divide these procedures into four categories which are classical techniques, based on network centrality, overlapping in nature, and depending on stochastic optimization. Figure 5 illustrates the framework of community detection algorithms. In the following subsections, we have discussed all of them in detail.

#### 4.3.1. Classical community detection

Some of the classical community detection methods are utilized for information diffusion techniques. The techniques typically rely on network structure, node similarity, and community structures. It is



**Figure 5.** The community detection framework under information diffusion model (Das & Biswas, 2021b)

suggested to use a hierarchical diffusion method to identify the community structure in Shen *et al.* (2010). The information is diffused from one community to another community in a hierarchical manner based on their placement in the network. In another work, Das and Biswas (2021a), the similarities in the network structure are considered for the information diffusion-based community detection (CSID) approach. The method has utilized the topology of the networks and considered only the importance of the central nodes. Information propagation with raw trajectory data is constructed for the diffusion process based on multiple similarity metrics in Liu and Wang (2016). A decentralized community detection algorithm based on the information diffusion method is proposed in Chen (2011). Ruan *et al.* (2013) have defined an edge strength measuring technique to find out communities utilizing edge-based data diffusion information. In another work, Yang *et al.* (2013) have employed several common metrics and traits along with edge strength measures between communities to define the structural boundaries. Some recent work (Bhattacharya & Sarkar, 2021) has demonstrated contagion and homophily-based information diffusion framework. These methods, however, have not taken into account the centrality considerations; instead, they may have uncovered communities based on network forms and structure.

#### 4.3.2. Network centrality

To identify the core or influential node in any network, a number of network topology measurement metrics are utilized, including degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, page rank, and clustering coefficient. Wang *et al.* (2015) have utilized centrality measures and structural holes to differentiate the important nodes from the dynamic networks utilizing individual nodes' influence. These detected nodes are employed to forecast the data diffusion hub in the network. Bo *et al.* (2014) and Wu *et al.* (2015) have used edge strength measures and mutual information between the nodes for defining information propagation strategy. Based on the shared characteristics of the vertices, Li *et al.* (2015) have suggested a method for selecting seed vertices. In one of the recent works, seed nodes are selected by using different centrality measures and the k-core decomposition method (Devi & Tripathi, 2020). All the methods mentioned above have considered the structure of the network but ignored the overlapping formations. In the next subsection, we have emphasized the overlapping communities and network structures for information diffusion.

#### 4.3.3. Overlapping community

Most social networks in real-life applications have shown overlapping parts within their structural formation. In recent times, a cascade information diffusion model has been proposed to simulate the evolution of communities (He *et al.*, 2021). The fitness function of nodes is updated based on node similarity and uses a clustering methodology for discovering overlapping communities in dynamic networks. In another work, the overlapping communities are discovered based on the communication and sharing information of a node with its neighbors (Sun *et al.*, 2018). A community-based seed selection method has considered the role of overlapping communities based on the gain in the nodes with respect to information diffusion (Wang *et al.*, 2021). Kalantari *et al.* (2020) have proposed a node overlapping community

detection technique based on node probing to distinguish the central nodes from the overlapping communities utilizing only the edge strength. Sun *et al.* (2020) have employed structural similarity measures to define the overlapping parts. In Ramezani *et al.* (2018), cosine similarities are used to find the number of cascades that belong among networks for discovering overlapping parts. One of the works (Alvari *et al.*, 2014) has utilized a game theory-based method to optimize the overlapping part. Though the method has been revised, more efficient optimizing methods have been developed for community detection.

#### 4.3.4. Stochastic optimization

There are several optimized community detection methods proposed for efficient information diffusion. In one of the popular methods, Hajibagheri *et al.* (2012), each node of the network is considered an agent. The interactions among the agents with their neighbors are optimized based on the proposed utility function. In Hajibagheri *et al.* (2013), the particle swarms method is used in network neighborhoods for optimizing objective functions. In another work, an approach is developed to discover information diffusion processes from online social networks based on modularity maximization (Obregon *et al.*, 2019). Another technique based on a genetic algorithm is utilized in all types of networks for optimized community formation (Mozafari & Hamzeh, 2015). Recently, a multi-objective optimization model utilizing key node identification has been introduced (Kalantari *et al.*, 2020). Through the study mentioned above, we have shown that most of the diffusion techniques, irrespective of their discovering nature, heavily depend on the network properties for high accuracy and efficiency.

In Table 5, we have compared all the community detection algorithms based on the information dissemination model. Four diffusion models—the linear threshold (LT), independent cascade (IC), threshold model (TM), and time-aware model (TAM)—have been taken into account. We have defined the time complexity of each algorithm, solving perspective, social facets, and network properties individually. The comparison mentioned above among the algorithms gives us a clear idea about the impact of the techniques in this domain.

## 5. Evaluation metrics

Evaluation metrics are crucial for assessing the quality and impact of research innovations. These metrics provide objective measures for assessing the importance, originality, and impact of the research on the field. Through the use of suitable evaluation metrics, researchers can assess the effectiveness of information diffusion models, deployment strategies, and proposed applications in a quantitative manner. In addition, these metrics allow for easy comparisons with other methodologies, making it easier to identify strengths, weaknesses, and areas that can be improved. Ultimately, the careful use of evaluation metrics ensures a high level of rigor and credibility in scholarly discourse, which leads to advancements in understanding and implementing information diffusion processes. This section is devoted to performance measures of different applications of information diffusion analysis. So we will discuss the evaluation metrics of influence maximization, link prediction, and community detection as illustrated in Table 6.

### 5.1. Influence maximization evaluation metrics

Four matrices quality, efficiency, memory footprint, and robustness analyze the performance of IM algorithms.

1. **Quality.** The influence spread estimates the quality of an IM approach, that is, the number of product adoption or users who a piece of information, news, idea, innovation, and product has influenced. Most of the traditional approaches ignore the effectiveness of seed users. Therefore, some of the recent works (Barbieri *et al.*, 2012; Zhuang *et al.*, 2013; Wang *et al.*, 2017; Singh *et al.*, 2019a) focus on contextual features to improve the effectiveness of seed users.

**Table 5.** The comparison of community detection algorithm based on information dissemination model (Das & Biswas, 2021b)

Algorithm	Diffusion models				Time complexity	Solving perspective	Social facets	N/W properties
	LT	IC	TM	TAM				
Hierarchical diffusion algorithm (Shen <i>et al.</i> , 2010)	✗	✓	✗	✗	$O(N + M)$	Disjoint Community	Common Neighbours	Complex, Small
Game theory (Alvari <i>et al.</i> , 2014)	✗	✗	✓	✗	$O(N + M)$	Disjoint Community	Mutual Information	Complex, Small
Cascade label propagation (Sattari & Zamanifar, 2018)	✓	✓	✗	✗	$O(N + M)$	Overlapping Community	Information Propagation	Belonging Factor
Information diffusion (Ramezani <i>et al.</i> , 2018)	✓	✗	✗	✓	$O(N + M)$	Disjoint Community	Number of Cascade	Similarity among Nodes
Information dynamics (Sun <i>et al.</i> , 2018)	✓	✓	✓	✗	$O(kN + LNk + cN)$	Overlapping Community	Social Interaction	Belonging Factor
Information diffusion cascade (Zhang <i>et al.</i> , 2018)	✓	✓	✗	✗	$O(Ic + kN)$	Disjoint Community	Social Interaction	Euclidean Distance
Information dynamics (Sun <i>et al.</i> , 2020)	✓	✓	✓	✓	$O(LNk + M + Lnk)$	Disjoint Community	Information	Structural equivalence

**Table 6.** The comparison of evaluation attributes (Kumar et al., 2020; Singh et al., 2021; Das & Biswas, 2021b)

Attribute type	Attribute	Computation of metric	Representative
Influence maximization			
Quality	Influence spread	Number of product adoption, $V_A(S)$	Singh et al. (2021), Arora et al. (2017)
Efficiency	Running time	Seed nodes identification time	Singh et al. (2021), Arora et al. (2017)
Memory footprint	Memory space	Memory consumption to run the seeding algorithm	Singh et al. (2021), Arora et al. (2017)
Robustness	Rate of change	Measuring stochastic behaviour of solution	Singh et al. (2021), Arora et al. (2017)
Link prediction			
Quality	AUC	Geometric mean of the area under the receiver operator's characteristic curve	Song & Meyer (2015)
Accuracy	AUPR	Area under the precision-recall curve (AUPRC)	Yang et al. (2015)
	Avg precision	Average of precision values	Morstatter et al. (2016)
	Recall	Fraction of relevant instances	Morstatter et al. (2016)
	F1-Score	harmonic mean of the precision and recall	
	Balanced accuracy	arithmetic mean of sensitivity and specificity	Chicco et al. (2021)
Community detection			
Quality	Modularity	Strength of division of a network into groups	Kumar et al. (2020), Kumar et al. (2019a)
	f-Modularity	Fuzzy measure of the strength of division of a network into groups	Kumar et al. (2020), Kumar et al. (2019a)
	ANUI	Strength of overlapping communities based on mutual information	Kumar et al. (2020), Kumar et al. (2019a)



Table 6. (Continued)

Attribute type	Attribute	Computation of metric	Representative
Efficiency	Conductance	Counts the number of cliques in connected graph	Kumar <i>et al.</i> (2019b)
	Permanence	Distribution of a vertex in internal and external community	Kumar <i>et al.</i> (2020), Kumar <i>et al.</i> (2019b)
	Purity	Measure the similarity between detected and ground truth communities	Kumar <i>et al.</i> (2019b)
	NMI	Define strength of communities based on mutual information	Kumar <i>et al.</i> (2020), Kumar <i>et al.</i> (2019b)
	ONMI	Measures similarity among communities without consider the number of partitions	Kumar <i>et al.</i> (2020), Kumar <i>et al.</i> (2019a)
	F-Measure	Measures the accuracy by considering precision and recall	Kumar <i>et al.</i> (2020), Kumar <i>et al.</i> (2019b)
	F1-Score	Measures the similarity between two overlapping communities	Kumar <i>et al.</i> (2020), Kumar <i>et al.</i> (2019a)
	ENMI	Find the strength of overlapping communities based on mutual information	Kumar <i>et al.</i> (2019a)

2. **Efficiency.** The efficiency of IM algorithms has been measured by the running time of algorithms to produce the desired seed users. In general, running time will increase with the seed size. However, there are some exceptions like TIM (Tang *et al.*, 2014) and IMM (Tang *et al.*, 2015). The simulation-based approaches don't have a better performance in terms of efficiency. Therefore some heuristic-based (Singh *et al.*, 2019b), meta-heuristics (Singh *et al.*, 2019a; Singh *et al.*, 2019b) and sampling (Jiang *et al.*, 2011; Goyal *et al.*, 2011b) methods have been introduced to overcome the limitations of simulation-based methods.
3. **Memory Footprint.** The memory footprint is also a factor in addition to the time efficiency of an approach to measure the scalability of an algorithm. The memory footprint is evaluated by analyzing the memory required to run a seeding strategy for seed identification. The large size of the dataset leads to sampling methods (Jiang *et al.* 2011; Goyal *et al.* 2011b) to take a sample of datasets for identifying seed users based on structural and topological features.
4. **Robustness.** Most IM algorithms are not robust as a slight modification in diffusion models leads to major changes in inactivation results. These IM algorithms only focus on quality, efficiency, and scalability while ignoring robustness. However, few works (Jung *et al.*, 2012; Galhotra *et al.*, 2016; He & Kempe, 2016; Mehmood *et al.*, 2016) have been presented in the literature to ensure robustness on diffusion analysis.

## 5.2. Link prediction evaluation metrics

There are numerous conventional, straightforward, and fundamental connection expectation metrics used to evaluate Link Prediction. Mainly, it can be divided into two major parts, which are quality and accuracy as defined in Table 6.

### 5.2.1. Quality

Any link prediction quality calculation means the Precision of the calculation concerning the ground truth. Likewise, it signifies the strength of the networks. The accompanying boundaries are utilized to measure the nature of the local area discovery strategies. The metrics used to calculate the quality of link prediction are (Kumar *et al.*, 2019a; Kumar *et al.*, 2019b; Kumar *et al.*, 2020):

1. **AUC (Area under the ROC Curve).** A ROC (recipient working trademark bend) is a diagram showing the presentation of a grouping model at all order limits. This ROC plots two boundaries: True Positive Rate (TPR) and False Positive Rate (FPR). A ROC plots TPR versus FPR at various order edges. To register the focuses in an ROC, we could assess a calculated relapse model ordinarily with different characterization limits, yet all the same, this would be wasteful. Luckily, there's a productive, arranging-based calculation that can give this data to us, called AUC (Fawcett, 2006).
2. **AUPR.** It is the area under the precision-recall curve. It is a helpful metric to present imbalanced information in an issue setting where very much attention is in tracking down the positive models. The AUPRC is determined as the space under the PR bend. A PR bend shows the compromise between accuracy and review across various choice limits. The x-pivot of a PR bend is the review, and the y-hub is the accuracy. Rather than ROC bends, this is where the y-pivot is the review, and the x-hub is FPR. Like plotted ROC bends, the choice edges are verifiable in a plotted PR bend and are not displayed as a different hub. To compute AUPRC, we ascertain the region under the PR bend. There are numerous techniques for computation space under the PR bend, including the lower trapezoid assessor, the introduced middle assessor, and the average Precision (Saito & Rehmsmeier, 2015).
3. **Average Precision.** It is the average precision values.

### 5.2.2. Accuracy

For any link prediction tracking method, accuracy signifies the precision of the calculation in regard to the calculation time. It relies a lot upon the size of the organization. A few boundaries are used to measure the productivity of the link prediction over a social network (Kumar *et al.*, 2019b; Kumar *et al.*, 2019a; Kumar *et al.*, 2020).

1. **Recall.** It is a metric that evaluates the quantity of right sure link forecasts (true Positive) made from all certain expectations (True Positive + False Negative) that might have been made in the complete network.

$$recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (1)$$

2. **F1-Score.** It is the harmonic mean of precision and recall.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2)$$

3. **Balanced Accuracy.** It is the arithmetic mean of Sensitivity (true positive rate or recall) and specificity (true negative rate or 1 – false positive rate)

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2} \quad (3)$$

### 5.3. Community detection evaluation metrics

The evaluation metrics of community detection are used to define the quality of the detected communities. Mostly, it can be divided into two major parts, which are quality and efficiency as defined in Table 6.

#### 5.3.1. Quality

For any community detection algorithm, quality denotes the accuracy of the algorithm with respect to the ground truth. Also, it denotes the strength of the communities. The following parameters are used to measure the quality of the community detection techniques.

1. **Modularity.** Modularity (Muhuri & Mukhopadhyay, 2021) measures the robustness of the partitions in a network. Modularity finds the density of connections between the nodes of the same community rest of the communities. The modularity  $Q$  of a graph with  $m$  number of edges can be defined as

$$Q = \frac{1}{2m} \sum_{vw} \left[ A_{vw} - \frac{k_v k_w}{2m} \right] \delta(c_v, c_w) \quad (4)$$

$A_{vw}$  represents the connectivity of node  $i$  and  $j$ ,  $k_v$  denotes the sum of the weights of the edges incident with node  $v$ ,  $c_v$  is the community of node  $v$ ,  $\delta(c_v, c_w)$  is equals to 1 when  $v$  and  $w$  are in the same community, 0 otherwise.

2. **f-Modularity.** It is a fuzzy measure that finds the mutual information among communities (Guo *et al.*, 2021). The value decreases with the contraction of the community structures. f-modularity of a graph  $G$  can be denoted as

$$f_{mod}(G) = \max_{D \in C} \sum_{u,v} (\sigma f(D_{u,v}) F_{u,v} - f^*(\sigma f(D_{u,v})) J_{u,v}) \quad (5)$$

Here  $D$  distinguishes between frequency matrix  $F$  and random matrix  $J$ .  $C$  is a constraint set.

3. **ANUI.** Average normalized unifiability and isolability (ANUI) analyses both unifiability and isolability to denote the quality of an overlapping community (Das & Biswas, 2021b). A graph  $G$  with  $k$  communities, unifiability of community  $C_i$  to community  $C_j$  is denoted as

$$\text{unifiability}(C_i) = \sum_{i=1}^k \text{Unifiability}(C_i, C_j) \quad (6)$$

$$\text{isolability}(C_i) = \frac{(u, v) | u \in C_i(v)}{(u, v); (u, w) | u \in C_i(v) \& w \notin C_i} \quad (7)$$

ANUI is denoted as

$$\text{ANUI}(C_i) = \frac{\text{unifiability}(C_i) + \text{isolability}(C_i)}{2} \quad (8)$$

4. **Conductance.** It finds structural cohesiveness by incorporating external connections of the communities. Conductance of forecast community  $C$  is denoted as

$$f(w) = \frac{E_w^{\text{out}}}{2.E_w^{\text{in}} + E_w^{\text{out}}} \quad (9)$$

where  $E_w^{\text{out}}$  and  $E_w^{\text{in}}$  denote the number of edges that are outside and inside of the community  $C$ , respectively.

5. **Permanence.** Permanence (Chakraborty *et al.*, 2014) is a measure to find the pull of the neighborhood node on a particular node  $v$ . The internal clustering coefficient  $c_{in}$  is also utilized to find the permanence  $\text{perm}(v)$ .

$$\text{perm}(v) = \left[ \frac{I(v)}{E_{\max}(v)} \times \frac{1}{D(v)} \right] - [1 - c_{in}(v)] \quad (10)$$

where  $I(v)$  is internal connections,  $E_{\max}(v)$  is the strength of internal community, and  $D(v)$  is the degree of node  $v$ .

### 5.3.2. Efficiency

For any community finding technique, efficiency denotes the accuracy of the algorithm with respect to the computation time. Also, it depends on the size of the network. Some parameters are utilized to measure the efficiency of the community detection algorithms.

1. **Purity.** It assigns the detected community to the most frequent baseline (Lin *et al.*, 2005).

$$\text{Purity}(\Omega, C) = \frac{1}{N} \sum_k \max_j |\omega_k, c_j| \quad (11)$$

where  $\Omega$  is the set of ground truth communities.

2. **NMI.** Normalized Mutual Information (NMI) finds the mutual dependence among the communities (Xie *et al.*, 2013). NMI detects the similarity and quality between two divisions. Let the two communities are  $C$  and  $C'$ . Then NMI can be represented as

$$\text{NMI}(C, C') = \frac{2I(C, C')}{H(C) + H(C')} \quad (12)$$

where  $H(\cdot)$  represents entropy function and the mutual information  $I(C, C') = H(C) + H(C') - H(C, C')$ . No similarity and maximum similarity of two communities are denoted as the NMI value 0 and 1, respectively.

3. **ONMI.** Overlap-Normalized Mutual Information (ONMI) is proposed in Van Lierde *et al.* (2019) for overlapping communities. The true and the predicted partitions are denoted by  $C_l$  and  $\bar{C}_l$ , respectively, where  $1 \leq l \leq K, \bar{K}$ . The number of true and predicted partitions are  $K$  and

$\bar{K}$ , respectively. Then ONMI can be represented as:

$$ONMI(X, \bar{X}) = 1 - \frac{1}{2}(H(X|\bar{X}) + H(\bar{X}|X)) \quad (13)$$

where  $X_i = 1$  denotes the event that a node belongs to  $C_i$  and  $\bar{X}_i = 1$  denotes the event that a node belongs to  $\bar{C}_i$ . Here normalized entropy of all  $X_k$  given all  $\bar{X}_i$  is

$$H(X|\bar{X}) = \frac{1}{k} \sum_k \frac{H(X_k|\bar{X})}{H(X_k)} \quad (14)$$

High ONMI value is expected for the overlapping communities.

4. **F-Measure.** it is the harmonic mean of both types of purity measures or between Precision and recall (Chakraborty *et al.*, 2017).

$$F - measure = \frac{2.Purity(\Omega, C).Purity(C, \Omega)}{Purity(\Omega, C) + Purity(C, \Omega)} \quad (15)$$

5. **F1-score.** It measures the similarities between two overlapping communities (Chakraborty *et al.*, 2017).

$$F1 = \frac{1}{2} \left( \frac{1}{\phi} \sum_{\phi_i \in \phi} F1(\phi_i, C_{g(i)}) + \frac{1}{C} \sum_{c_i \in C} F1(\phi_g, (i), c_i) \right) \quad (16)$$

It represents the average of the F1-score of the best-detected partition and ground truth communities and the F1-score of the ground truth communities and best-detected partition.

6. **ENMI.** Extended normalized mutual information (ENMI) compares the partitions with the baseline (Sun *et al.*, 2018). ENMI is defined as

$$ENMI(X|Y) = 1 - [H(X|Y) + H(Y|X)]/2 \quad (17)$$

where  $H(X|Y)$  is denoted as

$$H(X|Y) = 1 - \frac{1}{C'} \sum \frac{H(X_k|Y)}{H(X_k)} \quad (18)$$

where  $X$  and  $Y$  denotes the random variables connected with the communities  $C$  and  $C'$ , respectively.  $H(X|Y)$  is the normalized entropy of community  $X$  given community  $Y$ .

## 6. Research challenges and future directions

The research challenges of information diffusion analysis have been categorized based on the applications. This section will discuss the challenges of information diffusion on influence maximization, link prediction, and community detection. These challenges are stated below in future directions.

### 6.1. Influence maximization

The computation of approximate influence spread is dependent on the diffusion process, which is stochastic. Therefore, it activates a different set of users in each diffusion round. Few efforts have been made to tackle these uncertainties, illustrated as follows.

- **Diffusion Process Modeling.** The first challenge of information diffusion analysis corresponding to the IM problem is how to adapt the diffusion process for viral marketing. It is hard to accommodate user behavior about product adoption as the diffusion process is random. Therefore, some efforts (Kempe *et al.*, 2003; Guille *et al.*, 2013) have been made to tackle the adaption of the diffusion process for the IM problem by incorporating and extending traditional diffusion models.

**Table 7.** Illustration of information diffusion challenges over different applications (Kumar et al., 2020; Das & Biswas, 2021b; Singh et al., 2021)

	Framework	Challenges	Representative
Research challenges	Influence maximization	Stability and robustness	Jung et al. (2012)
		Adjusting submodularity	Galhotra et al. (2016)
		Adopting conformity	Li et al. (2011), Li et al. (2013), Tang et al. (2013)
		Overlapping users identification	Buccafurri et al. (2012), Suganya et al. (2017)
		Network coupling	Zhang et al. (2016), Wang et al. (2016)
		Incorporating heterogeneous diffusion models	Li et al. (2012), Zhan et al. (2015)
		Budget fixing	Sun et al. (2016), Singh et al. (2019b)
		Incorporating product influence	Sun et al. (2016)
		Adopting contextual features	Su et al. (2018), Lee & Chung (2015), Bozorgi et al. (2017)
		Incorporating uncertainty and incompleteness	Lei et al. (2015), Tong et al. (2017)
	Link prediction	Efficiency within link prediction	Zhang et al. (2019)
		Sparsity of data	Beigi et al. (2019)
		Trusted link prediction	Golzardi et al. (2019)
		Effectively breaking of complex networks into simple networks	Aslan et al. (2019)
		Level-wise link prediction	Kumar et al. (2019b)
	Community detection	Meaningful community	Shen et al. (2010)
		Optimizing running time	Das and Biswas (2021a)
		Noise removal	Chen (2011)
		Model generalization	Li et al. (2015), Devi & Tripathi (2020)
		Sparse network	Alvari et al. (2014)

- Complexity and Optimality.** Kempe et al. (2003) have proved that the IM problem is NP-hard under classical diffusion models. The influence spread computation under these models is computationally complex. Due to this, most simulation-based IM algorithms are inefficient and have some approximation guarantee to the optimal solution. However, heuristics and sampling methods are more time-efficient but less optimal. It is tough to find a near-optimal scalable solution. Therefore, most of the algorithms (Chen et al., 2010; Singh et al., 2019a; Singh et al., 2019b) focus on the trade-off between optimality and scalability.
- Stochasticity of Diffusion Process.** Most of the diffusion models are stochastic. Therefore, IM algorithms produce a different set of active users in each iteration from seed users due to the randomness of the diffusion process. So it is very tough to collect the final product



adoption. However, some strategies like the mean and median of repeated iteration outcomes are applied to tackle the IM problem. For example, simulation-based methods (Kempe *et al.*, 2003; Leskovec *et al.*, 2007) perform  $r = 10\,000$  simulations while sampling methods (Cheng *et al.*, 2013; Ohsaka *et al.*, 2014) iterate multiple times on different sample episodes.

- **Stability and Robustness.** Most of the IM methods are based on the assumption that the network structure is fixed while the network structure keeps on changing in the real world. Therefore, a slight change in the network can lead to a significant update in seed users. The authors of He & Kempe (2014) also stated that noisy influence probabilities could generate the inaccurate seed. Therefore some efforts (Jung *et al.*, 2012; Galhotra *et al.*, 2016; He & Kempe, 2016; Mehmood *et al.*, 2016) have been made to find a robust solution for IM problem.
- **Incorporating Group Norms.** Most of the literature considers only individual influence under the IC diffusion model and avoids group norms. However, people are also influenced by acquaintances and group norms such as trust, conformity, and context. People who share common interests, backgrounds, cultures, etc., are more inclined to each other's thoughts. Therefore, some efforts (Li *et al.*, 2011; Tang *et al.*, 2013; Zhang *et al.*, 2014) have been made to incorporate group norms with diffusion models. There are few possible future directions with group norms that can consider user profile information with group characteristics such as trust and conformity.
- **Adoption of Weaker Submodularity.** The evaluation function is strict with the definition in general submodularity of approximate influence. This reduces the number of MC simulations under classical diffusion models for the IM problem. However, some of the advances (Li *et al.*, 2011; Gionis *et al.*, 2013; Galhotra *et al.*, 2016) in the IM problem states that strictness of submodularity leads to no approximation guarantee of the solution. Because these advances are based on opinion formation and change. Therefore, it is required to adopt weaker submodularity in influence diffusion function to ensure some theoretical guarantee of the solution. There are some efforts (Das & Kempe, 2011) have made to weaken the submodularity of influence diffusion.
- **Incorporating Heterogeneous Diffusion Process.** Most of the algorithms under the IM2 and MIM2 framework assume that every layer, relationship, or network has used the same diffusion models. However, in reality, each network has its diffusion method. Therefore, one possible future work can be the consideration of different diffusion processes across the layers and networks to achieve a more realistic scenario. However, some efforts (Zhan *et al.*, 2015; Kuhnle *et al.*, 2018) have been made to ensure heterogeneity across layers for the IM problem.
- **Considering Contextual Diffusion Models.** Most of the classical IM algorithms ignore context-aware diffusion models by incorporating conformity, semantics, location, etc., with diffusion models for novel viral marketing applications. However, some efforts (Lee & Chung, 2015; Bozorgi *et al.*, 2017; Su *et al.*, 2018) have been made to consider contextual features with diffusion models. Still, most of these contextual features are unexplored.
- **Incorporating Conformity.** Conformity is a crucial factor in how social influence spreads throughout societies. It describes how people tend to conform to the attitudes, beliefs, and behaviors of a majority group or societal norms. This phenomenon can occur through different mechanisms, such as informational conformity, where individuals conform because they believe others have relevant knowledge, or normative conformity, where individuals conform to gain social approval or avoid rejection. It supports processes of influence diffusion, where the adoption of certain behaviors or attitudes by a few individuals within a social network can lead to their widespread acceptance and adoption by others. Gaining insight into conformity is essential for understanding the development and evolution of social norms, as well as their influence on individual and collective behavior. Thus, to enhance the efficiency of the seed nodes, the authors in Li *et al.* (2011) made some attempts, yet many of these aspects remain unexplored.

## 6.2. Link prediction

This section discusses the challenges and efforts made to tackle these challenges, illustrated as follows.

- Efficiency with Link Prediction.** Because of the huge measure of information in interpersonal organizations, efficient and reliable link prediction addresses a critical test. To resolve this issue, researchers have proposed two calculations for interface expectation. The proposed calculations tackled the productivity issue by embracing low-rank factorization models while demonstrating extremely proficient contrasts with different techniques. This way, the review addresses a huge advance forward in the test of fostering a productive connection forecast model. Another significant issue that should be tended to worries the precision of connection forecast draws near. The test discoveries uncovered that the proposed approach beat numerous different techniques as far as exactness and versatility, and the related runtime were not exactly as seen in past examinations. Be that as it may, albeit the proposed approach improved connection forecast precision, it stays untested on large network information, which might see its proficiency become sabotaged (Fire *et al.*, 2014).
- Sparsity of Data.** One more issue identified with interface forecast in marked SNs concerns the sparsity of information. To reduce this issue, researchers have proposed some effective methods, for instance, an imaginative methodology that is fit for investigating the client's character utilizing web-based media. The acquired test results demonstrated an integral relationship between the marked connection forecast issue and character data. An intriguing cutting-edge technique for working with connect expectation is the utilization of worldly consistency in relational correspondence to focus on weighted edges in network diagrams. Another approach used is contrasted and different techniques, this strategy predicts interfaces regardless of whether there is a shortage in the quantity of edges required for investigation. A differentiating similitude-based connection expectation strategy is based on fuzzy connection significance. The technique performed well, utilizing two methodologies to accomplish its targets. Initially, for the determination of the neighbor, the distance between hubs was utilized. Besides, the fluffy connection significance was utilized to track down the pertinent connection. By utilizing these systems, the technique has acquired sound outcomes. The challenge for a sparse network is now been addressed but still needs a lot of attention as the network is getting sparse exponentially (Beigi *et al.*, 2019).
- Trusted Link Prediction.** With the improvement of the Internet, an ever-increasing number of people or associations will more often than not discuss and communicate on the organizational stage. Through friendly stages, individuals can not just offer their sentiments about various items yet, in addition, express their perspectives on others, which extraordinarily improves individuals' social exercises. Nonetheless, the fast improvement of social stages has filled them with many futile or bogus data and records. To rapidly and effectively peruse the substance for intrigued, clients as a rule add clients who have normal interests to the 'trust list.' simultaneously, to try not to peruse the substance not intrigued, clients generally add clients with inverse interests to the 'doubt list.' For instance, and have comparative interests and have some contention in a specific region. In this way, Twitter may follow and add to the boycott. We can assemble a marked interpersonal organization by catching the trust and doubt connections between clients. Through the trust/doubt connections in the marked informal community, we can not just know which clients the objective client has social associations with yet additionally know what sort of mentality the objective client embraces toward these clients. In any case, the social connections in an interpersonal organization are excessively inadequate, which truly upsets the extension of the client's group of friends and further improves the social stage. Thus, it has become important to assist clients with finding all the more new companions or confided in clients. Luckily, clients have left a lot of verifiable conduct information (e.g., client's appraising and remarks) on friendly stages, which gives great conditions to assessing the trust connections and comparability of inclinations between various clients. Notwithstanding, the

way to deal with observing likely companions through shared interests actually faces many difficulties (Fan *et al.*, 2019).

- **Level-wise Link Prediction.** In complex networks, level-wise link prediction is used as it breaks the network into smaller networks and then predicts the link for each small network. The biggest challenge is dividing the complex network into smaller networks and then combining the link prediction of all networks and concluding it. The networks are so dynamic, and thus, predicting the link is a challenge! Trust, efficiency, and accuracy have to be maintained at every level as the next level depends on the previous link prediction. Thus it becomes a very challenging task (Kumar *et al.*, 2020).

### 6.3. Community detection

There are several challenges associated with community detection concerning information diffusion. The challenges are mostly divided into five categories discussed here in detail.

- **Meaningful Communities.** In an online social network, all the detected communities are not meaningful at all. The community detection methods might create some partitions that have no importance from the application point of view. In Shen *et al.* (2010), the problem has been addressed.
- **Running Time Optimization.** Running time optimization is another important challenge in community detection as discussed in Das and Biswas (2021b). Though game-theoretic methods have performed better theoretically, their computational complexity should be reduced to achieve more efficiency.
- **Noise Removal.** All the community detection algorithms are sensitive to noise. Different noise removal filters can be deployed to increase the technique's accuracy. The problem has been discussed in Chen (2011) concerning a decentralized network.
- **Model Generalization.** Researchers are very much focused on generalizing the model for different types of complex networks, including the independent cascade, biological, linear threshold, and epidemics models. In Li *et al.* (2015) and Devi and Tripathi (2020), authors have discussed the issue for encouraging futuristic models.
- **Sparse Network.** If in a network, the number of 0 is greater than the number of 1, or the network consists of a huge number of isolated vertex. The community detection algorithms have failed to generate efficient communities as shown in Alvari *et al.* (2014). This challenge should be addressed for the refinement of the existing algorithms.

## 7. Open problems

In this section, we have discussed the open problems associated with information diffusion analysis.

- *How to aggregate information dissemination with reality?* Most information dissemination models are not much closer to reality as these are the extension of classical diffusion models used across the domains. Therefore, it is necessary to strengthen these models to acquire actual word-of-mouth effects. Apart from extension, some new diffusion models are introduced for social network analysis. However, these models are more specific to applications and scenarios. Validation and correctness are also a concern as it is based on feedback from the Internet. Therefore, aggregating reality with information diffusion analysis is still much explored.
- *How to attain equivalence between feasibility and generality?* Some diffusion models, like the competitive diffusion models, are not monotonic and submodular. So applicability of greedy-based methods is limited under such models. However, the literature lists some directions to handle these diffusion settings, such as nature-inspired optimization, deep learning, etc. The generality of existing methods under these settings is still a problem to explore.

- *How to obtain stochasticity of information diffusion process?* Some studies suggest that noisy and inaccurate data for activation and influence probabilities can lead to incorrect adoption of information diffusion behavior (He & Kempe, 2014). Therefore, it is necessary to adopt real dynamics of information diffusion such as structural features, promotional impact, activation, and influence behavior to achieve the stochasticity of diffusion models. However, it has attracted the attention of researchers in recent years but is still much open to exploring this estimation.
- *How to acquire real-world complexity with information diffusion analysis?* Most of the information diffusion analysis advances focuses on the topological structure and classical diffusion process. Therefore, these models are not consistent with reality as they ignore social, political, cultural, psychological, geographical, and environmental factors. So it is still an open problem to explore new possibilities of the information diffusion process to align with the complex reality.
- *How to deal with opinion change and negative influence in information dissemination?* Most of the literature does not consider the adverse effect and opinion change in information diffusion analysis. These works consider positive impact only, and once a user is influenced by a piece of information, always intact with the same. This is not the case with actual scenarios where users are constantly changing and influenced by new information. Therefore, it is required to adopt negative and opinion change with information diffusion analysis. However, some studies (Li et al., 2014; Shen et al., 2015) are introduced regarding this but are still exploring.
- *How to tackle multiple, complementary, and competing ideas with information diffusion?* In recent years, the applicability of information diffusion to some practical applications has attracted researchers to focus on more realistic assumptions. For example, viral marketing applications need to be more specific toward product and user characteristics rather than only topological structure. Similarly, considering users' influence in a single online social network will ignore the user's influence over other networks, leading to inaccurate estimation of influence. Therefore, incorporating multiple, complementary, and competitive ideas closer to reality needs to be explored.
- *How to assure ethical use of information propagated?* The rapid growth of online social networks and their benefits leads to the urgency of users' privacy protection. Users and network service providers need to understand the importance of privacy-preserving and ethical use of social media platforms. Ignoring the ethical use of platforms can have a severe effect on users and service providers. Therefore, privacy-preserving and ethical use can be considered with information-sharing perspective for exploring.

## 8. Concluding remarks

This paper surveyed the role, emergence, and significance of information dissemination for social network analysis. Information diffusion plays a key role in developing and proposing new solutions for social network applications. This study analyzed the various diffusion models with their vulnerabilities. Primarily, a brief discussion about the diffusion process component is presented. An in-depth information diffusion deployment and application is studied. A comparative analysis of the literature corresponding to influence maximization, link prediction, and community detection algorithms is discussed keeping the viewpoint of information diffusion. The study reveals the performance evaluation metrics for diffusion analysis applications. Finally, research gaps, challenges, and future prospects have been discussed, along with some open problems. This study will provide naive researchers with the basic and recent trends in information diffusion analysis with a better understanding and a good starting point for their research.

**Funding.** This article does not contain any studies with human participants or animals performed by any of the authors. The article presents a review of information diffusion analysis approaches. No funding has been received for this work.

**Competing interests.** The authors declare no conflicts of interest.

**Authors contribution statement.** Shashank Sheshar Singh: Conceptualization, Formal analysis, Writing—original draft, Revision. Divya Srivastava: Formal Analysis, Writing—original draft. Madhushi Verma: Formal analysis, Writing—original draft. Samya Muhuri: Formal Analysis, Writing—original draft.

**Data availability and access.** Our work does not have any data to explore. Data sharing is not applicable to this article, as no datasets were generated or analyzed during the current study.

## References

- Abdullah, S. & Wu, X. 2011. An epidemic model for news spreading on twitter. In *2011 IEEE 23rd International Conference on Tools with Artificial Intelligence*, 163–169. IEEE.
- Aggarwal, C. C., Lin, S. & Yu, P. S. 2012. On influential node discovery in dynamic social networks. In *SDM*.
- Al Hasan, M., Chaoji, V., Salem, S. & Zaki, M. 2006. Link prediction using supervised learning. In *SDM06: Workshop on Link Analysis, Counter-Terrorism and Security*, 30, 798–805.
- Aleta, A., Tuninetti, M., Paolotti, D., Moreno, Y. & Starnini, M. 2020. Link prediction in multiplex networks via triadic closure. *Physical Review Research* 2(4), 042029.
- Ally, A. F. & Zhang, N. 2018. Effects of rewiring strategies on information spreading in complex dynamic networks. *Communications in Nonlinear Science and Numerical Simulation* 57, 97–110.
- Alvari, H., Hajibagheri, A. & Sukthankar, G. 2014. Community detection in dynamic social networks: A game-theoretic approach. In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, 101–107. IEEE.
- Arora, A., Galhotra, S. & Ranu, S. 2017. Debunking the myths of influence maximization: An in-depth benchmarking study. In *Proceedings of the 2017 ACM International Conference on Management of Data, SIGMOD'17*, 651–666. ACM. <http://doi.acm.org/10.1145/3035918.3035924>
- Aslan, S., Kaya, B. & Kaya, M. 2019. Predicting potential links by using strengthened projections in evolving bipartite networks. *Physica A: Statistical Mechanics and its Applications* 525, 998–1011.
- Ayoub, J., Lotfi, D., El Marraki, M. & Hammouch, A. 2020. Accurate link prediction method based on path length between a pair of unlinked nodes and their degree. *Social Network Analysis and Mining* 10(1), 1–13.
- Azaouzi, M., Mnasri, W. & Romdhane, L. B. 2021. New trends in influence maximization models. *Computer Science Review* 40, 100393.
- Backstrom, L. & Leskovec, J. 2011. Supervised random walks: Predicting and recommending links in social networks. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 635–644.
- Barbieri, N., Bonchi, F. & Manco, G. 2012. Topic-aware social influence propagation models. In *2012 IEEE 12th International Conference on Data Mining*, 81–90.
- Beigi, G., Ranganath, S. & Liu, H. 2019. Signed link prediction with sparse data: The role of personality information. In *Companion Proceedings of The 2019 World Wide Web Conference*, 1270–1278.
- Bhagat, S., Goyal, A. & Lakshmanan, L. V. 2012. Maximizing product adoption in social networks. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, WSDM'12*, 603–612. ACM. <http://doi.acm.org/10.1145/2124295.2124368>
- Bharathi, S., Kempe, D. & Salek, M. 2007. Competitive influence maximization in social networks. In *Internet and Network Economics*, Deng, X. & Graham, F. C. (eds). Springer Berlin Heidelberg, 306–311.
- Bhattacharya, S. & Sarkar, D. 2021. Study on information diffusion in online social network. In *Proceedings of International Conference on Frontiers in Computing and Systems*, 279–288. Springer.
- Bhattacharjee, A. & Sanford, C. 2006. Influence processes for information technology acceptance: An elaboration likelihood model. *MIS Quarterly* 30(4), 805–825.
- Bliss, C. A., Frank, M. R., Danforth, C. M. & Dodds, P. S. 2014. An evolutionary algorithm approach to link prediction in dynamic social networks. *Journal of Computational Science* 5(5), 750–764.
- Bo, C., Tang, X.-y., Ling, Y. & LIU, Y.-s. 2014. Identifying method for opinion leaders in social network based on competency model. *Journal on Communications* 35(11), 12.
- Borgs, C., Brautbar, M., Chayes, J. & Lucier, B. 2014. Maximizing social influence in nearly optimal time. In *Proceedings of the Twenty-fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA'14*. Society for Industrial and Applied Mathematics, 946–957. <http://dl.acm.org/citation.cfm?id=2634074.2634144>
- Borodin, A., Filmus, Y. & Oren, J. 2010. Threshold models for competitive influence in social networks. In *Internet and Network Economics*, Saberi, A. (ed.), 539–550. Springer Berlin Heidelberg.
- Bozorgi, A., Samet, S., Kwisthout, J. & Wareham, T. 2017. Community-based influence maximization in social networks under a competitive linear threshold model. *Knowledge-Based Systems* 134, 149–158. <http://www.sciencedirect.com/science/article/pii/S0950705117303519>
- Braun, N. 1995. Individual thresholds and social diffusion. *Rationality and Society* 7(2), 167–182.
- Buccafurri, F., Lax, G., Nocera, A. & Ursino, D. 2012. Discovering links among social networks. In *Machine Learning and Knowledge Discovery in Databases*, Flach, P. A., De Bie, T. & Cristianini, N. (eds). Springer Berlin Heidelberg, 467–482.



- Cao, Z., Wang, L. & De Melo, G. 2018. Link prediction via subgraph embedding-based convex matrix completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 32.
- Carnes, T., Nagarajan, C., Wild, S. M. & van Zuylen, A. 2007. Maximizing influence in a competitive social network: A follower's perspective. In *Proceedings of the Ninth International Conference on Electronic Commerce, ICEC'07*, ACM, 351–360. <http://doi.acm.org/10.1145/1282100.1282167>
- Chakraborty, T., Dalmia, A., Mukherjee, A. & Ganguly, N. 2017. Metrics for community analysis: A survey. *ACM Computing Surveys (CSUR)* **50**(4), 1–37.
- Chakraborty, T., Srinivasan, S., Ganguly, N., Mukherjee, A. & Bhowmick, S. 2014. On the permanence of vertices in network communities. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1396–1405.
- Chang, B., Xu, T., Liu, Q. & Chen, E.-H. 2018. Study on information diffusion analysis in social networks and its applications. *International Journal of Automation and Computing* **15**(4), 377–401.
- Chen, J., Xu, X., Wu, Y. & Zheng, H. 2018. Gc-lstm: Graph convolution embedded lstm for dynamic link prediction. arXiv preprint arXiv:1812.04206.
- Chen, M., Mao, S. & Liu, Y. 2014. Big data: A survey. *Mobile Networks and Applications* **19**(2), 171–209.
- Chen, W. 2011. Discovering communities by information diffusion. In *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2, 1123–1132. IEEE.
- Chen, W., Collins, A., Cummings, R., Ke, T., Liu, Z., Rincon, D., Sun, X., Wei, W., Wang, Y. & Yuan, Y. 2011. Influence maximization in social networks when negative opinions may emerge and propagate. In *Proceedings of the 2011 SIAM International Conference on Data Mining (SDM'2011)*.
- Chen, W., Lu, W. & Zhang, N. 2012. Time-critical influence maximization in social networks with time-delayed diffusion process.
- Chen, W., Wang, C. & Wang, Y. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'10*, 1029–1038. ACM.
- Chen, W., Wang, Y. & Yang, S. 2009. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'09*, 199–208. ACM.
- Chen, W., Yuan, Y. & Zhang, L. 2010. Scalable influence maximization in social networks under the linear threshold model. In *2010 IEEE International Conference on Data Mining*, 88–97.
- Cheng, S., Shen, H., Huang, J., Chen, W. & Cheng, X. 2014. Imrank: Influence maximization via finding self-consistent ranking. CoRR abs/1402.3939. <http://arxiv.org/abs/1402.3939>
- Cheng, S., Shen, H., Huang, J., Zhang, G. & Cheng, X. 2013. Staticgreedy: Solving the scalability-accuracy dilemma in influence maximization. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, CIKM'13*, 509–518. ACM. <http://doi.acm.org/10.1145/2505515.2505541>
- Chhabra, A., Vashishth, V. & Sharma, D. K. 2017. A game theory based secure model against black hole attacks in opportunistic networks. In *2017 51st Annual Conference on Information Sciences and Systems (CISS)*, 1–6. IEEE.
- Chicco, D., Tötsch, N. & Jurman, G. 2021. The matthews correlation coefficient (mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining* **14**(1), 1–22.
- Clauset, A., Moore, C. & Newman, M. E. 2008. Hierarchical structure and the prediction of missing links in networks. *Nature* **453**(7191), 98–101.
- Cohen, E., Delling, D., Pajor, T. & Werneck, R. F. 2014. Sketch-based influence maximization and computation: Scaling up with guarantees. CoRR abs/1408.6282. <http://arxiv.org/abs/1408.6282>
- Daneshmand, S. M., Javari, A., Abtahi, S. E. & Jalili, M. 2015. A time-aware recommender system based on dependency network of items. *The Computer Journal* **58**(9), 1955–1966.
- Das, A. & Kempe, D. 2011. Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. *Computing Research Repository - CORR*.
- Das, S. & Biswas, A. 2021a. Community detection in social networks using local topology and information exchange. In *2021 International Conference on Intelligent Technologies (CONIT)*, 1–7. IEEE.
- Das, S. & Biswas, A. 2021b. Deployment of information diffusion for community detection in online social networks: A comprehensive review. *IEEE Transactions on Computational Social Systems*.
- Devi, K. & Tripathi, R. 2020. Information diffusion within a limited budget using node centralities and community detection. In *2020 21st Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 197–202. IEEE.
- Dey, P., Goel, K. & Agrawal, R. 2020. P-simrank: Extending simrank to scale-free bipartite networks. In *Proceedings of The Web Conference 2020*, 3084–3090.
- Divakaran, A. & Mohan, A. 2020. Temporal link prediction: A survey. *New Generation Computing* **38**(1), 213–258.
- Easley, D., Kleinberg, J. et al. 2012. Networks, crowds, and markets: Reasoning about a highly connected world. *Significance* **9**(1), 43–44.
- Eiselt, H. & Laporte, G. 1989. Competitive spatial models. *European Journal of Operational Research* **39**(3), 231–242. <http://www.sciencedirect.com/science/article/pii/0377221789901616>
- Elsweiler, D., Ruthven, I. & Jones, C. 2007. Towards memory supporting personal information management tools. *Journal of the American Society for Information Science and Technology* **58**(7), 924–946.
- Erlandsson, F., Bródka, P. & Borg, A. 2017. Seed selection for information cascade in multilayer networks. CoRR abs/1710.04391. <http://arxiv.org/abs/1710.04391>
- Fan, T., Xiong, S., Zhao, W. & Yu, T. 2019. Information spread link prediction through multi-layer of social network based on trusted central nodes. *Peer-to-Peer Networking and Applications* **12**, 1028–1040.

- Fawcett, T. 2006. An introduction to roc analysis. *Pattern Recognition Letters* **27**(8), 861–874.
- Fire, M., Tenenboim-Chekina, L., Puzis, R., Lesser, O., Rokach, L. & Elovici, Y. 2014. Computationally efficient link prediction in a variety of social networks. *ACM Transactions on Intelligent Systems and Technology (TIST)* **5**(1), 1–25.
- Fortunato, S. 2007. ‘Barth  ’ lemy m, *Resolution limit in community detection. Proceedings of the National Academy of Sciences of the United States of America* **104**(1), 36–41.
- Fortunato, S. 2010. Community detection in graphs. *Physics reports* **486**(3–5), 75–174.
- Freeman, L. C. 1978. Centrality in social networks conceptual clarification. *Social Networks* **1**(3), 215–239. <http://www.sciencedirect.com/science/article/pii/0378873378900217>
- Fukumizu, K., Bach, F. R. & Jordan, M. I. 2004. Dimensionality reduction for supervised learning with reproducing kernel hilbert spaces. *Journal of Machine Learning Research* **5**, 73–99.
- Galhotra, S., Arora, A. & Roy, S. 2016. Holistic influence maximization: Combining scalability and efficiency with opinion-aware models. In *Proceedings of the 2016 International Conference on Management of Data, SIGMOD’16*, 743–758. ACM. <http://doi.acm.org/10.1145/2882903.2882929>
- Ge, H., Huang, J., Di, C., Li, J. & Li, S. 2017. Learning automata based approach for influence maximization problem on social networks. In *2017 IEEE Second International Conference on Data Science in Cyberspace (DSC)*, 108–117.
- Gionis, A., Terzi, E. & Tsaparas, P. 2013. Opinion maximization in social networks, CoRR abs/1301.7455. <http://arxiv.org/abs/1301.7455>
- Girvan, M. & Newman, M. E. 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* **99**(12), 7821–7826.
- Goldenberg, J., Libai, B. & Muller, E. 2001. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters* **12**(3), 211–223.
- Goldenberg, J., L. 2002. Mullere. *Riding the Saddle: How Cross-market communications can Create a Major slump in sales* **66**(2), 1–16.
- Golzardi, E., Sheikahmadi, A. & Abdollahpouri, A. 2019. Detection of trust links on social networks using dynamic features. *Physica A: Statistical Mechanics and its Applications* **527**, 121269.
- Gomez-Rodriguez, M., Balduzzi, D. & Sch  lkopf, B. 2011. Uncovering the temporal dynamics of diffusion networks. CoRR abs/1105.0697. <http://arxiv.org/abs/1105.0697>
- Gong, M., Yan, J., Shen, B., Ma, L. & Cai, Q. 2016. Influence maximization in social networks based on discrete particle swarm optimization. *Information Sciences* **367–368**, 600–614. <http://www.sciencedirect.com/science/article/pii/S002002551630500X>
- Goyal, A., Lu, W. & Lakshmanan, L. V. 2011a. Celf++: Optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th International Conference Companion on World Wide Web, WWW’11*, 47–48. ACM.
- Goyal, A., Lu, W. & Lakshmanan, L. V. S. 2011b. Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining, ICDM’11*, 211–220. IEEE Computer Society.
- Granovetter, M. 1978. Threshold models of collective behavior. *American Journal of Sociology* **83**(6), 1420–1443.
- Granovetter, M. S. 1973. The strength of weak ties. *American Journal of sociology* **78**(6), 1360–1380.
- Greenhalgh, T., Stramer, K., Bratan, T., Byrne, E., Mohammad, Y. & Russell, J. 2008. Introduction of shared electronic records: Multi-site case study using diffusion of innovation theory. *BMJ* **337**.
- Grover, A. & Leskovec, J. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 855–864.
- Guille, A., Hacid, H., Favre, C. & Zighed, D. A. 2013. Information diffusion in online social networks: A survey. *ACM Sigmod Record* **42**(2), 17–28.
- Guo, J., Zhang, P., Zhou, C., Cao, Y. & Guo, L. 2013. Personalized influence maximization on social networks. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, CIKM’13*, ACM, 199–208. <http://doi.acm.org/10.1145/2505515.2505571>
- Guo, Y., Huang, Z., Kong, Y. & Wang, Q. 2021. Modularity and mutual information in networks: Two sides of the same coin. arXiv preprint arXiv:2103.02542.
- Hajibagheri, A., Alvari, H., Hamzeh, A. & Hashemi, S. 2012. Community detection in social networks using information diffusion. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 702–703. IEEE.
- Hajibagheri, A., Hamzeh, A. & Sukthankar, G. 2013. Modeling information diffusion and community membership using stochastic optimization. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 175–182.
- Han, K., Xu, C., Gui, F., Tang, S., Huang, H. & Luo, J. 2018. Discount allocation for revenue maximization in online social networks. In *Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing, Mobihoc’18*, 121–130. ACM. <http://doi.acm.org/10.1145/3209582.3209595>
- Hargittai, E. & Walejko, G. 2008. The participation divide: Content creation and sharing in the digital age. *Information, Community and Society* **11**(2), 239–256.
- He, L., Guo, W., Chen, Y., Guo, K. & Zhuang, Q. 2021. Discovering overlapping communities in dynamic networks based on cascade information diffusion. *IEEE Transactions on Computational Social Systems* **9**(3), 794–806.
- He, X. & Kempe, D. 2014. Stability of influence maximization. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD’14*, 1256–1265. ACM. <http://doi.acm.org/10.1145/2623330.2623746>
- He, X. & Kempe, D. 2016. Robust influence maximization. CoRR abs/1602.05240. <http://arxiv.org/abs/1602.05240>



- Jiang, C., Chen, Y. & Liu, K. R. 2014. Graphical evolutionary game for information diffusion over social networks. *IEEE Journal of Selected Topics in Signal Processing* **8**(4), 524–536.
- Jiang, Q., Song, G., Cong, G., Wang, Y., Si, W. & Xie, K. 2011. Simulated annealing based influence maximization in social networks. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI'11*, 127–132. AAAI Press. <http://dl.acm.org/citation.cfm?id=2900423.2900443>
- Jung, K., Heo, W. & Chen, W. 2012. Irie: Scalable and robust influence maximization in social networks. In *Proceedings of the 2012 IEEE 12th International Conference on Data Mining, ICDM'12*, 918–923. IEEE Computer Society.
- Kalantari, H., Ghazanfari, M., Fathian, M. & Shahanaghi, K. 2020. Multi-objective optimization model in a heterogeneous weighted network through key nodes identification in overlapping communities. *Computers & Industrial Engineering* **144**, 106413.
- Kashima, H. & Abe, N. 2006. A parameterized probabilistic model of network evolution for supervised link prediction. In *Sixth International Conference on Data Mining (ICDM'06)*, 340–349. IEEE.
- Katz, L. 1953. A new status index derived from sociometric analysis. *Psychometrika* **18**(1), 39–43.
- Kempe, D., Kleinberg, J. & Tardos, E. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'03*, 137–146. ACM.
- Kempe, D., Kleinberg, J. & Tardos, É. 2005. Influential nodes in a diffusion model for social networks. In *Automata, Languages and Programming*, Caires, L., Italiano, G. F., Monteiro, L., Palamidessi, C. & Yung, M. (eds). Springer Berlin Heidelberg, 1127–1138.
- Kermack, W. & McKendrick, A. 1991. Contributions to the mathematical theory of epidemics–i. *Bulletin of Mathematical Biology* **53**(1), 33–55. <http://www.sciencedirect.com/science/article/pii/S0092824005800400>
- Kermack, W. O. & McKendrick, A. G. 1927. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **115**(772), 700–721. <http://www.jstor.org/stable/94815>
- Kermack, W. O. & McKendrick, A. G. 1932. Contributions to the mathematical theory of epidemics. ii.—the problem of endemicity. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **138**(834), 55–83.
- Khan, A., Zehnder, B. & Kossmann, D. 2016. Revenue maximization by viral marketing: A social network host's perspective. In *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, 37–48.
- Khelil, A., Becker, C., Tian, J. & Rothmel, K. 2002. An epidemic model for information diffusion in manets. In *Proceedings of the 5th ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems*, 54–60.
- Kim, C. & Galliers, R. D. 2004. Toward a diffusion model for internet systems. *Internet Research* **14**(2), 155–166.
- Kim, J., Kim, S. K. & Yu, H. 2013. Scalable and parallelizable processing of influence maximization for large-scale social networks? In *2013 IEEE 29th International Conference on Data Engineering (ICDE)*, 266–277.
- Kimura, M. & Saito, K. 2006. Tractable models for information diffusion in social networks. In *Knowledge Discovery in Databases: PKDD 2006*, Fürnkranz, J., Scheffer, T. & Spiliopoulou, M. (eds). Springer Berlin Heidelberg, 259–271.
- Koren, Y., Bell, R. & Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer* **42**(8), 30–37.
- Kuhnle, A., Alim, M. A., Li, X., Zhang, H. & Thai, M. T. 2018. Multiplex influence maximization in online social networks with heterogeneous diffusion models. *IEEE Transactions on Computational Social Systems* **5**, 418–429.
- Kumar, A., Mishra, S., Singh, S. S., Singh, K. & Biswas, B. 2019a. Link prediction in complex networks based on significance of higher-order path index (shopi). *Physica A: Statistical Mechanics and its Applications*, 123790. <http://www.sciencedirect.com/science/article/pii/S0378437119321107>
- Kumar, A., Singh, S. S., Singh, K. & Biswas, B. 2019b. Level-2 node clustering coefficient-based link prediction. *Applied Intelligence* **49**(7), 2762–2779.
- Kumar, A., Singh, S. S., Singh, K. & Biswas, B. 2020. Link prediction techniques, applications, and performance: A survey. *Physica A: Statistical Mechanics and its Applications*, 124289. <http://www.sciencedirect.com/science/article/pii/S0378437120300856>
- Kumar, M., Mishra, S., Singh, S. S. & Biswas, B. 2024. Community-enhanced link prediction in dynamic networks. *ACM Transactions on the Web* **18**(2). <https://doi.org/10.1145/3580513>
- Kundu, S., Murthy, C. A. & Pal, S. K. 2011. A new centrality measure for influence maximization in social networks. In *Pattern Recognition and Machine Intelligence*, Kuznetsov, S. O., Mandal, D. P., Kundu, M. K. & Pal, S. K. (eds). Springer Berlin Heidelberg, 242–247.
- Lee, J. & Chung, C. 2015. A query approach for influence maximization on specific users in social networks. *IEEE Transactions on Knowledge and Data Engineering* **27**(2), 340–353.
- Lee, W., Kim, J. & Yu, H. 2012. Ct-ic: Continuously activated and time-restricted independent cascade model for viral marketing. In *2012 IEEE 12th International Conference on Data Mining*, 960–965.
- Lei, S., Maniu, S., Mo, L., Cheng, R. & Senellart, P. 2015. Online influence maximization (extended version). CoRR abs/1506.01188. <http://arxiv.org/abs/1506.01188>
- Leicht, E. A., Holme, P. & Newman, M. E. 2006. Vertex similarity in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* **73**(2), 026120.
- Leskovec, J., Adamic, L. & Huberman, B. A. 2007. The dynamics of viral marketing. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, 228–237.
- Leskovec, J., Huttenlocher, D. & Kleinberg, J. 2010. Predicting positive and negative links in online social networks. In *Proceedings of the 19th International Conference on World Wide Web*, 641–650.

- Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J. & Glance, N. 2007. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'07*, 420–429. ACM.
- Li, C.-T., Lin, S.-D. & Shan, M.-K. 2012. Influence propagation and maximization for heterogeneous social networks. In *Proceedings of the 21st International Conference on World Wide Web, WWW'12 Companion*, ACM, 559–560. <http://doi.acm.org/10.1145/2187980.2188126>
- Li, D., Xu, Z.-M., Chakraborty, N., Gupta, A., Sycara, K. & Li, S. 2014. Polarity related influence maximization in signed social networks. *PLOS ONE* **9**, 1–12. <https://doi.org/10.1371/journal.pone.0102199>
- Li, D., Zhang, Y., Xu, Z., Chu, D. & Li, S. 2016. Exploiting information diffusion feature for link prediction in sina weibo. *Scientific Reports* **6**(1), 1–8.
- Li, G., Chen, S., Feng, J., Tan, K.-I. & Li, W.-s. 2014. Efficient location-aware influence maximization. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, SIGMOD'14*, 87–98. ACM. <http://doi.acm.org/10.1145/2588555.2588561>
- Li, H., Bhowmick, S. S. & Sun, A. 2011. Casino: Towards conformity-aware social influence analysis in online social networks. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM'11*, ACM, 1007–1012. <http://doi.acm.org/10.1145/2063576.2063721>
- Li, H., Bhowmick, S. S. & Sun, A. 2013. Cinema: Conformity-aware greedy algorithm for influence maximization in online social networks. In *Proceedings of the 16th International Conference on Extending Database Technology, EDBT'13*, 323–334. ACM. <http://doi.acm.org/10.1145/2452376.2452415>
- Li, L., Liu, Y., Zhou, Q., Yang, W. & Yuan, J. 2020. Targeted influence maximization under a multifactor-based information propagation model. *Information Sciences* **519**, 124–140. <http://www.sciencedirect.com/science/article/pii/S0020025520300438>
- Li, S., Huang, J., Zhang, Z., Liu, J., Huang, T. & Chen, H. 2018. Similarity-based future common neighbors model for link prediction in complex networks. *Scientific Reports* **8**(1), 1–11.
- Li, W., Liu, W., Chen, T., Qu, X., Fang, Q. & Ko, K.-I. 2017. Competitive profit maximization in social networks. *Theoretical Computer Science* **694**, 1–9. <http://www.sciencedirect.com/science/article/pii/S0304397517305388>
- Li, Y., Chen, W., Wang, Y. & Zhang, Z. 2011. Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. CoRR abs/1111.4729. <http://arxiv.org/abs/1111.4729>
- Li, Y., Zhang, D. & Tan, K.-L. 2015. Real-time targeted influence maximization for online advertisements. *Proceedings of the VLDB Endowment* **8**(10), 1070–1081.
- Lin, T. Y., Ohsuga, S., Liau, C.-J. & Hu, X. 2005. *Foundations and Novel Approaches in Data Mining*, 9. Springer Science & Business Media.
- Liu, B., Cong, G., Zeng, Y., Xu, D. & Chee, Y. M. 2014. Influence spreading path and its application to the time constrained social influence maximization problem and beyond. *IEEE Transactions on Knowledge and Data Engineering* **26**(8), 1904–1917.
- Liu, Q., Xiang, B., Chen, E., Xiong, H., Tang, F. & Yu, J. X. 2014. Influence maximization over large-scale social networks: A bounded linear approach. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM'14*, 171–180. ACM. <http://doi.acm.org/10.1145/2661829.2662009>
- Liu, S. & Wang, S. 2016. Trajectory community discovery and recommendation by multi-source diffusion modeling. *IEEE Transactions on Knowledge and Data Engineering* **29**(4), 898–911.
- Loeliger, H.-A., Dauwels, J., Hu, J., Korl, S., Ping, L. & Kschischang, F. R. 2007. The factor graph approach to model-based signal processing. *Proceedings of the IEEE* **95**(6), 1295–1322.
- Lü, L., Jin, C.-H. & Zhou, T. 2009. Similarity index based on local paths for link prediction of complex networks. *Physical Review E-Statistical, Nonlinear, and Soft Matter Physics* **80**(4), 046122.
- Lu, W. & Lakshmanan, L. V. S. 2012. Profit maximization over social networks, CoRR abs/1210.4211.
- Ma, C., Zhu, C., Fu, Y., Zhu, H., Liu, G. & Chen, E. 2017. Social user profiling: A social-aware topic modeling perspective. In *International Conference on Database Systems for Advanced Applications*, 610–622. Springer.
- Mack, P. 1985. Diffusion of innovations by everett m. rogers. *Technology and Culture* **26**(1), 109–110.
- Magal, P. & Ruan, S. 2014. Susceptible-infectious-recovered models revisited: From the individual level to the population level. *Mathematical Biosciences* **250**, 26–40.
- Manapat, M. L. & Rand, D. G. 2012. Delayed and inconsistent information and the evolution of trust. *Dynamic Games and Applications* **2**(4), 401–410.
- Mehmood, Y., Bonchi, F. & Garca-Soriano, D. 2016. Spheres of influence for more effective viral marketing. In *Proceedings of the 2016 International Conference on Management of Data, SIGMOD'16*, 711–726. ACM. <http://doi.acm.org/10.1145/2882903.2915250>
- Meng, Y., Yi, Y., Xiong, F. & Pei, C. 2019. Txonehop approach for dynamic influence maximization problem. *Physica A: Statistical Mechanics and its Applications* **515**, 575–586. <http://www.sciencedirect.com/science/article/pii/S0378437118312937>
- Mishra, S., Singh, S. S., Kumar, A. & Biswas, B. 2022. Mnerlp-mul: Merged node and edge relevance based link prediction in multiplex networks. *Journal of Computational Science* **60**, 101606. <https://www.sciencedirect.com/science/article/pii/S1877750322000369>
- Mishra, S., Singh, S. S., Mishra, S. & Biswas, B. 2024. Multi-objective based unbiased community identification in dynamic social networks. *Computer Communications* **214**, 18–32. <https://www.sciencedirect.com/science/article/pii/S0140366423004188>
- Morstatter, F., Wu, L., Nazer, T. H., Carley, K. M. & Liu, H. 2016. A new approach to bot detection: Striking the balance between precision and recall. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 533–540, IEEE.

- Mozafari, N. & Hamzeh, A. 2015. An enriched social behavioural information diffusion model in social networks. *Journal of Information Science* **41**(3), 273–283.
- Muhuri, S., Chakraborty, S. & Chakraborty, S. N. 2018. Extracting social network and character categorization from bengali literature. *IEEE Transactions on Computational Social Systems* **5**(2), 371–381.
- Muhuri, S. & Mukhopadhyay, D. 2021. A hypergraph clustering-based technique for detecting fake news from broadcasting network. In *2021 Asian Conference on Innovation in Technology (ASIANCON)*, 1–5. IEEE.
- Nasiri, E., Berahmand, K., Rostami, M. & Dabiri, M. 2021. A novel link prediction algorithm for protein-protein interaction networks by attributed graph embedding. *Computers in Biology and Medicine* **137**, 104772.
- Nazemian, A. & Taghiyareh, F. 2012. Influence maximization in independent cascade model with positive and negative word of mouth. In *6th International Symposium on Telecommunications (IST)*, 854–860. IEEE.
- Nguyen, D. T., Das, S. & Thai, M. T. 2013. Influence maximization in multiple online social networks. In *2013 IEEE Global Communications Conference (GLOBECOM)*, 3060–3065.
- Nguyen, H. T., Dinh, T. N. & Thai, M. T. 2016. Cost-aware targeted viral marketing in billion-scale networks. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, 1–9.
- Nguyen, H. T., Thai, M. T. & Dinh, T. N. 2016. Stop-and-stare: Optimal sampling algorithms for viral marketing in billion-scale networks. In *Proceedings of the 2016 International Conference on Management of Data, SIGMOD'16*, 695–710. ACM. <http://doi.acm.org/10.1145/2882903.2915207>
- Obregon, J., Song, M. & Jung, J.-Y. 2019. Inflow: Mining information flow based on user community in social networking services. *IEEE Access* **7**, 48024–48036.
- Ohsaka, N., Akiba, T., Yoshida, Y. & Kawarabayashi, K.-I. 2014. Fast and accurate influence maximization on large networks with pruned monte-carlo simulations. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, AAAI'14*, 138–144. AAAI Press. <http://dl.acm.org/citation.cfm?id=2893873.2893897>
- Ohsaka, N., Akiba, T., Yoshida, Y. & Kawarabayashi, K.-i. 2016. Dynamic influence analysis in evolving networks. *Proceedings of the VLDB Endowment* **9**(12), 1077–1088. <https://doi.org/10.14778/2994509.2994525>
- Page, L., Brin, S., Motwani, R. & Winograd, T. 1999. The pagerank citation ranking: Bringing order to the web, Technical Report 1999-66, Stanford InfoLab. Previous number = SIDL-WP-1999-0120. <http://ilpubs.stanford.edu:8090/422/>
- Parshani, R., Carmi, S. & Havlin, S. 2010. Epidemic threshold for the susceptible-infectious-susceptible model on random networks. *Physical Review Letters* **104**(25), 258701.
- Pathak, N., Banerjee, A. & Srivastava, J. 2010. A generalized linear threshold model for multiple cascades. In *2010 IEEE International Conference on Data Mining*, 965–970. IEEE.
- Peccli, A., Giovanini, B., Pacheco, C. C., Moreira, C., Ferreira, F., Tosta, F., Tesolin, J., Dias, M. V., Lima Filho, S. P., Cavalcanti, M. C. et al. 2015. Dimensionality reduction for supervised learning in link prediction problems. In *ICEIS (I)*, 295–302.
- Perozzi, B., Al-Rfou, R. & Skiena, S. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 701–710.
- Rajamma, R. K., Zolfagharian, M. A. & Pelton, L. E. 2011. Dimensions and outcomes of b2b relational exchange: A meta-analysis. *Journal of Business & Industrial Marketing* **26**(2), 104–114.
- Ramezani, M., Khodadadi, A. & Rabiee, H. R. 2018. Community detection using diffusion information. *ACM Transactions on Knowledge Discovery from Data (TKDD)* **12**(2), 1–22.
- Razaque, A., Rizvi, S., Almiani, M., Al Rahayfeh, A. et al. 2019. State-of-art review of information diffusion models and their impact on social network vulnerabilities. *Journal of King Saud University-Computer and Information Sciences* **34**(1), 1275–1294.
- Richardson, M. & Domingos, P. 2002. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'02*, 61–70. ACM. <http://doi.acm.org/10.1145/775047.775057>
- Rodriguez, M. G., Balduzzi, D. & Schölkopf, B. 2011. Uncovering the temporal dynamics of diffusion networks. arXiv preprint arXiv:1105.0697.
- Ruan, Y., Fuhry, D. & Parthasarathy, S. 2013. Efficient community detection in large networks using content and links. In *Proceedings of the 22nd International Conference on World Wide Web*, 1089–1098.
- Saito, K., Kimura, M., Ohara, K. & Motoda, H. 2012. Efficient discovery of influential nodes for sis models in social networks. *Knowledge and Information Systems* **30**(3), 613–635.
- Saito, K., Nakano, R. & Kimura, M. 2008. Prediction of information diffusion probabilities for independent cascade model. In *International conference on knowledge-based and intelligent information and engineering systems*, 67–75. Springer.
- Saito, T. & Rehmsmeier, M. 2015. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* **10**(3), e0118432.
- Sattari, M. & Zamanifar, K. 2018. A cascade information diffusion based label propagation algorithm for community detection in dynamic social networks. *Journal of Computational Science* **25**, 122–133.
- Schelling, T. C. 2006. *Micromotives and Macrobehavior*, WW Norton & Company.
- Scholz, C., Atzmueller, M., Barrat, A., Cattuto, C. & Stumme, G. 2013. New insights and methods for predicting face-to-face contacts. In *Seventh International AAAI Conference on Weblogs and Social Media*.
- Schütz, G. M., Brandaut, M. & Trimper, S. 2008. Exact solution of a stochastic susceptible-infectious-recovered model. *Physical Review E* **78**(6), 061132.
- Seddiki, M. S. & Frikha, M. 2012. A non-cooperative game theory model for bandwidth allocation in network virtualization. In *2012 15th International Telecommunications Network Strategy and Planning Symposium (NETWORKS)*, 1–6. IEEE.

- Serban, I., Sordoni, A., Bengio, Y., Courville, A. & Pineau, J. 2016. Building end-to-end dialogue systems using generative hierarchical neural network models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 30.
- Shen, C., Nishide, R., Piumarta, I., Takada, H. & Liang, W. 2015. Influence maximization in signed social networks. In *Web Information Systems Engineering – WISE 2015*, Wang, J., Cellary, W., Wang, D., Wang, H., Chen, S.-C., Li, T. & Zhang, Y. (eds). Springer International Publishing, 399–414.
- Shen, K., Song, L., Yang, X. & Zhang, W. 2010. A hierarchical diffusion algorithm for community detection in social networks. In *2010 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, 276–283. IEEE.
- Sheshar, S., Srivastva, S. D., Verma, M. & Singh, J. 2021. Influence maximization frameworks, performance, challenges and directions on social network: A theoretical study. *Journal of King Saud University-Computer and Information Sciences* **34**(9), 7570–7603.
- Singh, N. et al. 2019. Improved link prediction using pca. *International Journal of Analysis and Applications* **17**(4), 578–585.
- Singh, S. S., Kumar, A., Singh, K. & Biswas, B. 2019a. C2im: Community based context-aware influence maximization in social networks. *Physica A: Statistical Mechanics and its Applications* **514**, 796–818. <http://www.sciencedirect.com/science/article/pii/S0378437118312822>
- Singh, S. S., Kumar, A., Singh, K. & Biswas, B. 2019b. Lapso-im: A learning-based influence maximization approach for social networks. *Applied Soft Computing*, 105554. <http://www.sciencedirect.com/science/article/pii/S1568494619303345>
- Singh, S. S., Kumar, A., Singh, K. & Biswas, B. 2020. Im-ss0: Maximizing influence in social networks using social spider optimization. *Concurrency and Computation: Practice and Experience* **32**(2), e5421. e5421 cpe.5421. <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.5421>
- Singh, S. S., Mishra, S., Kumar, A. & Biswas, B. 2020. Clp-id: Community-based link prediction using information diffusion. *Information Sciences* **514**, 402–433. <http://www.sciencedirect.com/science/article/pii/S0020025519310734>
- Singh, S. S., Muhuri, S., Mishra, S., Srivastava, D., Shakya, H. K. & Kumar, N. 2024. Social network analysis: A survey on process, tools, and application. *ACM Computing Surveys* **56**(8). <https://doi.org/10.1145/3648470>
- Singh, S. S., Singh, K., Kumar, A. & Biswas, B. 2019a. Aco-im: Maximizing influence in social networks using ant colony optimization. *Soft Computing*, 1–23.
- Singh, S. S., Singh, K., Kumar, A. & Biswas, B. 2019b. Mim2: Multiple influence maximization across multiple social networks. *Physica A: Statistical Mechanics and its Applications* **526**, 120902. <http://www.sciencedirect.com/science/article/pii/S037843711930500X>
- Singh, S. S., Singh, K., Kumar, A., Shakya, H. K. & Biswas, B. 2019. A survey on information diffusion models in social networks. In *Advanced Informatics for Computing Research*, Luhach, A. K., Singh, D., Hsiung, P.-A., Hawari, K. B. G., Lingras, P. & Singh, P. K. (eds). Springer Singapore, 426–439.
- Singh, S. S., Srivastava, V., Kumar, A., Tiwari, S., Singh, D. & Lee, H.-N. 2023. Social network analysis: A survey on measure, structure, language information analysis, privacy, and applications. *ACM Transactions on Asian and Low-Resource Language Information Processing* **22**(5). <https://doi.org/10.1145/3539732>
- Singh, S. S., Srivastava, D., Kumar, A. & Srivastava, V. 2022. Flp-id: Fuzzy-based link prediction in multiplex social networks using information diffusion perspective. *Knowledge-Based Systems* **248**, 108821. <https://www.sciencedirect.com/science/article/pii/S0950705122003859>
- Singh, S. S., Srivastava, D., Verma, M. & Singh, J. 2021. Influence maximization frameworks, performance, challenges and directions on social network: A theoretical study. *Journal of King Saud University - Computer and Information Sciences* **34**(9), 7570–7603. <https://www.sciencedirect.com/science/article/pii/S1319157821002123>
- Singh, S. S., Muhuri, S. & Srivastava, V. 2024. B+ tree-inspired community-based link prediction in dynamic social networks. *Arabian Journal for Science and Engineering* **49**, 4039–4060.
- Song, D. & Meyer, D. A. 2015. Recommending positive links in signed social networks by optimizing a generalized auc. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- Stanley, N., Bonacci, T., Kwitt, R., Niethammer, M. & Mucha, P. J. 2019. Stochastic block models with multiple continuous attributes. *Applied Network Science* **4**(1), 1–22.
- Su, S., Li, X., Cheng, X. & Sun, C. 2018. Location-aware targeted influence maximization in social networks. *Journal of the Association for Information Science and Technology* **69**(2), 229–241. <http://https://doi.org/10.1002/asi.23931>
- Suganya, T., Thennammai, S. & Velusamy, R. 2017. Unique user identification across multiple social network. *International Journal of Research in Marketing* **8**, 137–142.
- Sun, H., Gao, X., Chen, G., Gu, J. & Wang, Y. 2016. Multiple influence maximization in social networks. In *Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication, IMCOM'16*, 44:1–44:8. ACM.
- Sun, Z., Sheng, J., Wang, B., Ullah, A. & Khawaja, F. 2020. Identifying communities in dynamic networks using information dynamics. *Entropy* **22**(4), 425.
- Sun, Z., Wang, B., Sheng, J., Yu, Z. & Shao, J. 2018. Overlapping community detection based on information dynamics. *IEEE Access* **6**, 70919–70934.
- Susarla, A., Oh, J.-H. & Tan, Y. 2012. Social networks and the diffusion of user-generated content: Evidence from youtube. *Information Systems Research* **23**(1), 23–41.
- Sviridenko, M. 2004. A note on maximizing a submodular set function subject to a knapsack constraint. *Operations Research Letters* **32**(1), 41–43.
- Tang, J., Tang, X. & Yuan, J. 2017. Towards profit maximization for online social network providers. CoRR abs/1712.08963.
- Tang, J., Tang, X. & Yuan, J. 2018. Profit maximization for viral marketing in online social networks: Algorithms and analysis. *IEEE Transactions on Knowledge and Data Engineering* **30**(6), 1095–1108.
- Tang, J., Wu, S. & Sun, J. 2013. Confluence: Conformity influence in large social networks.



- Tang, Y., Shi, Y. & Xiao, X. 2015. Influence maximization in near-linear time: A martingale approach. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, SIGMOD'15*, ACM, 1539–1554. <http://doi.acm.org/10.1145/2723372.2723734>
- Tang, Y., Xiao, X. & Shi, Y. 2014. Influence maximization: Near-optimal time complexity meets practical efficiency. In *SIGMOD Conference*.
- Tong, G., Wu, W., Tang, S. & Du, D. 2017. Adaptive influence maximization in dynamic social networks. *IEEE/ACM Transactions on Networking* **25**(1), 112–125.
- Van Lierde, H., Chen, G. & Chow, T. W. 2019. Scalable spectral clustering for overlapping community detection in large-scale networks. *IEEE Transactions on Knowledge and Data Engineering* **32**, 754–767.
- Varshney, D., Kumar, S. & Gupta, V. 2014. Modeling information diffusion in social networks using latent topic information. In *Intelligent Computing Theory*, Huang, D.-S., Bevilacqua, V. & Premaratne, P. (eds), 137–148. Springer International Publishing.
- Varshney, D., Kumar, S. & Gupta, V. 2017. Predicting information diffusion probabilities in social networks: A Bayesian networks based approach. *Knowledge-Based Systems* **133**, 66–76.
- Wang, C., Guan, X., Qin, T. & Zhou, Y. 2015. Modelling on opinion leader's influence in microblog message propagation and its application. *Journal of Software* **26**, 1473–1485.
- Wang, C., Huang, P., Yang, D. & Chen, W. 2016. Cross-layer design of influence maximization in mobile social networks. CoRR abs/1604.02796. <http://arxiv.org/abs/1604.02796>
- Wang, P., Xu, B., Wu, Y. & Zhou, X. 2015. Link prediction in social networks: The state-of-the-art. *Science China Information Sciences* **58**(1), 1–38.
- Wang, Q., Gong, M., Song, C. & Wang, S. 2017. Discrete particle swarm optimization based influence maximization in complex networks. In *2017 IEEE Congress on Evolutionary Computation (CEC)*, 488–494.
- Wang, X., Zhang, Y., Zhang, W., Lin, X. & Chen, C. 2017. Bring order into the samples: A novel scalable method for influence maximization. *IEEE Transactions on Knowledge and Data Engineering* **29**(2), 243–256.
- Wang, Y., Cong, G., Song, G. & Xie, K. 2010. Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In *KDD*.
- Wang, Y. & Feng, X. 2009. A potential-based node selection strategy for influence maximization in a social network. In *Advanced Data Mining and Applications*, Huang, R., Yang, Q., Pei, J., Gama, J., Meng, X. & Li, X. (eds), 350–361. Springer Berlin Heidelberg.
- Wang, Y., Zhu, J. & Ming, Q. 2017. Incremental influence maximization for dynamic social networks. In *Data Science*, Zou, B., Han, Q., Sun, G., Jing, W., Peng, X. & Lu, Z. (eds). Springer Singapore, 13–27.
- Wang, Z., Lei, Y. & Li, W. 2020. Neighborhood attention networks with adversarial learning for link prediction. *IEEE Transactions on Neural Networks and Learning Systems* **32**(8), 3653–3663.
- Wang, Z., Sun, C., Xi, J. & Li, X. 2021. Influence maximization in social graphs based on community structure and node coverage gain. *Future Generation Computer Systems* **118**, 327–338.
- Weersink, A. & Fulton, M. 2020. Limits to profit maximization as a guide to behavior change. *Applied Economic Perspectives and Policy* **42**(1), 67–79. <https://onlinelibrary.wiley.com/doi/abs/10.1002/aep.13004>
- Williamson, S. A. 2016. Nonparametric network models for link prediction. *The Journal of Machine Learning Research* **17**(1), 7102–7121.
- Wu, J., Shen, J., Zhou, B., Zhang, X. & Huang, B. 2019. General link prediction with influential node identification. *Physica A: Statistical Mechanics and its Applications* **523**, 996–1007.
- Wu, X., Zhang, H., Zhao, X., Li, B. & Yang, C. 2015. Mining algorithm of microblogging opinion leaders based on user-behavior network. *Applied Computing and Informatics* **32**, 2678–2683.
- Xie, J., Kelley, S. & Szymanski, B. K. 2013. Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Computing Surveys (CSUR)* **45**(4), 43.
- Xu, T., Zhu, H., Chen, E., Huai, B., Xiong, H. & Tian, J. 2014. Learning to annotate via social interaction analytics. *Knowledge and Information Systems* **41**(2), 251–276.
- Xu, T., Zhu, H., Zhao, X., Liu, Q., Zhong, H., Chen, E. & Xiong, H. 2016. Taxi driving behavior analysis in latent vehicle-to-vehicle networks: A social influence perspective. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1285–1294.
- Yang, J., McAuley, J. & Leskovec, J. 2013. Community detection in networks with node attributes. In *2013 IEEE 13th International Conference on Data Mining*, 1151–1156. IEEE.
- Yang, W., Brenner, L. & Giua, A. 2018. Influence maximization by link activation in social networks. In *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, 1, 1248–1251.
- Yang, Y., Lichtenwalter, R. N. & Chawla, N. V. 2015. Evaluating link prediction methods. *Knowledge and Information Systems* **45**(3), 751–782.
- Yao, L., Wang, L., Pan, L. & Yao, K. 2016. Link prediction based on common-neighbors for dynamic social network. *Procedia Computer Science* **83**, 82–89.
- Yao, Y., Zhang, R., Yang, F., Tang, J., Yuan, Y. & Hu, R. 2018. Link prediction in complex networks based on the interactions among paths. *Physica A: Statistical Mechanics and its Applications* **510**, 52–67.
- Yin, D., Hong, L. & Davison, B. D. 2011. Structural link analysis and prediction in microblogs. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, 1163–1168.

- Zhan, Q., Zhang, J., Wang, S., Yu, P. S. & Xie, J. 2015. Influence maximization across partially aligned heterogeneous social networks. In *Advances in Knowledge Discovery and Data Mining*, Cao, T., Lim, E.-P., Zhou, Z.-H., Ho, T.-B., Cheung, D. & Motoda, H. (eds). Springer International Publishing, 58–69.
- Zhang, H., Dinh, T. N. & Thai, M. T. 2013. Maximizing the spread of positive influence in online social networks. In *2013 IEEE 33rd International Conference on Distributed Computing Systems*, 317–326.
- Zhang, H., Nguyen, D. T., Zhang, H. & Thai, M. T. 2016. Least cost influence maximization across multiple social networks. *IEEE/ACM Transactions on Networking* **24**(2), 929–939.
- Zhang, H., Wu, G. & Ling, Q. 2019. Distributed stochastic gradient descent for link prediction in signed social networks. *EURASIP Journal on Advances in Signal Processing* **2019**(1), 1–11.
- Zhang, J., Tang, J., Zhuang, H., Wing-ki Leung, C. & Li, J. 2014. Role-aware conformity influence modeling and analysis in social networks.
- Zhang, Y. 2015. Influence maximization on multi-phased multi-layered network.
- Zhang, Y., Lyu, T. & Zhang, Y. 2018. Cosine: Community-preserving social network embedding from information diffusion cascades. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 32.
- Zhao, X., Xu, T., Liu, Q. & Guo, H. 2016. Exploring the choice under conflict for social event participation. In *International Conference on Database Systems for Advanced Applications*, 396–411. Springer.
- Zhou, C., Zhang, P., Zang, W. & Guo, L. 2015. On the upper bounds of spread for greedy algorithms in social network influence maximization. *IEEE Transactions on Knowledge and Data Engineering* **27**(10), 2770–2783.
- Zhou, T., Cao, J., Liu, B., Xu, S., Zhu, Z. & Luo, J. 2015. Location-based influence maximization in social networks. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM'15*, ACM, New York, NY, USA, pp. 1211–1220. <http://doi.acm.org/10.1145/2806416.2806462>
- Zhuang, H., Sun, Y., Tang, J., Zhang, J. & Sun, X. 2013. Influence maximization in dynamic social networks. In *2013 IEEE 13th International Conference on Data Mining*, 1313–1318.