

Review

# A Survey on Information Diffusion in Online Social Networks: Models and Methods

Mei Li \*, Xiang Wang, Kai Gao and Shanshan Zhang

School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China; wangxiang@hebust.edu.cn (X.W.); gaokai@hebust.edu.cn (K.G.); zshanshanmiss@gmail.com (S.Z.)

\* Correspondence: limei@hebust.edu.cn; Tel.: +86-139-3317-5921

Received: 16 August 2017; Accepted: 22 September 2017; Published: 29 September 2017

**Abstract:** By now, personal life has been invaded by online social networks (OSNs) everywhere. They intend to move more and more offline lives to online social networks. Therefore, online social networks can reflect the structure of offline human society. A piece of information can be exchanged or diffused between individuals in social networks. From this diffusion process, lots of latent information can be mined. It can be used for market predicting, rumor controlling, and opinion monitoring among other things. However, the research of these applications depends on the diffusion models and methods. For this reason, we survey various information diffusion models from recent decades. From a research process view, we divide the diffusion models into two categories—explanatory models and predictive models—in which the former includes epidemics and influence models and the latter includes independent cascade, linear threshold, and game theory models. The purpose of this paper is to investigate the research methods and techniques, and compare them according to the above categories. The whole research structure of the information diffusion models based on our view is given. There is a discussion at the end of each section, detailing related models that are mentioned in the literature. We conclude that these two models are not independent, they always complement each other. Finally, the issues of the social networks research are discussed and summarized, and directions for future study are proposed.

**Keywords:** social networks; information diffusion; theoretical diffusion model; explanatory models; predictive models

## 1. Introduction

The term “social networks” (SNS) was first used by Barnes [1] in the Human Relations Journal in 1954. Social networks originated from e-mail and are now the most widely used applications. With the evolution of social networks, there are more and more new platforms, e.g., Facebook and Flickr in 2004, YouTube in 2005, Twitter in 2006, and Sina Micro-blog in 2009. The ways in which people obtain information have changed. In the past, individuals were passive receivers of information yet now they are its active publishers and communicators.

When a piece of information flows from one individual or community to another in a network, then an information diffusion process—also known as information propagation, information spread, or information dissemination—has occurred. Much research effort has been put into analyzing information diffusion, with most studies investigating which factors affect information diffusion, which information diffuses most quickly, and how information is disseminated [2,3]. These questions are answered using information diffusion models and other methods, which play an important role in understanding the diffusion phenomenon. We do not know why the information flows to this direction in social networks, although we have seen the advantages of a social network in information diffusion. If, using information diffusion models, we can work out who the important users are,

and which factors are influencing the information diffusion process, then we can better understand this phenomenon. A good performance model is very important for understanding how to predict and influence information diffusion, and has significant reference value to various applications, e.g., rumor controlling [4–6], behavior analysis [7], gauging public opinion, the study of psychological phenomena [8], and for resource allocation in public health care systems [9].

Although some surveys exist that illustrate information diffusion, they each refer to only one issue or other views [10,11] in social networks. Guille et al. [12] analyze topic detection, information diffusion modeling, and identification methods for influential spreaders. Dong et al. [13] divide current models into theoretical diffusion models and information diffusion cascade models based on the research data, some of which is from a real social network and some of which is not. Wani et al. [14] discuss the parameters which affect information diffusion. These parameters include the tie-strength, homophily, communities, opinion, user roles, and topics. Kumaran et al. [15] compare the methods, algorithms, and techniques for influence spreader detection. Their research view is related to influence diffusion. Dey et al. [16] survey the related references of topic analysis, information diffusion, and the properties of social connections in the context of online social networks. Although the researchers' views are different from the literature analysis, all of them are different facets in information diffusion research. However, there is not enough illustration to understand the information diffusion process. From our investigation of the literature published over the past few years, we conclude that basic information diffusion issues can be classified as a “3W issue”, the 3 Ws being “What”, “Why” and “Where”.

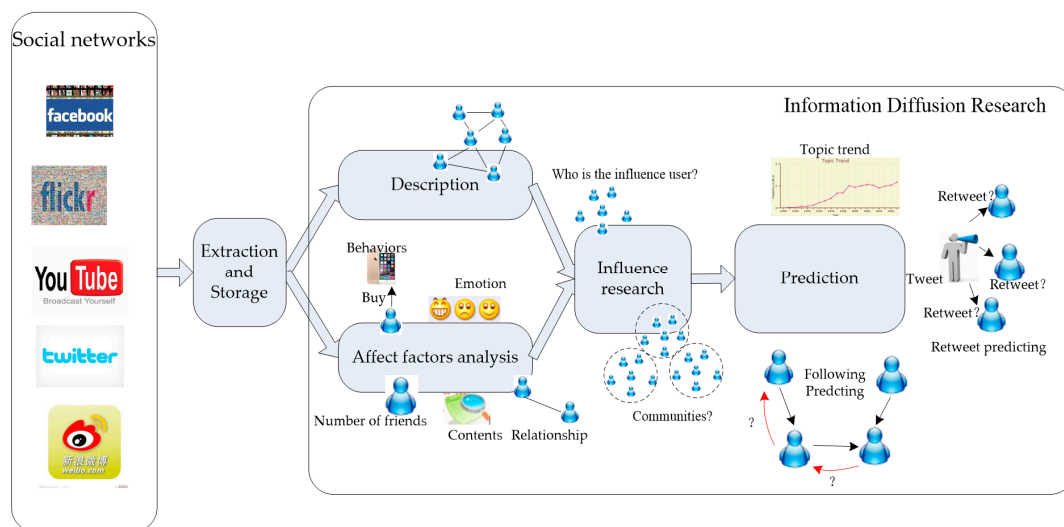
The first W, “What”, refers to the question “what latent information is there to be found in social networks?” For example, a large volume of consumer data will contain some interesting findings, e.g., how an individual's consumption habits relate to their profession.

The second W is “Why”, referring to the question “why has the information propagated in this way?” When we see data analysis visualizations, we always want to know which factors have affected the visualization result, e.g., interactions or information.

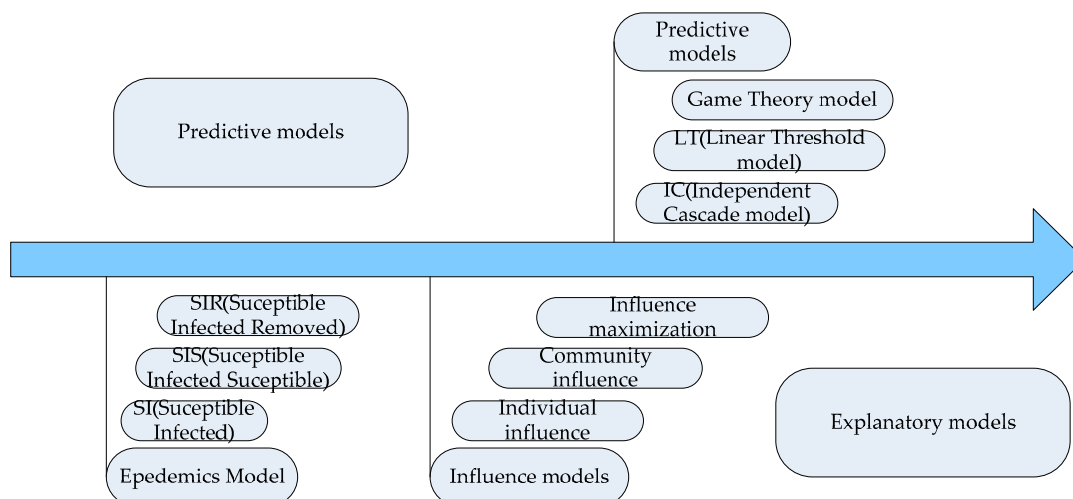
The third W is “Where”, referring to the question “where will the information be diffused to in the future?” For example, user A has two friends, B and C, in a social network. Both B and C are influential users. If A posts some information, both B and C will each have a different perspective on the information, which will influence how they respond and whether they further propagate it through the network. These factors aid in the understanding of which node will be the destination for the future diffusion of the information, i.e., the prediction of information diffusion.

The “What” and “Why” questions involve explaining aspects of information diffusion and the “Where” question is related to prediction. Actually, the “3W issue” represents the different stages in information diffusion research. Therefore, the research roadmap on information diffusion can be described as depicted in Figure 1. After the data extraction and storage, the first basic job is to describe the diffusion process and analyze the influence factors that influence information diffusion. From this analysis, the future diffusion process can then be predicted.

Based on this roadmap, the literature related to these issues can be classified into two categories, i.e., explanatory models and predictive models. Then, the most widely used basic models and their scalable models in these two categories are reviewed, analyzed, discussed, and compared. Finally, future challenges and methods for research into information diffusion are presented. The information diffusion models based on our view are depicted as Figure 2. This is the basis for this paper. Actually, these two models are not independent completely from one another, although we define them separately. The comparison, along with the review of the literature, shows that these two models always complement each other.



**Figure 1.** Research Roadmap on Information Diffusion.



**Figure 2.** Categorization of Information Diffusion Models.

The remainder of this paper is organized as follows. Section 2 describes and compares the explanatory models. In Section 3, we describe and compare the predictive models. Future challenges are discussed in Section 4, and finally, we discuss and conclude the paper in Section 5.

## 2. Explanatory Models

### 2.1. Aims of the Explanatory Models

Information is spread by way of interactions between different individuals in society. These individuals can be regarded as nodes in social networks. A node in a social network is an abstract representation of a user in “real” society. The interactions between two users may be regarded as relations, which are represented by edges running between two nodes in social networks. Therefore, a real social group can be mapped by a huge social network and a piece of information can be disseminated by these nodes within it. This raises many questions about the information diffusion process, such as: What are the main factors that affect information diffusion? Which node has the most influence? Why does the information diffuse the way it does? For example, some nodes refuse to accept information, some refuse to spread information, and some both accept and spread information [17].

Different groups of nodes also have different characteristics: some are homogenous and some are heterogeneous [18]. The explanatory models presented in this paper aim to examine the information diffusion process and elucidate the factors that affect it in an attempt to explain this phenomenon.

## 2.2. The Basic Epidemics Model

The information diffusion process can be considered in the same way as an epidemic spread process. In epidemics transmission, there are both users infected with pathogens and users who are susceptible to the pathogens. The virus can spread from infected users to susceptible users, and information can be diffused from communicators to recipients in a similar fashion. To investigate information diffusion, it makes sense to learn from the basic epidemic models. In the compartment model of epidemics, the basic models are SI (Susceptible Infected) model, SIS (Susceptible Infected Susceptible) model, SIR (Susceptible Infected Removed) model and SIRS (Susceptible Infected Removed Susceptible) model, and these are described below.

### 2.2.1. The SI Model

Pastoratorras [19] proposed the SI model for complex networks in 2001. The impact of birth and death rates on the total number of people is not regarded. The model assumes that the total number of people is  $N$ .  $N$  is divided into two categories:  $S$  (susceptible) and  $I$  (infected). At time  $t$ ,  $s(t)$  represents the susceptible proportion of the total population,  $i(t)$  represents the infected proportion, and  $\lambda$  represents the daily contact rate, which means the proportion of the susceptible users infected by infected users in the total population, where  $s(t) + i(t) = 1$ . Hence,  $Ns(t) + Ni(t) = N$ . From this, we can see that there will be  $\lambda s(t)$  susceptible users infected. If the infected users are  $Ni(t)$ , then there will be  $\lambda s(t)Ni(t)$  susceptible users infected a day.  $\lambda Nsi$  represents the increment of the increase in the number of patients per day; that is,  $N \frac{di}{dt} = \lambda Nsi$ ,  $\frac{di}{dt} = \lambda si$ , and  $s + i = 1$ . At time  $t = 0$ , the proportion of patients is  $i_0$ , and the SI model can be described by Equations (1) and (2).

$$\frac{di}{dt} = \lambda i(1 - i), \quad (1)$$

$$i(0) = i_0. \quad (2)$$

### 2.2.2. The SIS Model

The SI model is not practical, as it does not allow for infected users to be cured after having been infected. The SIS model addresses this issue. Newman [20,21] and Gross et al. [22] have done some research into complex and adaptive networks. The parameters of  $N$ ,  $s(t)$ ,  $i(t)$ , and  $\lambda$  are the same as in the SI model. The SIS model additionally assumes that  $\mu$  represents the daily rates of the cured patients. In other words,  $\mu$  represents the proportion of the infected users that have been cured in the total population. The increment of the change in the number of patients can be expressed as  $N \frac{di}{dt} = \lambda Nsi - \mu Ni$ , where  $\lambda Nsi$  is the increment of the increase in the number of patients per day, and  $\mu Ni$  is the increment of the increase in the number of cured patients per day. The SIS model can be described by Equations (3) and (4). When  $\mu = 0$ , the SIS model can be simplified and expressed as the SI model.

$$\frac{di}{dt} = \lambda i(1 - i) - \mu i, \quad (3)$$

$$i(0) = i_0. \quad (4)$$

### 2.2.3. The SIR Model

The SIR model, combined with differential dynamic equations, was established by Kermack and McKendrick [23]. When an individual is cured, he/she can be an immune user. This is not taken into account by the SI or SIS model. The SIR model divides the total population  $N$  into three categories:  $S$ ,  $I$ , and  $R$  (Removed), where  $S$  and  $I$  represent the susceptible and infected users as described in the

previous two models,  $R$  represents immune users, and  $s(t) + i(t) + r(t) = 1$ . It assumes that  $s(0) = s_0$ ,  $i(0) = i_0$ ,  $r(0) = 0$ , and  $\frac{ds}{dt} + \frac{di}{dt} + \frac{dr}{dt} = 0$ . The daily increment of the increasing number of immune users is expressed by  $N \frac{dr}{dt} = \mu Ni$ . The SIR model can be described by Equations (5)–(7).

$$\frac{ds}{dt} = -\lambda si, \quad (5)$$

$$\frac{di}{dt} = \lambda i(1 - i) - \mu i, \quad (6)$$

$$\frac{dr}{dt} = \mu i. \quad (7)$$

#### 2.2.4. The SIRS Model

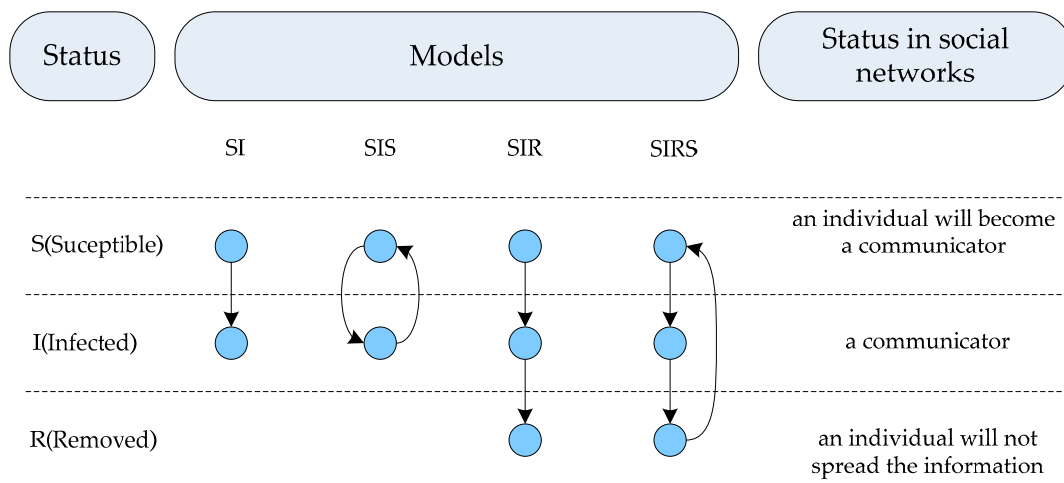
The SIRS [24] model is different to the above models. It believes that a cured user can become a susceptible user with probability  $\alpha$ . The SIRS model can be described by Equations (8)–(10).

$$\frac{ds}{dt} = -\lambda si + \alpha r, \quad (8)$$

$$\frac{di}{dt} = \lambda i(1 - i) - \mu i, \quad (9)$$

$$\frac{dr}{dt} = \mu i - \alpha r. \quad (10)$$

The comparison of basic epidemic models is shown in Figure 3, which demonstrates the diffusion process of a virus in an epidemic and also the status of the users in social networks, showing how the epidemic model can be used for information diffusion research.



**Figure 3.** Comparison of Four Basic Epidemic Models.

#### 2.2.5. Epidemic Models in Social Networks

Information diffusion is similar to the spread of an epidemic, but there are differences. Information diffusion is related to time, relationship strength, information content, social factors, network structure [25], etc. Researchers have made ongoing improvements based on classical models, developing new models such as SEIR (Suceptible Exposed Infected Removed) model [26], S-SEIR (Single layer-SEIR) [27], SCIR (Suceptible Contacted Infected Removed) model [28], irSIR (infection recovery SIR) model [29], FSIR (Fractional SIR) model [30] and ESIS (Emotional Suceptible Infected Suceptible) model [31].

Wang et al. [26] developed the SEIR model by adding Exposed (E) nodes based on the SIR model. They built a dynamical evolution equation to accurately describe the process of information propagation, analyzing the impact of user login frequency and number of friends on information diffusion. Their results demonstrated that user login frequency is directly proportional to the speed and range of information transmission. While users' behavioral characteristics were considered, different users with different behaviors were ignored; however, the behavioral characteristics of different users will have a great effect on information diffusion. When this is ignored, the propagation model will not be accurate enough.

Xu et al. [27] found that information diffusion is not only related to a user's behavior, but also to the value of the information itself. They therefore built an S-SEIR model for single layer social networks based on the SEIR model. They also proved that the transmission of information depends on users' behavior.

Xuejun [28] built an SCIR model for Micro-blogs by adding a Contacted (C) status. It assumes that all fans are assigned Contacted status when a user publishes a message. Then, the status of fans will change according to a certain probability, becoming either transmission or immune users after a while. This model can represent the regularity of online topic spreading well.

John and Joshua [29] proposed an irSIR model based on the SIR model to simulate the adoption and abandonment of user views by adding an infection recovery kinetics process. They then verify the validity of this model using Google and Facebook. Feng et al. [30] think that information diffusion is different from epidemics. The threshold for the spreading of infection is 0, whereas information diffusion will be affected by a threshold. When an individual is flooded with an excess of information, the information may not be better spread. They proposed a FSIR model to consider the effect of neighbors on an individual in the diffusion of information.

Wang et al. [31] proposed an ESIS model based on the SIS model. They considered that when information is transmitted between individuals, it also expresses a kind of emotional information. The proportion of forwarded information that has an emotional quality is used as an edge weight in this model. It proved that information diffusion is related to propagation probability and transmission intensity, and therefore its performance is better than that of the SIS model.

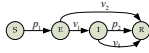
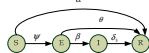
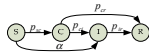
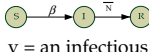
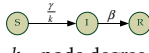
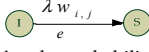
From the above analysis, we can see that the research on epidemic models is scalable research based on the basic epidemic model. The comparison of these models is shown in Table 1. All of these models can be divided into two categories, one is to simulate the information diffusion process [26–28], the other is to find those factors affecting information diffusion [26,29–31]. In the first category of model, a certain status [26–28] is always added to nodes in the model in order to describe the whole information diffusion process and predict future social network behaviors. The more statuses that are added to a node, the better the model fits the real diffusion process. However, the speed and scope of diffusion will be affected. Hence, the SEIR and S-SEIR with E node and SCIR with C node models are slower when approaching a stable state in information diffusion than is the basic SIR model.

However, actions can be taken to improve the diffusion speed, such as choosing a higher degree for the initial diffusion node. In the diffusion process, the distribution of S, E/C, I, and R are different over time. In the beginning, S will decrease rapidly to a stable number. E and C will increase rapidly, and then once they reach a highest point, they will decrease to zero in a short amount of time. The distribution of I is similar to E and C, but it changes more slowly. We can see the distribution of R as an upside down copy of the distribution of S. Understanding the distribution of these statuses is very useful in the innovation of models. In information diffusion research, the description of the diffusion process alone is not enough. Meaningful research finds those factors that can affect information diffusion: e.g., information weight value, user behavior, or emotion. In recent years, researchers have therefore considered the dynamic transmission rate [29,31–33] from one status to another with one or more factors. When we consider these factors, the new models always exhibit good performance in describing the social network. For this research, the status of nodes is based on the basic epidemic model. These factors are always involved in the recovery rate, infection rate, or a



spreader rate. It is useful for information diffusion in a social network to be based on our empirical knowledge. Most research focuses on a certain dataset that has come from a real social network. These models are therefore not adaptive. If we want to make them more adaptive, we should pay more attention to the parameters between different statuses in Table 1.

**Table 1.** Comparison of Epidemic Models in the Literature.

Scalable Model	Method	Consider the User's Different Behaviors	Expression of the Diffusion Process	Dynamic Infected Rate and Recovery Rate	Performance Metrics	Applications
SEIR [26]	add Exposed node	-		-	distribution of nodes density	detect the affect factors: login frequency and number of friends
S-SEIR [27]	information value is considered	✓	 $\delta = \text{user behavior}$	-	distribution of S, E, I, and R	simulate the diffusion process
SCIR [28]	add Contacted node	-		-	distribution of I and R	represent the regularity of online topic spreading
irSIR [29]	add Infection Recovery dynamics	-	 $v = \text{an infectious recovery rate}$	✓	degree of fitting with real data	describe OSN abandonment
FSIR [30]	consider the behavior of the neighbors	✓	 $k = \text{node degree}$	✓	degree of fitting with real data	detect the affect factors: information numbers and friends numbers
ESIS [31]	consider the information weight with emotion	-	 $\lambda = \text{the probability of I to S; } w_{i,j} = \text{the strength of edge } e \text{ from } i \text{ to } j$	✓	degree of fitting with real data	detect the affect factors: propagation probability and transmission intensity

OSN: online social network. S: susceptible. E: exposed. I: infected. R: removed.

### 2.3. Influence Models in Social Networks

Influence analysis is key in social networks [34]. Information diffusion based on influence is divided into three categories [35]: individual influence, community influence, and influence maximization. We can understand the mode of the information diffusion through influence research. The following describes the analysis and comparison of the relevant literature.

#### 2.3.1. Individual Influence

Individual influence refers to opinion leaders-related research. Opinion leaders are the nodes who can play a role as a bridge of information diffusion. They have a certain influence on other users in a social network. The influence of opinion leaders cannot be ignored in information diffusion research [36]. The research on the influence of opinion leaders includes methods based on network structure, mutual information, and user attributes [37]. It mainly uses centrality and structural holes to measure the importance of the nodes for the first method. PageRank and other algorithms are also used to rank the nodes. This method is simple, but the accuracy is not high. The second method focuses on mutual information exchange between users. Its result is more objective and accurate than the first, but it is difficult to use in large-scale data processing. The third method is based on users' behaviors, activities, or other factors. Although this method is more subjective, it is indispensable for individual influence research.

Chenxu et al. [38] proposed a method for modeling and measuring the influence of micro-blog opinion leaders based on information transmission. This method is based on network structure only. It mines the opinion leader by finding out who the tipping point node is in the information diffusion process. The process of information diffusion is described by the dynamic direct graph. It shows that information dissemination is weakly correlated with the number of opinion leaders. The initial influence of opinion leaders is positively correlated with the number of their fans. However, the influence duration is irrelevant to the number of fans. This model could be used to successfully predict popular information.

Bo et al. [39] proposed a method for finding opinion leaders based on a user's own behavior and interactions between different users. According to the competency model of management, the users in a social network can be divided into four categories, i.e., ordinary users, active users, subject opinion leaders, and network leaders. The opinion leaders can be found by using both of the dominant and implicit factors. The user's own behaviors, information analysis, and the interaction relationships between users are referred. Finally, the network opinion leaders can be obtained through three layers of screening. The influence of the information diffusion will be maximized through the mining of opinion leaders.

Jiaxin et al. [40] proposed a method for measuring social influence by predicting a user's ability to disseminate information. The influence evaluation is based on the retweet count. To get this parameter, the temporal distribution of individual retweet behaviors, the time-validity of a piece of information from a tweeter, and the preference of user retweet are analyzed. This method is based on both network structure and user's behaviors.

Xianhui et al. [41] proposed an algorithm based on Topic-Leader Rank. They combined the node weight (users' activity, relationships between two nodes, and topic), edge weight, a user interaction attribute, and a content attribute to mine the opinion leaders for a specific topic.

Ullah et al. [42] proposed an effective model for finding influential nodes, maximizing the diffusion of information, and minimizing the contagion time. This model not only considers the interactions between nodes, but also the topological structure of the network. First, it ranks the nodes based on the weights between different nodes. The influence between users is decided by the temporal interactions of the user and its neighbors. Second, the top-K influence nodes will be selected by the node neighbors, neighbors-of-neighbors, and the topological connections.

Most of the studies in individual influence research focus on mining opinion leaders. Generally speaking, if you want to find the opinion leaders you must know who the most influencing users are in a social network. The comparisons of the individual influence methods in the literature are shown in Table 2. We compare them from three aspects: network structure, user interactions, and user attributes. These are the main methods for individual influence research. There can be only one element included, e.g., tipping point node [38]. It can also include several combinations of elements, e.g., user activity and centrality [39], activity and access time distribution [40], and activity and relationships [41,42]. A user's activity is very important for individual influence research, but of course, using a combination of these three elements gives better results and produces more precise information [41]. Regardless of which method is used, they all need a quantitative criterion with which to weight the influence. The criterion can be the out-degree of a node [38], activists [39,42], centrality and intermediary [39], capability of diffusion [40], and coverage and coreratio [41]. These methods always start with a rough selection based on the network structure for opinion leaders first. Then, the interactions and user attributes are exploited to make the first result more precise. Interactions-based [41–43] individual influence research has been paid more attention in recent times.



**Table 2.** Comparison of the Individual Influence Methods.

Researcher	Network Structure	User Interactions	User Attributes		Method	Quantitative Criterion	Applications
			User behaviors	Other features			
Chenxu [38]	✓	-	-	-	social network analysis	out-degree	identify opinion leaders and prediction
Bo [39]	-	✓	✓	centrality	competency	activists, centrality and intermediary	identify opinion leaders and influence maximization
Jiaxin [40]	✓	-	✓	access time	social network analysis	capability of diffusion	influence predicting
Xianhui [41]	✓	✓	✓	topic and weight	page-rank	coverage and coreratio	mining topic opinion leader
Ullah [42]	✓	✓	✓	neighbors-of-neighbors	social network analysis	activists	identify influential nodes

### 2.3.2. Community Influence

A community is a group of people with some common properties. In social networks, individuals will form various communities on the basis of interests. A community is a subset of the network in which the users are densely connected and have similar attributes, e.g., they like to play badminton, or their research area is similar. Although the structure of social networks will change over time, communities remain relatively stable. The main challenge is how to detect those communities that have high influence within a social network, and many methods have been proposed to this end, mainly including links and attributes. Previous research has been able to detect communities by way of social links, content [44–46], node attributes [47], sentiment topics [48], and others [49].

As one example, Yang et al. [44] proposed a PCL-DC method based on a discriminative probabilistic model. This method is used to estimate communities combined with links and content. The link probability of two nodes is not only described by popularity, but also by content. It uses a two-stage Expectation-maximization (EM) algorithm to optimize community membership probabilities and content weights. In the end, each node will be distributed to a community with a maximum probability.

Zhou et al. [45] proposed an SA-Cluster-Inc method. They calculated the distance from one node to others by inserting virtual property nodes and property edges into the new attribute graph. Next, the K-means cluster algorithm is used to cluster the original nodes. According to the trend of clustering, a neighborhood random walk distance matrix will be updated in each iteration. However, only the increment matrix will be calculated instead of the full matrix calculation in the SA-Cluster model, making this model more efficient.

Ruan et al. [46] also support the viewpoint that links combine with contents, but suggest that the method is not as efficient as it could be. They proposed an efficient CODICIL method for detecting communities by way of combining links and contents. The link strength is decided by the likelihood that a relationship will remain within a community. The similarity of the content is estimated by way of the cosine similarity or Jaccard coefficient. First, it constructs the content edges. Second, the content edges and topological edges are used together to obtain the edge union. Finally, the edges that are relevant to graph node local neighborhoods will be retained using a biased edge sampling procedure, and the communities will be clustered using the Metis and Markov clustering algorithms. Yang et al. [47] share the perspective of the above researchers, but chose to give content the name “attributes”, and they proposed a Communities from Edge Structure and Node Attributes (CESNA) model.

Yang and Manandhar [48] think that links combined with content or attributes is not the ideal method for finding sentiment topic-based communities. They proposed a method that combines links, topics, and sentiment to find different communities with a different topic distribution. It also explains the structure of overlap communities. From the sentiment perspective, this method has a certain representative significance.

Peng et al. [49] think that there are 10 connections at least for each node in a community. Although the number is much less than that shown in the initial graph representation, it can represent the structure of the graph. They adopt the K-core algorithm to extract the K-core sub-graph first. Communities are detected using community detection algorithms and optimized algorithms.

Gurini et al. [50] think that the sentiment-based method is not objective. They proposed an SVO method. This method not only considers the target user's attitudes, but also the volume and objectivity of related generated contents. In this method,  $\theta = 0.8$  represents a similarity threshold. It will build an edge when the similarity value between two nodes is greater than  $\theta$ . The clustering algorithm includes two steps: the first step to find the natural partition of the network, and the second step is to find the global maximum of modularity where cliques are combined into two groups. These two steps are iterated to identify the latent sentiment communities.

Ullah et al. [51] proposed a model to detect the communities by way of trust and interest similarity. The trust between two nodes in a social network is measured by two hops between source node and target node. The trust value is bi-directional. Interest similarity is weighted by a cosine similarity function. In this model, the first set of nodes are selected as community centers according to a given initial number of communities using interest weighted by trust. Then, nodes are assigned to the communities based on a similarity threshold.

Community detection is the basis of community influence research. From the above analysis, we can see that research in community detection includes links-based, content- or attributes-based, and sentiment-based methods. However, the method based on links only is not accurate and not fit for the study of dynamic social networks. Irrelevant attributes or content will mislead the community detection process. Therefore, most researchers use a combination of these two, and the cluster-based method is predominantly used. In cluster methods, the first step is to construct the network structure based on content or attributes. Second, the initial structure will be updated by links iteratively. The main objective is to improve the accuracy of community detection and reduce time consumption. In terms of accuracy, various attributes are taken into account. From a literature analysis, we find that the performance of the model will not be improved when considering too many attributes. For example, sentiment analysis is not fit for detecting all communities or being useful for specific topics or situations. A comparison of community detection algorithms is presented in Table 3.

**Table 3.** Comparison of the Main Algorithms Aforesaid.

Model	Links	Attributes or Contents	Sentiment	Method	Quantitative Criterion
PCL-DC [44]	✓	✓	-	probability	-
SA-Cluster-Inc [45]	✓	prolific and topic	-	cluster	density and entropy function
CODICIL [46]	✓	stemmed words, title and context, tags	-	cluster	quality function
sentiment-topic based [48]	✓	user, text	✓	probability	sentiment-topic similarity
SVO [50]	✓	interests	✓	cluster	homophily
interest and trust based [51]	✓	interest, trust	-	both	quality function

### 2.3.3. Influence Maximization

Influence maximization is a recent focus of social networks research. The concept of influence maximization was first proposed by Kempe et al. [52], however its efficiency is limited. In order to

maximize influence, many researchers have done a lot of follow-up studies, e.g., the IRIE model [53] and IPA model [54]. Although the efficiency of these models is improved, they are still not accurate enough. Borgs et al. [55] proposed a method based on reverse influence sampling to improve the accuracy, but there are too many sampling times. Tang et al. [56] proposed a method that can ensure the accuracy of the model, but the time consumption is too much. According to the research up to now, the main challenge is finding the seed nodes in influence maximization [57]. These methods are based on influence probability, greedy algorithms, and heuristics algorithms. An analysis of the relevant literature is discussed below.

Lei et al. [58] believe that this influence probability is not available or is incomplete in some cases. They propose an Online Influence Maximization (OIM) model. In this model, the influence nodes are selected based on loop iteration method. The model first uses existing influence information to obtain the initial influence nodes. Second, seed nodes are selected by using an Explore–Exploit (EE) strategy based on influence competition. Third, according to the users' feedback, the influence information is updated to complete the iteration. Finally, the influence maximization nodes can be obtained through several iterations according to the users and the market budget. This model has great advantages for a choice of product sales strategy when there are several similar products made by different brands.

Lin et al. [59] propose STORM, STORM-Q, and STORM-QQ based on the Multi-Round Competitive Influence Maximization (MRCIM) method. In these models, the most influential group of nodes can be obtained by multiple iterations of several network groups.

Horel and Singer [60] consider that the research into influence nodes selection is limited in a specific range. We cannot choose any influence nodes in a whole social network, e.g., for online sales, the influence nodes can be selected from the users who have purchased the goods only. However, the users who have purchased the goods may not be the influential users based on the Heavy-tailed character in social networks. This means that to find the influence users in a certain range is unknown. To solve the problem, Horel and Singer propose an adaptive method. Users may not be influential users, but they may still know the surrounding influential users based on the Friendship Paradox.

Li et al. have focused on influence research for several years. In the WEAPON algorithm [61], interplay and individual conformity are used in influence research. They proposed an influence maximization method based on conformity awareness [62]. They have also done some research in competition influence maximization [63], and proposed a GETREAL model based on game theory. In this model, each network group is seen as a competitor that competes for influence in social networks. The selection strategy that is best for each network group is obtained by finding the Nash Equilibrium in each round game. Expected influence is viewed as the revenue in this social network game.

Morone and Makse [64] think that most researchers consider relatively large degree nodes only, but ignore weak relationship nodes. They proposed an optimal percolation model to analyze influence maximization. In this model, the function of the weak relationship nodes is emphasized and the potential individual relationships are revealed through the weak relationships.

From the above literature analysis, we can see that influence maximization research focuses on both the individual level and the community level. The common object of these two levels is to find seed nodes and maximize their influence. The comparison of influence maximization methods in the literature is shown in Table 4. Influence maximization research is always data and model driven. In a model-driven algorithm, a known influence diffusion model is given initially, then a certain heuristic algorithm can be used to choose seed nodes. It is not adaptive for some network topology. However, the analysis of the social networks is based on real social network data in data-driven models. The ultimate model will then be achieved by way of a learning process. Thus, these models are very adaptive.

The influence maximization model always includes two phases: selection phase (training phase) and action phase (competition phase). The seed nodes can be selected by single-round or multi-round methods. In the multi-round method, the historical influence will be used to update the seed nodes in the next round. The individual influence maximization is always based on a certain topic or a piece of

information. However, when it refers to competitive research, it is often based on multiple items or information. The main objective is to maximize the group's influence whether the opponent's strategy is known [59] or unknown [59,63]. If the opponent's strategy is not available, the model must be a learning-based algorithm to adapt to a variety of situations. The strategy for seeding is not the only one. There will be several strategies through multiple rounds, making it more flexible and adaptive. These basic models are often based on the Independent Cascade (IC) model, the Linear Threshold (LT) model, and the Game theory model [65].

**Table 4.** Comparison of Influence Maximization Method.

Model	Find Seeds	Techniques for Choosing Seed Nodes	Data/Model Driven	Multi-Round	Multi Innovations/Items/ Information	Application
OIM [58]	✓	explore-exploit, heuristic	model	✓	-	individual influence maximization
Adaptively Seeding [60]	✓	friendship paradox	data	-	-	
CASINO [62]	✓	conformity aware is mentioned	data	✓	-	
Optimal percolation [64]	✓	the important of weak nodes	data	-	-	
STORM [59]	✓	maximization the total gain	data	✓	✓	competitive influence maximization
GETREAL [63]	✓	game theory	model	-	✓	

### 3. Predictive Models

#### 3.1. Aims of Predictive Models

A book dealer will want to know which book will be most popular in next quarter: being able to accurately predict that information will be useful. In a social network, when a piece of important information is published by an individual, the information will be spread quickly throughout the social network. Especially in the case of “bad” information, a government will want to know how a situation will develop: being able to predict how information will spread throughout the network in the future will be useful. Predictive models are used to predict the future information diffusion process in social networks based on certain factors. These models are also often used for influence maximization. They are the IC model, the LT model, and the Game Theory model.

#### 3.2. Independent Cascade Model (ICM)

In the basic IC model, the inactive node  $v$  can be activated by the active node  $u$  independently with a probability of  $P_{u,v}$  at time  $t$ . If node  $v$  is activated, then it will be an active node at time  $t$ . Regardless of whether the  $u$  activates  $v$  at time  $t$ ,  $v$  will not be activated by  $u$  at the following time. The IC model is mainly used for prediction and influence research. Saito et al. [66] adopted an EM algorithm to predict propagation probability based on this model. It is not suitable for the application of a large amount of data in social networks due to its time consumption. Wang et al. [67] and Jung et al. [53] focus on influence maximization based on the IC model. They believe that the scalability of the algorithm is key for influence maximization research, in order to fit large-scale social networks. Arora et al. [68] proposed an ASIM algorithm to combine running time with memory consumption for influence maximization research, making it fit for the study of real social networks. The IC model is being applied more and more in the field of influence maximization. The research includes both direct applications of the IC model and scalable applications.

In scalable application research, Barbieri et al. [69] propose a topic-aware model, i.e., the TIC and TLT model, which realizes that the propagation of an item in a social network is related to its topic. The model is used to obtain a topic distribution.

Kim et al. [70] do not agree with the view that impact is a one-time thing in the basic IC model, as it is inconsistent with the actual data. They believe that activation is time limited and proposed a CT-IC model. In this model, each active node will activate its inactive neighborhoods repeatedly within a limited timeframe. Of course, this model is better than the basic IC model.

Zhu et al. [71] propose a Continuous Time Markov Chain model (CTMC-ICM)—based on the IC model—to find a small subset of nodes in a social network that could maximize the spread of influence.

Research on the IC model focuses mainly on scalability, e.g., ASIM, TIC, CT-IC, and CTMC-IC, and application research is focused on influence maximization.

### 3.3. Linear Threshold Model (LTM)

In the LT model, each active node  $v$  has an activation threshold at time  $t$ . All the neighborhoods of  $v$  try to activate  $v$ . When the influence degree of all active nodes exceeds the activation threshold of  $v$ , inactive node  $v$  will become an active node at time  $t + 1$ . All the active neighbors can activate  $v$  many times. As with the IC model, the LT model is used to study influence in social networks, with a focus on threshold behavior during the influence spreading process. In other words, it focuses on the cumulative effect of influence spread in the process. There has been work done on how to maximize influence based on the LT model [10,23,72].

Lagnier et al. [73] proposed a Decaying Reinforced User-centric (DRUC) model—based on the LT model—which combines information content with a user profile. At time  $t$ , the probability of information diffusion is decided by three factors: the user's interest in the content of the information; the user's intent with regard to spreading the information; and the influence of neighbor nodes that have "infected". However, this model is not related to the time delay of information propagation. The propagation process of the model is discrete on the time axis.

Chen and Yitong [74] proposed a heuristic algorithm based on an activation threshold using the LT model. It considers the influence of nodes and the node's activation threshold comprehensively. According to the dynamic threshold of each node in the activation process, the Potential Influence Nodes (PIN) will be calculated. In the heuristic stage, the nodes with the largest PIN value will be selected as seed nodes. In the greedy stage, the nodes that have the greatest influence on the increment of the maximum range are selected to expand the influence of the social network's information diffusion.

From the literature analysis, we know that the LT model tends to be used for influence spread and influence maximization research. It always adopts either the heuristic algorithm or greedy algorithm to find the seed nodes with the greatest influence.

### 3.4. Game Theory Model (GTM)

Game theory is a strategy that maximizes profit. The study of game theory is limited to multiple individuals or groups with specific restrictions. It always utilizes an opponent's strategy to maximize profit. A piece of information is either spread or not due to the effect of costs, benefits, and strategic choice. Game theory has been used in social networks research for several years. Camerer [75] utilized game theory to model and analyze the interactions between cascading behavior and individual and group effect in a network at Stanford University in the United States.

Qifa et al. [76] focused on the microscopic view. Using game theory, the relationship as an important variable is added to the model. The cost, benefit, and strategy choice are analyzed when the users choose to spread the information. The results show that as long as the profit from information diffusion is more than the cost of information, users will choose to spread the information. The closer is the relationship that exists between users, the more easily information is spread.

Yuanzhuo et al. [77] proposed an evolutionary game model of network group behaviors based on a game model. They believe that the features of individual information behavior on the micro level are much more complex than on the macro level. Because of the sociality and randomness,

group behaviors often show great uncertainty in a network. The evolution game model is suitable for solving the dynamic problem of information propagation in social networks.

Liu et al. [78] proposed a joint game model (Game Coalition) to predict future relationships between users when they consider the structure of a social network (e.g., fans, neighbors) and interactive features (e.g., @ interest and topic) comprehensively. They validated the performance of their model using Twitter and Sina Micro-blog data.

In prediction research, network structures and user behaviors are either involved in a model independently or they are combined. A social network is a dynamic network, therefore the prediction models for a social network must be robust. Through an analysis of the literature and the comparison of these three models, we know that IC models are sender-centered, considering only the senders of the information, whereas LT models are receiver-centered. Game Theory (GT) models are more neutral and consider the profit of the whole network, making them suitable for the study of dynamic networks. A comparison of these three models is presented in Table 5.

**Table 5.** Comparison of Independent Cascade Model (ICM), Linear Threshold Model (LTM) and Game Theory Model (GTM).

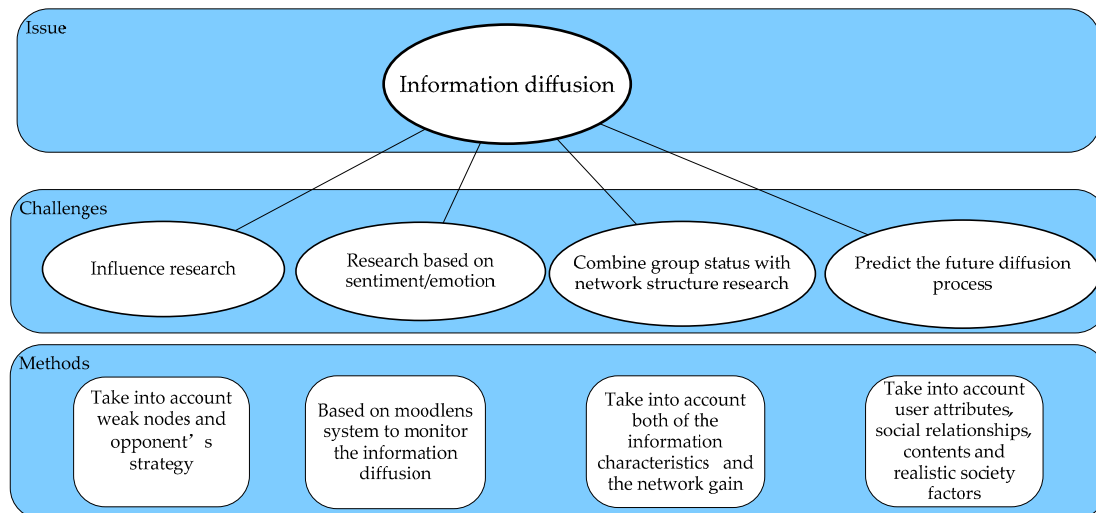
Model	Basic Model			Research Views	Application
	IC	LT	GT		
EM [66]	✓	-	-	the likelihood for information diffusion episodes	prediction of propagation probability
ASIM [68]	✓	-	-	combine running-time with memory-consumption	influence maximization
TIC, TLT [69]	✓	✓	-	Topic-aware	prediction of topic distribution
DRUC [73]	-	✓	-	information content and user profile	find affect factors
Heuristic and Greedy [74]	-	✓	-	influence of nodes and the node's activation threshold	select the greatest influence nodes
Microscopic [76]	-	-	✓	relationship and cost	prediction of the information spread
Evolutionary game [77]	-	-	✓	individual information behavior in micro level	prediction of information diffusion in dynamic network
Game Coalitional [78]	-	-	✓	structure of social network and interactive features	relationships prediction

From the investigation of the IC, LT, and GT models above, we can see that these three models are used not only for prediction, but also for influence spread or maximization. That is, the predictive models are not independent and the explanatory models are the basis for prediction research. The prediction models are the tools or methods, and these two models always need each other. In the case of prediction, the researcher's views can be classified as either macro and micro perspectives. The macro view analyzes future diffusion on the basis of a network structure, and the micro view analyzes future diffusion using user behavior. In prediction research, network structures and user behaviors are either involved in a model independently or they are combined. A social network is a dynamic network, therefore the prediction models for a social network must be robust. Through an analysis of the literature and a comparison of these three models, we know that IC models are sender-centered, considering only the senders of the information, whereas LT models are receiver-centered. GT models are more neutral and consider the profit of the whole network, making them suitable for the study of dynamic networks. The comparison of these three models in social networks is presented in Table 5.



#### 4. Future Challenges

Information diffusion has been a hot topic in social networks research in recent years. Although there have been many innovative studies in this field, there are still some issues that need to be resolved. The challenges and methods for overcoming them are presented in Figure 4 below.



**Figure 4.** The Future Challenges and Methods for Information Diffusion Research.

##### 4.1. Influence Analysis

- The importance of the weak nodes

The recent research on influence focuses on individuals, communities, and networks. The purpose of this research is to find seed nodes. Most researchers will consider degree-centrality nodes, closeness-centrality nodes, and betweenness-centrality nodes. These factors are explicit. While the implicit weak-nodes are not degree-centrality nodes, they know which nodes are either influence nodes or play an important role in information diffusion across the whole social network. In the future, it will be necessary to use machine learning to obtain parameters for the model from a realistic dataset.

- Competitive influence maximization

The objective of recent research into influence maximization often focuses on single pieces of information; however, there are various kinds of information, such as advertising information, that are of interest to stakeholders. How to make information contained in certain advertisements popular is referred to as competitive influence maximization. Research into maximizing competitive influence can also be applied to the community competition between positive and negative statements in a social network. In other words, one community's strategy will impact another community's strategic choice [79]. Research has been carried out in this area; however, an opponent's strategy is not often available, therefore the competitive influence model must be a learning-based algorithm that adapts to a variety of situations. That is to say, there is not only one strategy for seeding. There must be several strategies through multiple rounds. In our opinion, competitive influence research will be very useful for the marketing industry.

##### 4.2. Information Diffusion Based on Sentiment/Emotion

Opinions expressed within social networks are often related to an individual's current situation, and online social networks provide a way for users to express their emotions. The dissemination of information that has emotive content can have a huge impact on real society. Most researchers focus on

emotion/sentiment in social networks. Zhao et al. [80] think that traditional emotional analysis based on keywords is not suitable for a short Micro-blog, and that a method based on expression symbols is better than one based on keywords. They divided the traditional positive and negative emotions into four categories: anger, disgust, joy, and sadness. Approximately 95% of emotions can be classified into these four general categories. They built a Moodlens system to monitor the sentiment of users in social networks. Fan et al. [81] undertook research based on the work done by Zhao. They discussed the emotional relationship between interactive users in a Micro-blog. Their research shows that the “angry” relationship is stronger than the “joyful” one. This “angry” emotional relationship is related to real social life, e.g., food safety, corruption, and bribery. This research is a basis for emotional impact and emotion diffusion in social networks. Kramer et al. [82] have also certified information diffusion with emotion in their research. Information diffusion with sentiment analysis in social networks is rare, but useful, especially for building a good and adaptive moodlens system to monitor information diffusion. We believe that sentiment analysis is an important factor that should not be ignored in information diffusion.

#### *4.3. Combine Group Status with Network Structure Research*

From the review of the literature, we can see that most of the research refers to group status or network structure analysis. All the users are divided into several groups in group status-based models. A node with a particular status will be affected by other nodes that have different statuses around it, so that its status will be changed from one group to another. This model only pays attention to the group status impact of the surrounding nodes. However, information diffusion will be greatly affected by characteristics of the group and society. Information diffusion in this model is more objective. There are more factors that must be considered. In the network structure-based model, information diffusion mainly depends upon whether there will be a gain greater than a certain threshold in the whole social network. Although this model is subjective, it lacks the characteristics of information. In our opinion, the best way to solve the problem is combine these two models.

#### *4.4. Prediction of Information Diffusion*

Researchers forecast the trends of information diffusion based on the characteristics of information diffusion analysis [83,84]. Jiuxin et al. [85] propose a method that combines user attributes with social relationships and micro-blog contents to predict information diffusion. Zhao et al. [86] propose a Self-Exciting Model to predict the number of retweets a post will receive on Twitter. Generally speaking, although there has been some work, this research is still infrequent and most of it is theoretical in nature; the field of research and its application is still very limited. At the same time, the predictions refer not only to online social networks, but also to aspects of real society, e.g., the impacts of television, newspapers, other traditional media, and the reality of social activities. To move prediction research from theory to practice requires a long process, and much future work is needed in this direction.

### **5. Discussion and Conclusions**

Over recent years, a variety of social networks have emerged for individuals to keep in contact with others conveniently. Large amounts of information can therefore be produced within these social networks that can tell us who the most important person is, who a topic leader will be, why an event will unfold in a certain way, and other important things. To solve these issues requires multidisciplinary research that draws on the fields of not only computer science, but also sociology, psychology, economics, and others. Researchers have built certain models to explain the diffusion phenomenon, and others to predict future diffusion, all of which are based on machine learning. In this paper, the explanatory and predictive models for information diffusion analysis were investigated.

The primary goal of information diffusion analysis is to illustrate the diffusion process. The epidemic model is the first choice for this research (e.g., SI, SIS, SIR). It utilizes a model to simulate information diffusion but is not very accurate. Therefore, more statuses are added to the basic epidemic

model, and although the resulting models are more accurate, they are still not fit for real social networks when considering users' different behaviors, as an individual's behavior can affect information diffusion. Such factors are important when building a diffusion model. Many additional factors could be involved in information diffusion models, such as network structure, mutual information, and user behaviors. In influence research especially, more than one factor is always combined in order to maximize the social network's influence. In individual influence research, the focus is always on one type of information only. There are multiple types of information to be taken into account for research into competitive influence. Generally speaking, competitive research is often used in community influence.

With the development of information diffusion research, the future diffusion process can be predicted, and the IC, LT, and GT models are used for this purpose. From the perspective of research applications, these three models can be used not only for prediction, but also for influence research.

The predictive models are always based on the former basic explanatory models. This view is also supported by the literature in this paper. In the literature [29], the dynamical directed graph is employed to first model the propagation process, which reveals the influence of opinion leaders. The influence model is then used to predict how and where the information flows. By way of an effect factors analysis, a prediction model can be built to predict the influence of a node on a given social network [31]. Prediction models are also used as tools for explanatory research, especially in influence research. In references [47,51], we can see the use of the IC and LT models in influence maximization research. Hence, these two models are not independent. Of these three models, the GT model is the more adaptive and robust, and is therefore widely used in social networks research.

In this paper, a systematic analysis of the research into information diffusion in social networks is stated. It divides the information diffusion models into two categories: explanatory models and predictive models. The most widely used models are analyzed in detail. From the literature analysis arises three points. First, most of the studies on information diffusion models aim to explain the information diffusion process, to find factors, or to predict the future outcomes and direction of the process. Second, these two models are not independent; they always need each other. Third, we explore the current issues in these studies and the future research directions. From the literature, we can conclude that this type of research is very meaningful and its future applications may be able to provide decision support for public opinion monitoring, marketing, etc.

**Acknowledgments:** This paper is supported by Natural Science Foundation of Hebei Education Department (No. QN2015207), National Science Foundation of China (No. 61272362), the Key Research Project for University of Hebei Province (No. ZD2014029), the Basic Research Project of Hebei Province (No. F2017208012), and the Research Project for College of Information Science and Engineering, Hebei University of Science and Technology.

**Author Contributions:** Mei Li and Xiang Wang gathered all of the literature; Mei Li, Xiang Wang, Kai Gao and Shanshan Zhang read and analyzed the literature; Mei Li, Xiang Wang, Kai Gao and Shanshan Zhang wrote the paper; all of the authors have read and checked the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Barnes, J.A. Class and committees in a Norwegian island parish, human relations. *Hum. Relat.* **1954**, *7*, 39–58. [[CrossRef](#)]
2. Christakis, N.A.; Fowler, J.H. The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **2007**, *357*, 370–379. [[CrossRef](#)] [[PubMed](#)]
3. Zhang, Y.; Wu, Y. How behaviors spread in dynamic social networks. *Comput. Math. Organ. Theory* **2012**, *18*, 419–444. [[CrossRef](#)]
4. Fallahpour, R.; Chakouvari, S.; Askari, H. Analytical solutions for rumor spreading dynamical model in a social network. *Nonlinear Eng.* **2015**, *4*, 23–29. [[CrossRef](#)]
5. Zhao, Z.; Liu, Y.; Wang, K. An analysis of rumor propagation based on propagation force. *Physica A* **2015**, *443*, 469–474. [[CrossRef](#)]

6. Li, D.; Ma, J.; Tian, Z.; Zhu, H. An evolutionary game for the diffusion of rumor in complex networks. *Physica A* **2015**, *433*, 51–58. [[CrossRef](#)]
7. Lönnqvist, J.E.; Deters, F.G.E. Facebook friends, subjective well-being, social support, and personality. *Comput. Hum. Behav.* **2016**, *55*, 113–120. [[CrossRef](#)]
8. Dong, Y.H.; Chen, H.; Qian, W.N.; Zhou, A.Y. Micro-blog social moods and Chinese stock market: The influence of emotional valence and arousal on Shanghai composite index volume. *Int. J. Embed. Syst.* **2015**, *7*, 148–155. [[CrossRef](#)]
9. Cui, Q.; Qiu, Z.; Liu, W.; Hu, Z. Complex dynamics of an sir epidemic model with nonlinear saturate incidence and recovery rate. *Entropy* **2017**, *19*, 305. [[CrossRef](#)]
10. Guo, J.; Zhang, P.; Fang, B.X.; Zhou, C.; Cao, Y.; Guo, L. Personalized key propogating users mining based on LT model. *Chin. J. Comput.* **2014**, *37*, 809–818.
11. Binxing, F. *Online Social Network Analysis*; Publishing House of Electronics Industry: Beijing, China, 2014.
12. Guille, A.; Hacid, H.; Favre, C.; Zighed, D. Information diffusion in online social networks: A survey. In Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, New York, NY, USA, 22–27 June 2013; pp. 31–36.
13. Li, D.; Xu, Z.M.; Li, S.; Liu, T.; Wang, X.W. A survey on information diffusion in online social networks. *Chin. J. Comput.* **2014**, *254*, 31–36.
14. Wani, M.; Ahmad, M. Information diffusion modelling and social network parameters (A survey). In Proceedings of the International Conference on Advances in Computers, Communication and Electronic Engineering, Kashmir, India, 16–18 March 2015; pp. 245–249.
15. Kumaran, P.; Chitrakala, S. A survey on influence spreader identification in online social network. In Proceedings of the International Conference on Information Communication and Embedded Systems, Chennai, India, 25–26 February 2016; pp. 1–7.
16. Dey, K.; Kaushik, S.; Subramaniam, L.V. Literature survey on interplay of topics, information diffusion and connections on social networks. *arXiv* **2017**, arXiv:1706.00921.
17. Han, X.; Niu, L. On charactering of information propagation in online social networks. *J. Netw.* **2013**, *8*, 124–132.
18. Ou, C.; Jin, X.; Wang, Y.; Cheng, X. Modelling heterogeneous information spreading abilities of social network ties. *Simul. Model. Pract. Theory* **2017**, *75*, 67–76. [[CrossRef](#)]
19. Pastorsatorras, R. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **2001**, *86*, 3200–3203. [[CrossRef](#)] [[PubMed](#)]
20. Newman, M.E.J. The structure and function of complex networks. *Soc. Ind. Appl. Math.* **2003**, *45*, 167–256. [[CrossRef](#)]
21. Newman, M.E. Threshold effects for two pathogens spreading on a network. *Phys. Rev. Lett.* **2005**, *95*, 108701. [[CrossRef](#)] [[PubMed](#)]
22. Gross, T.; D’Lima, C.J.; Blasius, B. Epidemic dynamics on an adaptive network. *Phys. Rev. Lett.* **2006**, *96*, 208701. [[CrossRef](#)] [[PubMed](#)]
23. Liu, D.; Yan, E.W.; Song, M. Microblog information diffusion: Simulation based on sir model. *J. Beijing Univ. Posts Telecommun.* **2014**, *16*, 28–33.
24. Jin, Y.; Wang, W.; Xiao, S. An sirs model with a nonlinear incidence rate. *Chaos Solitons Fractals* **2007**, *34*, 1482–1497. [[CrossRef](#)]
25. Liu, C.; Zhang, Z.K. Information spreading on dynamic social networks. *Commun. Nonlinear Sci. Numer. Simul.* **2012**, *19*, 896–904. [[CrossRef](#)]
26. Wang, C.; Yang, X.Y.; Xu, K.; Ma, J.F. Seir-based model for the information spreading over SNS. *Tien Tzu Hsueh Pao/Acta Electron. Sin.* **2014**, *42*, 2325–2330.
27. Xu, R.; Li, H.; Xing, C. Research on information dissemination model for social networking services. *Int. J. Comput. Sci. Appl.* **2013**, *2*, 1–6.
28. Ding, X.J. Research on propagation model of public opinion topics based on SCIR in microblogging. *Comput. Eng. Appl.* **2015**, *51*, 20–26.
29. John, C.; Joshua, S.A. Epidemiological modeling of online social network dynamics. *arXiv* **2014**, arXiv:1401.4208.
30. Feng, L.; Hu, Y.; Li, B.; Stanley, H.E.; Havlin, S.; Braunstein, L.A. Competing for attention in social media under information overload conditions. *PLoS ONE* **2015**, *10*, e0126090. [[CrossRef](#)] [[PubMed](#)]

31. Wang, Q.; Lin, Z.; Jin, Y.; Cheng, S.; Yang, T. Esis: Emotion-based spreader-ignorant-stifler model for information diffusion. *Knowl.-Based Syst.* **2015**, *81*, 46–55. [[CrossRef](#)]
32. Qu, B.; Hanjalic, A.; Wang, H. Heterogeneous recovery rates against sis epidemics in directed networks. In Proceedings of the International Conference on Network Games, Control and Optimization, Trento, Italia, 29–31 October 2014.
33. Lu, D.; Yang, S.; Zhang, J.; Wang, H.; Li, D. Resilience of epidemics for sis model on networks. *Chaos* **2017**, *27*, 083105. [[CrossRef](#)] [[PubMed](#)]
34. Wasserman, S.; Faust, K. Social network analysis methods and applications. *Struct. Anal. Soc. Sci.* **1994**, *91*, 219–220.
35. Li, H.; Cui, J.; Ma, J. Social influence study in online networks: A three-level review. *J. Comput. Sci. Technol.* **2015**, *30*, 184–199. [[CrossRef](#)]
36. Fan, X.H.; Zhao, J.; Fang, B.X.; Li, Y.X. Influence diffusion probability model and utilizing it to identify network opinion leader. *Chin. J. Comput.* **2013**, *36*, 360–367. [[CrossRef](#)]
37. Wu, X.D.; Li, Y.; Li, L. Influence analysis of online social networks. *Chin. J. Comput.* **2014**, *37*, 735–752.
38. Wang, C.X.; Guan, X.H.; Qin, T.; Zhou, Y.D. Modelling on opinion leader's influence in microblog message propagation and its application. *J. Softw.* **2015**, *26*, 1473–1485.
39. Chen, B.; Tang, X.; Yu, L.; Liu, Y. Identifying method for opinion leaders in social network based on competency model. *J. Commun.* **2014**, *35*, 12–22.
40. Mao, J.X.; Liu, Y.Q.; Zhang, M.; Ma, S.P. Social influence analysis for micro-blog user based on user behavior. *Chin. J. Comput.* **2014**, *37*, 791–800.
41. Wu, X.; Zhang, H.; Zhao, X.; Li, B.; Yang, C. Mining algorithm of microblogging opinion leaders based on user-behavior network. *Appl. Res. Comput.* **2015**, *32*, 2678–2683.
42. Ullah, F.; Lee, S. Identification of influential nodes based on temporal-aware modeling of multi-hop neighbor interactions for influence spread maximization. *Physica A* **2017**, *486*, 968–985. [[CrossRef](#)]
43. Sheikahmadi, A.; Nematbakhsh, M.A.; Zareie, A. Identification of influential users by neighbors in online social networks. *Physica A* **2017**, *486*, 517–534. [[CrossRef](#)]
44. Yang, T.; Jin, R.; Chi, Y.; Zhu, S. *Combining Link and Content for Community Detection*; Springer: New York, NY, USA, 2014; pp. 190–201.
45. Zhou, Y.; Cheng, H.; Yu, J.X. Clustering large attributed graphs: An efficient incremental approach. In Proceedings of the 2010 IEEE International Conference on Data Mining, Sydney, Australia, 13–17 December 2010; pp. 689–698.
46. Ruan, Y.; Fuhry, D.; Parthasarathy, S. Efficient community detection in large networks using content and links. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2012; pp. 1089–1098.
47. Yang, J.; McAuley, J.; Leskovec, J. Community detection in networks with node attributes. In Proceedings of the 2013 IEEE 13th International Conference on Data Mining Workshops, Dallas, TX, USA, 7–10 December 2013; pp. 1151–1156.
48. Yang, B.; Manandhar, S. Community discovery using social links and author-based sentiment topics. In Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Beijing, China, 17–20 August 2014; pp. 580–587.
49. Peng, C.; Kolda, T.G.; Pinar, A. Accelerating community detection by using K-core subgraphs. *arXiv* **2014**, arXiv:11403.2226.
50. Gurini, D.F.; Gasparetti, F.; Micarelli, A.; Sansonetti, G. Analysis of sentiment communities in online networks. In Proceedings of the International Workshop on Social Personalisation & Search Co-Located with the ACM SIGIR Conference, Santiago, Chile, 9–13 August 2015; pp. 1–3.
51. Ullah, F.; Lee, S. Community clustering based on trust modeling weighted by user interests in online social networks. *Chaos Solitons Fractals* **2017**, *103*, 194–204. [[CrossRef](#)]
52. Kempe, D.; Kleinberg, J.; Tardos, É. Maximizing the spread of influence through a social network. In Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Washington, DC, USA, 24–27 August 2003; pp. 137–146.
53. Jung, K.; Heo, W.; Chen, W. Irie: Scalable and robust influence maximization in social networks. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining, Brussels, Belgium, 10–13 December 2012; pp. 918–923.



54. Kim, J.; Kim, S.K.; Yu, H. Scalable and parallelizable processing of influence maximization for large-scale social networks? In Proceedings of the 2013 IEEE 29th International Conference on Data Engineering, Brisbane, Australia, 8–12 April 2013; pp. 266–277.
55. Borgs, C.; Brautbar, M.; Chayes, J.; Lucier, B. Maximizing social influence in nearly optimal time. *arXiv* **2012**, arXiv:1212.0884.
56. Tang, Y.; Xiao, X.; Shi, Y. Influence maximization: Near-optimal time complexity meets practical efficiency. In Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, Snowbird, UT, USA, 22–27 June 2014; pp. 75–86.
57. Wang, Y.; Huang, W.J.; Zong, L.; Wang, T.J.; Yang, D.Q. Influence maximization with limit cost in social network. *Sci. China Inf. Sci.* **2013**, *56*, 1–14. [[CrossRef](#)]
58. Lei, S.; Maniu, S.; Mo, L.; Cheng, R.; Senellart, P. Online influence maximization. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, 10–13 August 2015; pp. 645–654.
59. Lin, S.C.; Lin, S.D.; Chen, M.S. A learning-based framework to handle multi-round multi-party influence maximization on social networks. In Proceedings of the 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, 10–13 August 2015; pp. 695–704.
60. Horel, T.; Singer, Y. Scalable methods for adaptively seeding a social network. In Proceedings of the 24th International World Wide Web Conference (WWW2015), Florence, Italy, 18–22 May 2015; pp. 1–14.
61. Hui, L.I.; Shen, B.; Cui, J.; Ma, J. Ugc-driven social influence study in online micro-blogging sites. *China Commun.* **2014**, *11*, 141–151.
62. Li, H.; Bhowmick, S.S.; Sun, A.; Cui, J. Conformity-aware influence maximization in online social networks. *VLDB J.* **2014**, *24*, 117–141. [[CrossRef](#)]
63. Li, H.; Bhowmick, S.S.; Cui, J.; Gao, Y.; Ma, J. Getreal: Towards realistic selection of influence maximization strategies in competitive networks. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Australia, 31 May–4 June 2015; pp. 1525–1537.
64. Morone, F.; Makse, H.A. Influence maximization in complex networks through optimal percolation. *Nature* **2015**, *524*, 65–68. [[CrossRef](#)] [[PubMed](#)]
65. Chen, W.; Wang, Y.; Yang, S. Efficient influence maximization in social networks. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Paris, France, 28 June–1 July 2009; pp. 199–208.
66. Saito, K.; Nakano, R.; Kimura, M. Prediction of information diffusion probabilities for independent cascade model. In Proceedings of the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Zagreb, Croatia, 3–5 September 2008; pp. 67–75.
67. Wang, C.; Chen, W.; Wang, Y. Scalable influence maximization for independent cascade model in large-scale social networks. *Data Min. Knowl. Discov.* **2012**, *25*, 545–576. [[CrossRef](#)]
68. Arora, A.; Galhotra, S.; Virinchi, S.; Roy, S. Asim: A scalable algorithm for influence maximization under the independent cascade model. In Proceedings of the 24th ACM International Conference on World Wide Web Companion, Florence, Italy, 18–22 May 2015; pp. 35–36.
69. Barbieri, N.; Bonchi, F.; Manco, G. Topic-aware social influence propagation models. *Knowl. Inf. Syst.* **2012**, *37*, 555–584. [[CrossRef](#)]
70. Kim, J.; Lee, W.; Yu, H. Ct-ic: Continuously activated and time-restricted independent cascade model for viral marketing. *Knowl.-Based Syst.* **2012**, *62*, 960–965. [[CrossRef](#)]
71. Zhu, T.; Wang, B.; Wu, B.; Zhu, C. Maximizing the spread of influence ranking in social networks. *Inf. Sci.* **2014**, *278*, 535–544. [[CrossRef](#)]
72. Chen, W.; Yuan, Y.; Zhang, L. Scalable influence maximization in social networks under the linear threshold model. In Proceedings of the 2010 IEEE International Conference on Data Mining, Sydney, Australia, 13–17 December 2010; pp. 88–97.
73. Lagnier, C.; Denoyer, L.; Gaussier, E.; Gallinari, P. *Predicting Information Diffusion in Social Networks Using Content and User's Profiles*; Springer: Berlin, Germany, 2013; pp. 74–85.
74. Chen, H.; Wang, Y.T. Threshold-based heuristic algorithm for influence maximization. *J. Comput. Res. Dev.* **2012**, *49*, 2181–2188.
75. Camerer, C.F. Behavioral game theory experiment in strategic interaction. *J. Socio-Econom.* **2003**, *32*, 135–146.



76. Hang, Q.F.; Zhu, J.M.; Song, B.; Zhang, N. Game model of information transmission in social networks. *J. Chin. Comput. Syst.* **2014**, *35*, 473–477.
77. Wang, Y.; Yu, J.; Qu, W.; Shen, H.; Cheng, X.; Lin, C. Everlutionary game model and analysis methods network group behavior. *Chin. J. Comput.* **2015**, *38*, 282–300.
78. Liu, D.; Wang, Y.; Jia, Y.; Li, J.; Yu, Z. From strangers to neighbors: Link prediction in microblogs using social distance game. In Proceedings of the 7th ACM International Conference on Web Search and Data Mining, New York, NY, USA, 24–28 February 2014.
79. Hu, Z.; Yao, J.; Cui, B.; Xing, E. Community level diffusion extraction. In Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD'15), Melbourne, Australia, 31 May–4 June 2015; pp. 1555–1569.
80. Zhao, J.; Dong, L.; Wu, J.; Xu, K. Moodlens: An emoticon-based sentiment analysis system for Chinese tweets. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; pp. 1528–1531.
81. Fan, R.; Zhao, J.; Chen, Y.; Xu, K. Anger is more influential than joy: Sentiment correlation in Weibo. *PLoS ONE* **2014**, *9*, e110184. [[CrossRef](#)] [[PubMed](#)]
82. Kramer, A.D.; Guillory, J.E.; Hancock, J.T. Experimental evidence of massive-scale emotional contagion through social networks. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 8788–8790. [[CrossRef](#)] [[PubMed](#)]
83. Chua, F.C.T.; Lauw, H.W.; Lim, E.P. Generative models for item adoptions using social correlation. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 2036–2048. [[CrossRef](#)]
84. Lee, J.R.; Chung, C.W. A new correlation-based information diffusion prediction. In Proceedings of the 23rd International Conference on World Wide Web Companion, Seoul, Korea, 7–11 April 2014; pp. 346–351.
85. Cao, J.X.; Wu, J.L.; Shi, W.; Liu, B.; Zheng, X.; Luo, J.Z. Sina microblog information diffusion analysis and prediction. *Chin. J. Comput.* **2014**, *37*, 779–790.
86. Zhao, Q.; Erdogdu, M.A.; He, H.Y.; Rajaraman, A.; Leskovec, J. Seismic: A self-exciting point process model for predicting tweet popularity. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, 10–13 August 2015; pp. 1513–1522.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).