

Diffusion Models for Social Analysis, Influence and Learning

Naglaa M.Badr
Information System Departement
Computers and Information Faculty
Menofia University, Egypt
naglaa.badr@ci.menofia.edu.eg

Assma H.elSayed
Information System Departement
Computers and Information Faculty
Menofia University, Egypt
asmaa.elsayed@ci.menofia.edu.eg

Hatem M.abdelkader
Information System Departement
Computers and Information Faculty
Menofia University, Egypt
hatem.abdelkader@ci.menofia.edu.eg

Abstract— Social networks are complicated by millions of users interacting and creating massive amounts of content. The problem is that any unobservable changes in network structure can result in dramatic swings in the spread of new ideas and behaviors between users. This diffusion process leads to numerous latent information that can be extracted, analyzed, and used in different applications, including market forecasting, rumor control, disease modeling, and opinion monitoring. Furthermore, mining social media big data helps to ease tracking propagated data and trends across the world. In this article, we address the study of diffusion models in social networks. We discuss three significant categories of diffusion models: contagion, social influence, and social learning models with different enhancements applied to improve performance. The aim is to study diffusion models in social networks to effectively understand how innovation and information spread over individuals and predict future trends. Also, identifying the most influential users in social networks is addressed to help spread knowledge faster and prevent harmful content like viruses or bad online behavior from spreading.

Keywords—Social Network, Information Diffusion, social influence, Predictive Models, Contusion.

I. INTRODUCTION

It's difficult to understand how information spreads over a network[1]. This information may be computer viruses, products, habits, or rumors and ideas propagated by social media. Information diffusion models are often used to improve viral marketing, social advertising, and disease modeling[2]. Diffusion is the process through which information is propagated between different locations via interaction. The diffusion process comprises three primary components: the sender, who initiates the process, the receiver, who receives diffusion information from the sender, and the medium. Typically, the number of recipients outnumbers the senders and mediums. This is the path that diffusion data travels from sender to recipient[3]. This can include social media (e.g., a tweet on Twitter).

The diffusion process differs as some users are more ready to adopt the innovation than others regarding their structural positions, personal characteristics, behaviors, and other parameters. Moreover, networks with varying connectivity patterns have unique propagation features that affect the diffusion of information, such as rumors[4].

Therefore, diffusion models can illustrate the spread process, show possible paths, predict a trend, and accelerate the spread process when needed.

To exploit the benefit of the enormous data diffused through social networks, we need first to understand three important aspects. The first is to extract the latent information found in social networks[5]. This information can be a user's habit, behaviors, emotions, relationships, or number of friends. Second, we need to know the reason behind information propagation in specific, which factors like interactions have affected data analysis results, and who initiates the adoption process of information and has a massive influence on other users causing information diffusion[6]. The final aspect is to investigate if there is a possibility that the information will be diffused in the future and use the user characteristic, extracted factors, and influence analysis to predict the future node destination. This means the ability to increase influence and predict trends like in viral marketing[1]. Fig. 1 explains these aspects of information diffusion, including a social network of users like Facebook or Twitter. After using social data analytics to store and extract data, The main task is to comprehend the information dispersion process and its influences. Then, use the analytical results to forecast future diffusion processes. Social network diffusion models can be classified into three models: contagion, social influence, and social learning[7]. The contagion diffusion model seems like an epidemic model that users adopt from others who already have it.

Meanwhile, In social influence, information spread depends on influential people as people adopt ideas or innovation when enough other people in the group have adopted. Finally, social learning means people adopt when they are convinced by sufficient empirical evidence that the information is worthwhile, where the outcomes of prior adopters provide the proof[4], [8]. Thus, Individuals may adopt at different times depending on their prior beliefs, knowledge, and costs.

This paper discusses basic diffusion models in social networks and their methods in these three classes. Furthermore, as these three models are not separated from one another, we show that these models may complete each other. Finally, we indicated no clear separation between different models and how they can be integrated to serve the application purpose.

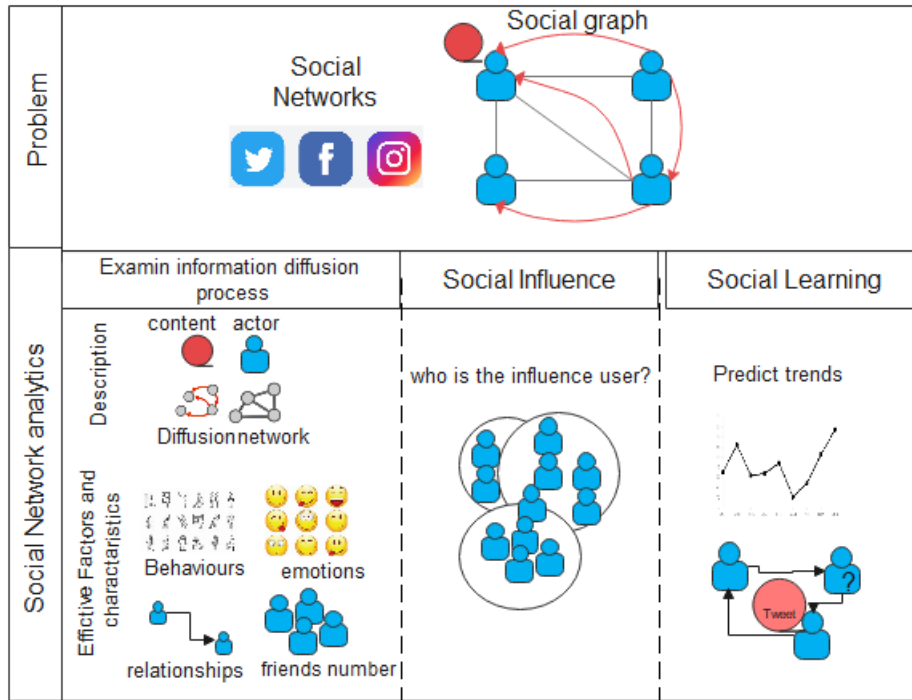


Fig. 1. Information Diffusion Process in Social Networks

this article makes the following contributions:

- 1) A more general perspective: we extend the diffusion models in social networks to include contagion, influence, and learning models.
- 2) Discuss different existing methods that included in these categories and addressing their pros and cons
- 3) Review information spreading in different social networks.

Every type of diffusion model will be discussed in the following sections, its methods, and related work. We will concentrate on three points., as illustrated in Fig. 2. Section 2: contagion diffusion models, epidemic models, and their applications will be studied. Epidemic models include Susceptible infected (SI), Susceptible infected Susceptible (SIS), Susceptible-Infected-Removal (SIR), Susceptible Infected Removed Susceptible (SIRS). Section 3: influence diffusion models, including independent cascade (IC) and Linear threshold, Influence maximization, and heuristic models. Section 4: Predictive diffusion models. And finally, we conclude the paper.

II. CONTAGION MODELS

When people are exposed to others who use or promote a new product or behavior, they can adopt it. In contagion diffusion models, the information adoption process is separated into several states to study the spread of information based on altering between them. This is similar to an epidemic-spreading process.[9], as clarified in Fig. 3. In the diffusion scenario, epidemic models assume N total persons. N has multiple primary states: S: susceptible. I: infected. R: recovered. From a social network perspective, e: exposed and c: contacted states are added.

A. Susceptible infected(SI)

In the SI model, at time t , $s(t)$ is the susceptible portion of the total population, $i(t)$ is the infected portion. This model

assumes that the total population is vulnerable to infection and that the daily contact rate is β . Equations (1) describe the SI model[10].

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

B. Susceptible infected Susceptible(SIS)

The issue is that the SI model prevents infected individuals from being cured. This is why a new model SIS represents cured patients with daily rates α . α indicates the portion of cured infected users who re-infected. From the SIS model, epidemic models are considered non-progressive models as activated nodes can be deactivated contrary to progressive models in social influence models where nodes, once activated always active. Equations (2) describe the SIS model.

$$\frac{di}{dt} = \beta i(1 - i) - \alpha i \quad (2)$$

C. The Susceptible-Infected-Recovered (SIR)

SIR model added a new state R: removed or recovered as A cured person can be an immune user. The recovered nodes cannot spread information. This means that the immune node can not be susceptible anymore. SIR model adds to SIS model, daily increase in immune users is expressed by $\frac{dr}{dt} = \lambda i$.

D. The SIRS (Susceptible Infected Removed Susceptible)

SIRS [10] model combines all the states as it assumes that a cured user can be again as a susceptible user with probability α . It demonstrates the diffusion process and the status of the users in social networks. Chao et al. [11] proposed the SEIR model by adding a new state to the SIR model E: Exposed, which means infected node but not infectious yet.

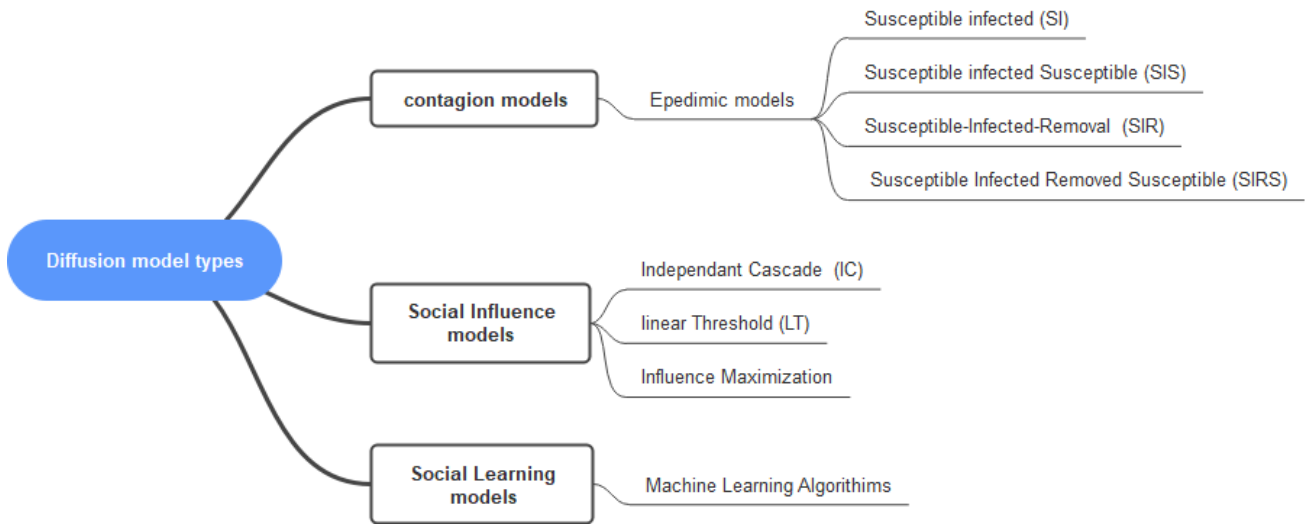


Fig.2. Diffusion Models types (contagion models, social influence models and social learning models)

They used a dynamical evolution equation to study user login frequency and friends numbers' effects on information spread. They found that user login frequency affects proportionally to data transfer. Also, individuals' behavioral features were examined, and they significantly impacted the information diffusion process. On the contrary, if it is ignored, it may result in an inaccurate diffusion model. Xu et al. [12] proved that information value as user behavior affects the diffusion process. As a result, they created an S-SEIR model for single-layer social networks. Wang et al. [13] extended the SCIR model to microblogs by adding a contacted (C) status that indicates whether a node is infectious or not, as shown in fig. 4. It assumes that when a user adopts an idea, all neighbors are Contacted. Then, depending on the probability, a neighbor will become an adopter or immune user. This model can accurately depict internet topic spreading.

Liang Mao [14] proposed a spatially explicit SIR-based model to visualize tree diffusion layers in one million-person city. As a result, a conceptual framework for the diffusion of influenza disease, information diffusion, and avoidance of harmful behavior in the social network was built. The model's results closely match influenza spread and information dispersion trends.

As mentioned before, contagion models are concerned with the dynamics of the process by dividing the population rumors. It involves information propagation, flow, and prevention behavior spread. These three processes interact and generate negative and positive feedback loops in human social networks of nodes into several statuses. Thus, Epidemic models aim to understand the diffusion of information.

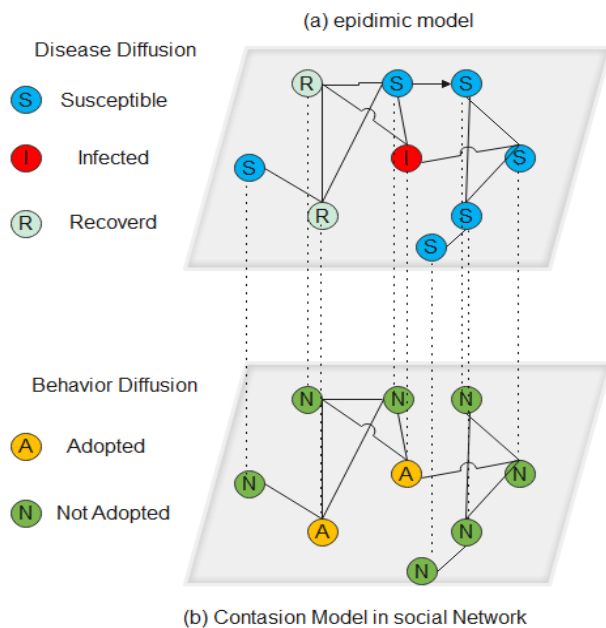


Fig. 3. The relation between Contagion model and information diffusion in social network

III. SOCIAL INFLUENCE

In social networks, users' interactions are represented as a directed graph with nodes representing users and edges representing relationships. When a user adopts information, the node will be active, and if not, it will be an inactive node [15]. Thus, nodes can influence each other and go from inactive to active but not the other way around. Social influence models, including Independent Cascade Model (IC), Linear Threshold Model (LT), and Influence Maximization (IM), will be discussed in the following sections.

A. Independent Cascade Model (IC)

IC assumes a user v is activated on every step by each of its connected neighbors u independently with probability $p_{u,v}$. In IC, u has a single chance to activate only one outgoing neighbor. Then u stops activating and stays active. Diffusion ends when no more nodes can be activated. Thus, IC is a progressive model as once a node is activated, it can not be deactivated again [15]. In some applications like opinion-aware, a submodularity function is needed to allow nodes to switch between positive and negative opinions through an influence graph

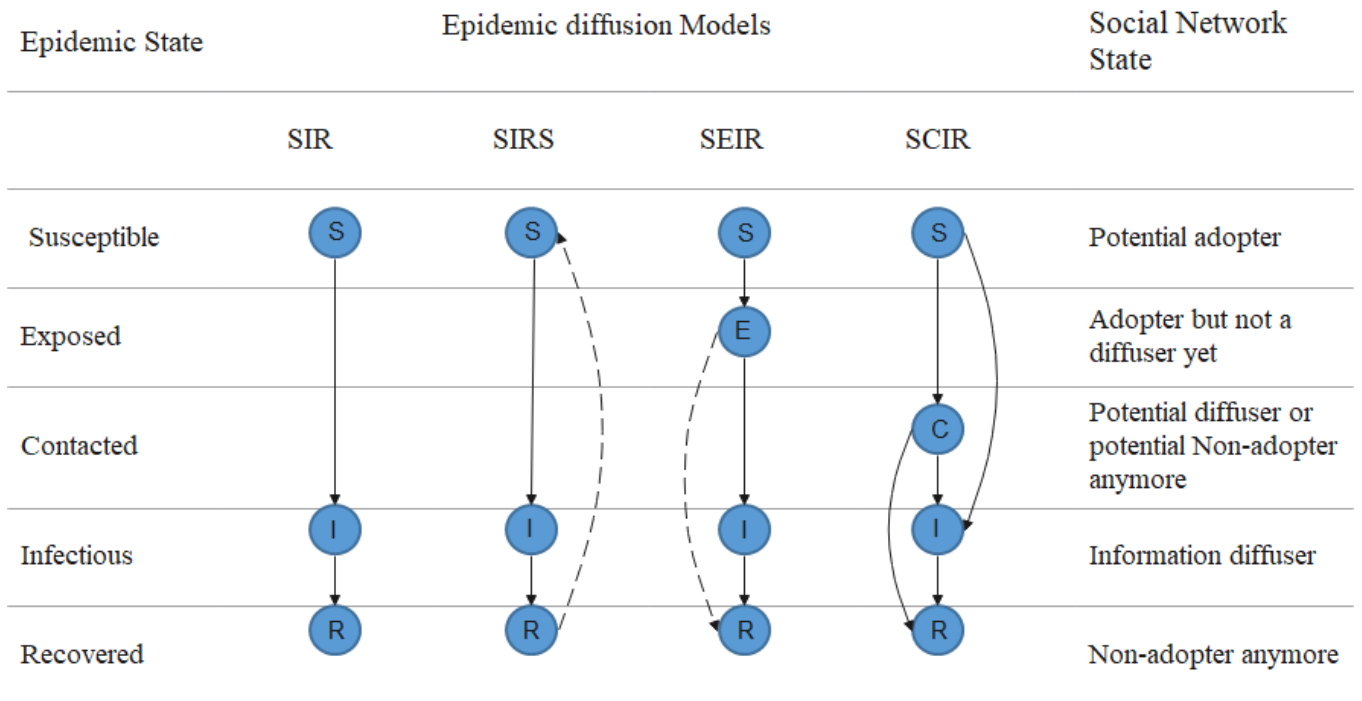


Fig. 4. Comparison of Epidemic models States

B. Linear Threshold Model (LT)

In threshold models, nodes behave differently if enough of their neighbors do by changing the threshold level, leading to different levels of adoption.

The LT model has an activation threshold for each active node v . When all active nodes' influence degree reaches v 's activation threshold, v becomes active at time $t + 1$. Thus, each neighbor can activate v times. The LT model studies influence in social networks, focusing on threshold behavior during influence spreading. The LT model has been used to optimize influence[17].

C. Influence Maximization (IM)

Social Influence involves identifying the most influential social network members. Kempe et al. [18] formulated this problem. Social influence aims to discover k nodes in a social network that maximize influence by activating neighboring nodes. Unfortunately, this maximum coverage optimization problem is non-deterministic polynomial-time (NP)-hard[8] for IC and LT models.

1) Greedy Algorithm

Most of the existing IM algorithms apply a simple greedy framework that chooses the active node with the highest marginal gain in each iteration. In a greedy algorithm, each option can significantly impact the node's impact value by using the optimal local solution to estimate the optimal global solution. Thus, the accuracy of the algorithm is relatively high. However, due to the algorithm's complexity and high execution time, it may be inefficient. Kempe et al. [18] developed a greedy approximation technique to address The maximum influence problem. Leskovec et al. [19] proposed

a greedy optimization method, the cost-effective lazy forward (CELFF) approach.

2) Heuristic algorithms

Heuristic algorithms have been introduced to increase efficiency by reducing complexity time, contrary to greedy algorithms that are computationally complex. However, instead of estimating each node's marginal gain, these heuristic algorithms pick them based on the connections between the nodes. The problem with heuristic algorithms is that they are pretty inaccurate. The next part states the most popular heuristic algorithms.

a) Degree centrality

Degree centrality is considered a local measure when computing influential users as it uses local measurements as the number of edges between nodes in a social graph. Degree centrality concerns the number of links connecting a node; the higher the degree node, the more evident the ability to increase information dissemination [20]. In social networks, degree counts represent the number of social relationships, friends, or followers and the number of interactions (retweets, shares) for each user. The problem with degree centrality is that it doesn't always provide high accuracy as users with a high degree are not necessarily considered influential seeds.

b) Closeness centrality

Closeness centrality determines a node's proximity to all other nodes in a social network. Closeness to any user is determined by the shortest path between two nodes in a graph. Closeness centrality is equal to the average distance between users.[20]

c) Betweenness centrality

The betweenness centrality is the count of the shortest paths that cross through a user's betweenness centrality [21]. For example, kourtellis et al. [22] applied betweenness centrality on Facebook data graphs to detect the network's central nodes representing the essential users.

d) Katz centrality

Katz centrality takes into account all network paths when calculating nodes' influence. It assigns a particular minimum score to every node in a network[23] to be calculated by all network links that pass through the node. Thus, Katz centrality considers all network links[24], not only the shortest path like closeness. However, Katz's centrality is the high computational complexity, making it difficult to use in extensive networks applications.

e) Eigenvector centrality

Eigenvector centrality calculate user influence by computing the influence score of connected users (3), taking into consideration the number of edges (adjacency matrix) [25]

$$x_i = \frac{1}{\lambda} A_{i,j}^T x_j \quad (3)$$

X_i is the influence for node i $A_{i,j}$ is the adjacency matrix, and λ happens to be the principal eigenvalue. Users with high scores are evidence that they are connected to influential users. User influence is proportional to the total of its associated users' influence scores[23].

f) PageRank-like algorithms

An iterative PageRank algorithm evaluates the value of a node based on the significant metric counts and associated link counts. Various applications utilize the PageRank algorithm or its variants to find essential nodes in a social graph. Yin et al.[26] extract unique features of users in the Sina Microblog such as their level of activity and readiness to retweet and then calculate user influence score using a weighted PageRank algorithm. They suggested a user interaction model that estimates pairwise influence rather than global influence.

g) Coreness-based Measures

The k-core decomposes the network to k part. In the k-core decomposition processes, all nodes with degrees (edge count for each node) less than the k are repeatedly eliminated.

The algorithm starts with users who have a degree one who is assigned to the 1-shell. All users with degree $k=1$ are first deleted [27]. Decomposition processes continue till no user with $k=1$ is found, as clarified in fig 5. The k-core algorithm gives more attention to node location than its degree; the deeper the node locates in the graph, the more it can be influential.

Zeng et al.[28] improved k-core by considering links between the remaining nodes and entirely deleting links connected to the deleted nodes to overcome the drawbacks of the k-shell decomposition process. Wei et al. [29] give high weight to specific nodes which connected to high degree nodes. Al-garadi et al. [30]exploit the importance of interaction in information diffusion and enhance k-core by using user interactions as a weighting factor for connections

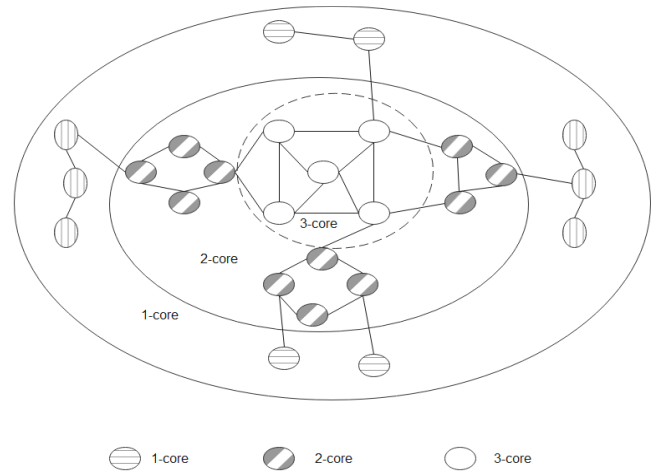


Fig. 5.K-core example with k=3

Table 1 summarizes the influence maximization problem methods regarding complexity, advantages, and disadvantages for each method.

IV. SOCIAL LEARNING

When users publish critical information on a social network, it rapidly propagates through the network. Especially in the case of negative news, It's important to know how things might turn out. Therefore, it's helpful to be able to predict future network information diffusion. Predictive models predict how information diffusion processes spread information throughout a network in the future. These models should learn to predict past diffusion traces in time and space. It is necessary to predict information spread to stop misinformation from being propagated in the network. Social learning diffusion models have two goals: first, to predict trendsetting users, and second, to predict the following influenced user who will retweet or repost information. There are two kinds of diffusion prediction models: macro and micro. Macroscopic diffusion prediction calculates the overall number of influenced users. Microscopic models seek to predict the next impacted person.

A. predicting influential users

Learning approaches anticipate influential users using machine learning algorithms, mainly supervised learning. Effective learning requires a rich set of characteristics with high discriminative power.

Mei et al. [31] added new features such as the number of reposts, sharing, retweets, tagging or mentioning others, the number of friends or followers, account information, the number of likes, comments, replies, number of posts, status, or tweets. They used a weighting entropy algorithm to train the previously eight mentioned features to predict user influence effectively. Liu et al. [32] extracted additional features like the number of other interactions (comments and responses) are extracted for a Support Vector Machine (SVM) training stage. Cossu et al. [33] investigated a large selection of standard features, which provided insignificant outcomes. New features were introduced to improve performance like hashtags, Shared URL links, indegree, closeness, eigenvalue, and betweenness centrality.

TABLE I. Comparison between influence maximization Models

References	GREEDY Algorithm		Heuristic Algorithm							Time Complexity	Pros	Cons
	IC	LT	DC	CC	BC	KC	EC	PR	K-CO			
[18] [19]	✓	✓	----	---	---	---	---	---	---	$O(krnm)$ Where k number of nodes and r number of rounds	<ul style="list-style-type: none"> accurate with approximation $(1-1/e-\epsilon)$ 	<ul style="list-style-type: none"> High computational time(inefficient) NP-hard
[20]	---	----	✓	---	---	---	---	---	---	$O(M)$, where, M represents number of edges.	<ul style="list-style-type: none"> Simple fast 	<ul style="list-style-type: none"> Measure local features of users
	---	----		✓	---	---	---	---	---	$O(N^3)$, where, N represents number of nodes.	<ul style="list-style-type: none"> Results in global impact 	<ul style="list-style-type: none"> Low efficiency as high computational run time Not suitable for massive social network
[21] [22]	---	----	----	---	✓	---	---	---	---	$O(N^3)$	<ul style="list-style-type: none"> Find shortest path Global measure 	<ul style="list-style-type: none"> Low efficiency as high computational run time Not suitable for massive social network
[23]	----	----	----	---	---	✓	✓	---	---	$O(N^3)$	<ul style="list-style-type: none"> Global measure 	<ul style="list-style-type: none"> Low efficiency as high computational run time Not suitable for massive social network
[24]	---	----	----	---	---	✓	---	---	---	$O(N^3)$	<ul style="list-style-type: none"> Simple effective 	<ul style="list-style-type: none"> Low efficiency as high computational run time Not suitable for massive social network
[25]	---	----	----	---	---	---	✓	---	---			
[26]	---	----	----	---	---	---	---	✓	---	$O(N+M)$	<ul style="list-style-type: none"> Simple Low complexity Global measure 	<ul style="list-style-type: none"> Unreliable Inaccurate noisy networks
[27:29]	--	--	--	--	--	--	--	--	✓	$O(N)$	<ul style="list-style-type: none"> Simple Low complexity Global measure 	<ul style="list-style-type: none"> Unweighted network

IC Information Cascade

LT Linear Threshold

DC Degree centrality

CC Closeness centrality

BC Betweenness centrality

KC Katz centrality

EC EigenVector Centrality

PR PageRank algorithms

K-CO K-corness

B. Retweet prediction

Information diffusion is affected by the speed of retweets and replies. Nguyen et al. [34] assume profiles with similar patterns are supposed to have the same behavior of tweeting and retweeting. They studied information strategies of users based on a set of features, including linguistics, freshness, trustability of information, and interest matching. The study also introduced a similarity model of the TF-IDF weighted

Bag of Words to predict retweeting behavior. Yang et al. [35] introduce reinforced recurrent networks with a structural model (FOREST) based on reinforcement learning and macroscopic predictions. First, it learns a cascade prediction model based on Reinforcement Neural Network (RNN) and then simulates results to predict influence in the social network. They used a Twitter dataset [36] that records the tweets spreading among users since October 2010. Also, they used the Douban data set [37], a Chinese social site where

individuals can update their reading statuses and follow others'. Finally, they used the Memetracker dataset [27], which contains a million news stories and blog posts and tracks the most frequent quotes, phrases, and memes. FOREST model was compared to multiple models. First, Topologically Long short-term memory (TopoLSTM) [38] extends the LSTM model by constructing the hidden states as a social graph. Second, DeepDiffuse [39] uses the embedding approach to utilize the infection timestamp data. Third, Neural Diffusion Model (NDM) [40] builds a microscopic cascade model to reduce long-term dependency using convolution neural networks (CNN). Finally, the Sequential neural information diffusion model with structure model (SNIDS) [41] computes pairwise similarities of all user pairs. FOREST enhanced results in comparison of mentioned methods [38–41] on information spread forecasting by more than 10% in terms of Mean Average Precision (MAP) scores and 12% in terms of Mean Square Log-transformed Error (MSLE)[42].

The shortcoming of social learning algorithms

We need enough training and testing data for learning approaches., which is difficult and expensive in both time and resources. It takes a lot of labeled training data to create a robust learning technique for locating significant users. To overcome this limitation, semi-supervised techniques are utilized with little labeled data.

Obtaining useful knowledge from small samples is mainly inaccurate [43]. Also, there is no evidence for suggested models in machine learning studies.[44]. Several factors affect machine-learning models' ability to measure user influence. Choosing the optimal attributes for discriminating between influential and noninfluential users is a complex issue. Most machine-learning algorithms select features[45] which can help determine the appropriate characteristics to train models. However, unlike average users, influential users are rare, causing class distributions to be distorted.

V. DISCUSSION AND CONCLUSION

This paper investigates contagion, Influence, and Learning models for information diffusion analysis in social networks. Despite adding more states to the basic epidemic model, the difficulty with contagion models is that they are still unsuitable for real-world social networks when varied user behaviors are taken into account. However, contagion models are pretty accurate regarding understanding the diffusion process. Contagion models are suitable for studying how information is spread. Therefore, these models are used in applications like disease modeling, news, opinion, rumors, natural disaster awareness[46], and trend-spreading. Also, these models play an excellent role in defining the source ex. Who spread the rumors?[47]

Meanwhile, in Social influence models, maximizing the social influence is the target. This can be achieved by combining multiple network and individual factors to discover influential users who can increase the spread process. Therefore, the influence model is often used in viral marketing and social advertising [48]. Social learning models are always based on the previous two models. In social learning, predictive models are used to predict diffusion in

the future. This can help restrain rumors, negative behaviors, malware from spreading through the social network.

From the previous discussion, we can conclude that these three models are not separated. They are always based on each other. We gave an outline of information diffusion models and social network analysis.. We proposed a taxonomy to categorize diffusion models related to literature, highlighting many effective methods in each category. We also addressed the pros and cons of different methods over time and outlined some challenges and applications.

REFERENCES

- [1] L. AlSuwaidan and M. Ykhlef, "Toward information diffusion model for viral marketing in business," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 2, pp. 637–646, 2016.
- [2] S. Jendoubi and A. Martin, "Evidential positive opinion influence measures for viral marketing," *Knowledge and Information Systems*, vol. 62, no. 3, pp. 1037–1062, 2020.
- [3] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: A survey," *ACM Sigmod Record*, vol. 42, no. 2, pp. 17–28, 2016.
- [4] J. Sun and J. Tang, "A survey of models and algorithms for social influence analysis," in *Social network data analytics*, Springer, 2011, pp. 177–214.
- [5] S. Dhama, K. Prabuchandran, and Y. Narahari, "Information diffusion in social networks in two phases," *IEEE Transactions on Network Science and Engineering*, vol. 3, no. 4, pp. 197–210, 2016.
- [6] Ç. Aslay and others, "From viral marketing to social advertising: ad allocation under social influence," Universitat Pompeu Fabra, 2016.
- [7] H. P. Young, "Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning," *American economic review*, vol. 99, no. 5, pp. 1899–1924, 2019.
- [8] W. Chen, L. V. Lakshmanan, and C. Castillo, "Information and influence propagation in social networks," *Synthesis Lectures on Data Management*, vol. 5, no. 4, pp. 1–177, 2013.
- [9] B. Qu, A. Hanjalic, and H. Wang, "Heterogeneous recovery rates against sis epidemics in directed networks," in *2014 7th International Conference on NETWORK Games, Control and Optimization (NetGCoop)*, 2014, pp. 173–179.
- [10] M. Li, X. Wang, K. Gao, and S. Zhang, "A survey on information diffusion in online social networks: Models and methods," *information*, vol. 8, no. 4, p. 118, 2017.
- [11] C. WANG, X. YANG, K. XU, and J. MA, "SEIR-based model for the information spreading over SNS," *ACTA ELECTRONICA SINICA*, vol. 42, no. 11, p. 2325, 2014.
- [12] R.-Z. Xu, H.-L. Li, and C.-M. Xing, "Research on information dissemination model for social networking services," *International Journal of Computer Science and Application*, vol. 2, no. 1, pp. 1–6, 2013.
- [13] C. Wang, X. Guan, T. Qin, and Y. Zhou, "Modelling on opinion leader's influence in microblog message propagation and its application," *J. Softw.*, vol. 26, pp. 1473–1485, 2015.
- [14] L. Mao, "Modeling triple-diffusions of infectious diseases, information, and preventive behaviors through a metropolitan social network—an agent-based simulation," *Applied Geography*, vol. 50, pp. 31–39, 2014.
- [15] W. Wang and W. N. Street, "Modeling and maximizing influence diffusion in social networks for viral marketing," *Applied network science*, vol. 3, no. 1, pp. 1–26, 2018.
- [16] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," *Marketing letters*, vol. 12, no. 3, pp. 211–223, 2001.
- [17] X. He and D. Kempe, "Stability and robustness in influence maximization," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 12, no. 6, pp. 1–34, 2018.
- [18] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 137–146.
- [19] A. Goyal, W. Lu, and L. V. Lakshmanan, "Celf++ optimizing the greedy algorithm for influence maximization in social networks," in *Proceedings of the 20th international conference companion on World wide web*, 2011, pp. 47–48.

- [20] D. L. Hansen, B. Shneiderman, M. A. Smith, and I. Himelboim, "Chapter 3 - Social network analysis: Measuring, mapping, and modeling collections of connections," in *Analyzing Social Media Networks with NodeXL (Second Edition)*, Second Edition., D. L. Hansen, B. Shneiderman, M. A. Smith, and I. Himelboim, Eds. Morgan Kaufmann, 2020, pp. 31–51.
- [21] L. Zhang, M. Luo, and R. J. Boncella, "Product information diffusion in a social network," *Electronic Commerce Research*, vol. 20, no. 1, pp. 3–19, 2020.
- [22] N. Kourtellis, T. Alahakoon, R. Simha, A. Iamnitchi, and R. Tripathi, "Identifying high betweenness centrality nodes in large social networks," *Social Network Analysis and Mining*, vol. 3, no. 4, pp. 899–914, 2013.
- [23] H. Liao, M. S. Mariani, M. Medo, Y.-C. Zhang, and M.-Y. Zhou, "Ranking in evolving complex networks," *Physics Reports*, vol. 689, pp. 1–54, 2017.
- [24] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, "Vital nodes identification in complex networks," *Physics Reports*, vol. 650, pp. 1–63, 2016.
- [25] E. J. Bienenstock and P. Bonacich, "Eigenvector centralization as a measure of structural bias in information aggregation," *The Journal of Mathematical Sociology*, vol. 0, no. 0, pp. 1–19, 2021.
- [26] Z. Yin, Y. Zhang, W. Chen, and R. Zong, "Discovering patterns of advertisement propagation in sina-microblog," in *Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy*, 2012, pp. 1–9.
- [27] W. Khaouid, M. Barsky, V. Srinivasan, and A. Thomo, "K-core decomposition of large networks on a single pc," *Proceedings of the VLDB Endowment*, vol. 9, no. 1, pp. 13–23, 2015.
- [28] A. Zeng and C.-J. Zhang, "Ranking spreaders by decomposing complex networks," *Physics Letters A*, vol. 377, no. 14, pp. 1031–1035, 2013.
- [29] B. Wei, J. Liu, D. Wei, C. Gao, and Y. Deng, "Weighted k-shell decomposition for complex networks based on potential edge weights," *Physica A: Statistical Mechanics and its Applications*, vol. 420, pp. 277–283, 2015.
- [30] Al-garadi Mohammed Ali, K. D. Varathan, and S. D. Ravana, "Identification of influential spreaders in online social networks using interaction weighted K-core decomposition method," *Physica A: Statistical Mechanics and its Applications*, vol. 468, pp. 278–288, 2017.
- [31] Y. Mei, Y. Zhong, and J. Yang, "Finding and analyzing principal features for measuring user influence on Twitter," in *2015 IEEE First International Conference on Big Data Computing Service and Applications*, 2015, pp. 478–486.
- [32] N. Liu, L. Li, G. Xu, and Z. Yang, "Identifying domain-dependent influential microblog users: A post-feature based approach," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2014, vol. 28, no. 1.
- [33] J.-V. Cossu, N. Dugué, and V. Labatut, "Detecting real-world influence through twitter," in *2015 Second European Network Intelligence Conference*, 2015, pp. 83–90.
- [34] D.-A. Nguyen, S. Tan, R. Ramanathan, and X. Yan, "Analyzing information sharing strategies of users in online social networks," in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016, pp. 247–254.
- [35] C. Yang, J. Tang, M. Sun, G. Cui, and Z. Liu, "Multi-scale Information Diffusion Prediction with Reinforced Recurrent Networks," in *IJCAI*, 2019, pp. 4033–4039.
- [36] N. O. Hodas and K. Lerman, "The simple rules of social contagion," *Scientific reports*, vol. 4, no. 1, pp. 1–7, 2014.
- [37] E. Zhong, W. Fan, J. Wang, L. Xiao, and Y. Li, "Comsoc: adaptive transfer of user behaviors over composite social network," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012, pp. 696–704.
- [38] J. Wang, V. W. Zheng, Z. Liu, and K. C.-C. Chang, "Topological recurrent neural network for diffusion prediction," in *2017 IEEE International Conference on Data Mining (ICDM)*, 2017, pp. 475–484.
- [39] M. R. Islam, S. Muthiah, B. Adhikari, B. A. Prakash, and N. Ramakrishnan, "Deepdiffuse: Predicting the 'who' and 'when' in cascades," in *2018 IEEE International Conference on Data Mining (ICDM)*, 2018, pp. 1055–1060.
- [40] C. Yang, M. Sun, H. Liu, S. Han, Z. Liu, and H. Luan, "Neural diffusion model for microscopic cascade prediction," *arXiv preprint arXiv:1812.08933*, 2018.
- [41] Z. Wang, C. Chen, and W. Li, "A sequential neural information diffusion model with structure attention," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018, pp. 1795–1798.
- [42] Q. Cao, H. Shen, K. Cen, W. Ouyang, and X. Cheng, "Deephawkes: Bridging the gap between prediction and understanding of information cascades," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 1149–1158.
- [43] M. A. Al-Garadi, K. D. Varathan, S. D. Ravana, E. Ahmed, G. Mujtaba, M. U. S. Khan, and S. U. Khan, "Analysis of online social network connections for identification of influential users: Survey and open research issues," *ACM Computing Surveys (CSUR)*, vol. 51, no. 1, pp. 1–37, 2018.
- [44] S. Rübiger and M. Spiliopoulou, "A framework for validating the merit of properties that predict the influence of a twitter user," *Expert Systems with Applications*, vol. 42, no. 5, pp. 2824–2834, 2015.
- [45] P. Shao and H. Chen, "Driving factors for opinion diffusion behavior in consumers on online social networks: A study of network characteristics," *IEEE Access*, vol. 7, pp. 118509–118518, 2019.
- [46] R. Dong, L. Li, Q. Zhang, and G. Cai, "Information diffusion on social media during natural disasters," *IEEE transactions on computational social systems*, vol. 5, no. 1, pp. 265–276, 2018.
- [47] G. Song, Y. Li, X. Chen, X. He, and J. Tang, "Influential node tracking on dynamic social network: An interchange greedy approach," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 2, pp. 359–372, 2016.
- [48] H. T. Nguyen, T. N. Dinh, and M. T. Thai, "Cost-aware targeted viral marketing in billion-scale networks," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, 2016, pp. 1–9.