Social Influence Analysis in Online Social Networks for Viral Marketing: A Survey

Halima Baabcha, Meriem Laifa, and Samir Akhrouf

Abstract One of the most exciting developments of the last few years has been the rise of online social networks. The richness of this network's content provides unprecedented opportunities for data analytics, which can be taken advantage of. One of the most important areas of social network analysis is the study of social influence. It can be used in a variety of ways, from viral marketing to advertising. In addition to identifying influential nodes in a social network, the modeling of influence diffusion and influence maximization in social networks is an important challenge in this area. There has been a lot of research done on the influence of online social networks, particularly in viral marketing contexts, for various applications. Methods for influence modeling, maximization, and identifying influential nodes are discussed in this chapter. Using cutting-edge research on viral marketing's impact on social influence, we hope to serve as a resource for aspiring researchers.

Keywords Online social networks · Influence analysis · Viral marketing · Influence maximization · Influential spreaders

1 Introduction

The recent rapid increasing popularity of online social networks (OSN) and new communication technologies availability (smartphone, tablet, etc.) provided the world population with a great facility to share and to communicate with others throughout the whole world and to share and exchange information, opinions, products, ideas, and services. The rich content of OSN that can be in diverse

---

H. Baabcha (✉) · M. Laifa
Department of Computer Science, University Mohammed El Bachir El Ibrahimi, Bordj Bou Arreridj, Algeria
e-mail: halima.baabcha@univ-bba.dz; meriem.laifa@univ-bba.dz

S. Akhrouf
Department of Computer Science, University Mohamed Boudiaf, M'sila, Algeria
e-mail: samir.akhrouf@univ-msila.dz

formats (text, image, video, audio, etc.) has been attracting researchers to study and analyze large-scale social structures and users' behaviors in order to – among other purposes – understand the flow of information through social networks. A lot of real-world applications like public opinion monitoring, recommender systems, and political campaigns often make use of OSNs for influence diffusion (Fu et al. 2016; Can and Alatas 2019; Saxena and Saxena 2020). Many existing models for influence diffusion have been proposed for various applications. In this chapter, we look into influence diffusion in the setting of viral marketing.

In essence, viral marketing is an efficient solution to advertisement for commercial companies through OSN where companies try to promote their products and services through word-of-mouth propagation among friends or followers. One of the fundamental objective of viral marketing is to find a set of users with the maximum influence in the network where the output is called K-seed set users with K as the optimal number of users chosen for influencing other users in the network.

The goals of this chapter are to highlight some of the most used and most recent work that has been done in social influence diffusion and influence maximization model for viral marketing and to call attention to the topic of deep learning for social influence. To the best of our knowledge, this work is a literature review of influence analysis for viral marketing in online social networks. Firstly, we give a better understanding of the preliminary knowledge concerning social influence analysis, and we illustrate the categories of relevant research works on influence analysis in the context of viral marketing. After that, we categorize and compare a number of relevant research works on influence maximization algorithms in social networks and the identification of influential spreaders.

In the following section, we describe the main concepts that will be addressed throughout the chapter, namely, online social networks, social network analysis, social influence analysis, viral marketing, and the definition problem of influence maximization. Section 3 presents related work to diffusion influence modeling, identification of influential spreaders, and influence maximization. Section 4 exhibits the current methodology using deep learning for influence spreading modeling. Finally, we conclude the chapter in Sect. 5.

2 Background

In this section, we begin with an overview of the basic terminology.

2.1 Online Social Networks

OSNs can be defined within the context of systems, but in general, they can be defined as a network of interactions or relationships, where nodes consist of actors or persons and the links consist of the relationships or the interactions between

the actors through the network. According to (Kemi 2016), an OSN alternatively referred to as a virtual community or profile site, a social network is a website on the Internet that brings people together in a central location to talk, share ideas and interests, or make new friends. With the emergence of the World Wide Web (WWW), OSNs have dramatically expanded in popularity around the world. For further details about online social networks and their features, we recommend readers to check (Fu et al. 2016; Boyd and Ellison 2007).

2.2   Social Network Analysis

Social network analysis (SNA) or social network mining is the research of social relations between nodes or people. The study of social network mining technologies focuses on the level of individuals, groups, organizations, and whole networks. It is the drawing and determining associations and drifts among users and other interlinked information entities using networks and graph theory (Otte and Rousseau 2002). Many research applications benefit from SNA techniques such as community or group detection (Cai et al. 2016; Dabaghi-Zarandi and Rafsanjani 2019), expert finding (Yuan et al. 2020), link prediction (Daud et al. 2020), recommender systems (Pourhojjati-Sabet and Rabiee 2020), predicting trust and distrust among individuals (Towhidi et al. 2020; Girdhar et al. 2019), influence propagation (Ortiz-Gaona et al. 2020; Abd Al-Azim et al. 2020; Singh et al. 2019; Gulati and Eirinaki 2018), etc.

2.3   Social Influence Analysis

According to (Sun and Tang 2011), influence is usually reflected in changes in social action patterns (i.e., user behavior) in a social network. Typically, it refers to the phenomenon that an individual's emotions, opinions, or behaviors are affected by others (Qiu et al. 2018). This means that people are influenced by their social circle and tend to imitate the actions of those in their immediate vicinity. As a result of its wide range of real-world applications, social influence analysis has received considerable attention in the past (Peng et al. 2016) such as domain expert finding (Al-Taie et al. 2018), personal recommendation (Cheng et al. 2016), emotion prediction (Qiyao et al. 2016), and viral marketing (Talukder et al. 2017; Bhattacharya et al. 2019; Menta and Singh 2017) for which it became an important strategy.

2.4 Viral Marketing

Viral marketing (VM) is one of the various real-world applications of social influence analysis. According to (Wang and Street 2018), VM is a process of influence diffusion over social networks. It is a relatively recent solution to advertisement within online social networks. It has been applied to business-to-consumer transactions.

Using a social network to spread the word about a product or service is a form of viral marketing. In other words, it employs customers in a market to promote a product (McKay et al. 2019). For example, if a company has a certain number of new products, they could hand them out to a customer, and then the influence model maximization can predict optimal objective to get these products in order to spread the product's influence over a specific network. There is a dearth of new diffusion methods in the literature, particularly for dynamic and massive networks. Additionally, it provides information on the various mining techniques that can be used for viral marketing.

On the other hand, the goal of viral marketing is to minimize marketing cost while maximizing the profit. The main idea of viral marketing is to find a set of customers for giving free samples within the budget B to maximize the expected total sales of the product, in other words use the K-seed set users for influencing other users in the network. Because the majority of the promotional work is done by customers, this type of "word-of-mouth" advertising can be far more cost-effective than more traditional ones. Friends' recommendations are more trustworthy than those made by a company selling the product (Richardson and Domingos 2002). Figure 1 shows the viral marketing process.

2.5 Influence Maximization Problem

A social network is depicted as a graph $G (V, E)$ with a set of nodes V that represents individuals and a set of edges E that represents the relationship shared among the nodes in the graph. The influence maximization problem takes as input a graph $G (V, E)$. The goal of this problem is to identify a subset of users in graph G who have the greatest amount of influence. An initial influence on this problem should yield the maximum number of nodes in the graph G that are influenced by a K-sized seed set, which is the solution to this problem's problem (Du et al. 2019). Figure 2 shows the input and output for influence maximization problem. The important objective of the problem of influence maximization is to find a set of users with that maximum influence in a graph.
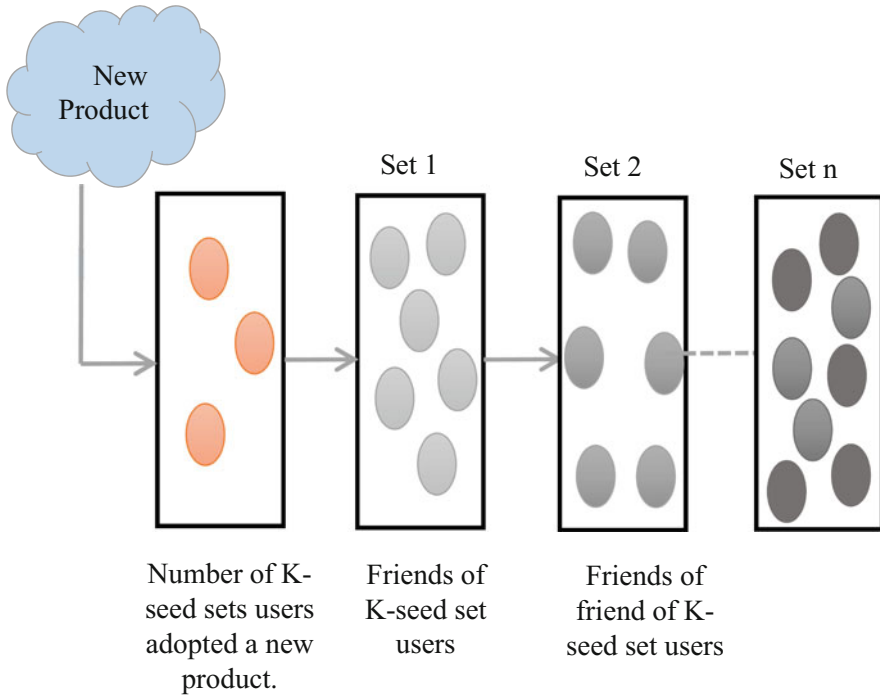
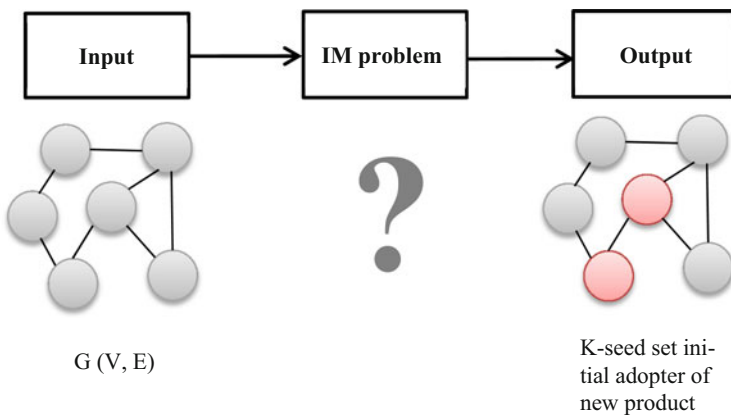**Fig. 1** Viral marketing via social network process

Fig. 2 Input and output of influence maximization problem

3 State of the Art

In this section, we review the most important and most cited existing approaches for influence diffusion, influence maximization, and identification of influential users.

3.1 Influence Diffusion Model

Influence diffusion has become an important technique for viral marketing. The Oxford Dictionary defines diffusion as "the spread of something." In social network analysis, diffusion is the process of information diffusion via the network. The majority of current research in SNA focuses on information and influence diffusion in online social networks (Guille et al. 2013; Arnaboldi et al. 2014; Xu et al. 2014; Sun et al. 2019; Dhamal et al. 2016; Gaeta 2018; Kong et al. 2020). According to (AlSuwaidan and Ykhlef 2016), diffusion models were originally used in social networks to simulate the process of information and influence propagation in the network. Many models and algorithms for influence diffusion have been proposed (More and Lingam 2019; Toalombo et al. 2020; Li and Liu 2019; Pan et al. 2020). In these models, each node is either active or inactive over iterations. An inactive node becomes active as more of its neighbors became active. In (Richardson and Domingos 2002), the authors provide the first algorithmic treatment to deal with the influence propagation problem. They built probabilistic models and used these models to choose the best viral marketing plan. Then the authors in (Kempe et al. 2003) studied influence propagation by focusing on the modeling influence by two fundamental stochastic influence cascade graph-based models, named independent cascade model (ICM) and linear threshold model (LTM). These models are based on directed graphs where each node can be activated or not with a monotonicity assumption (i.e., activated nodes cannot be deactivated). They formulated the problem as a network $G = (V, E, p)$, where V is the set of nodes and E is the set of edges between nodes and p is the probability that node v can successfully activate node u, denoted by $p(u,v)$. If a node accepts data from other nodes, it is considered active; otherwise, it is considered inactive (Du et al. 2019).

3.1.1 Linear Threshold Model (LTM)

In the linear threshold model (LTM), each edge or link $e(u,v)$ is associated with a weight $W(u, v)$, such that the sum of the weights of incoming neighbors of node v is less than or equal to 1 and each node v is also associated with a threshold θ_v. The linear threshold model starts with some active nodes with all other nodes being inactive and a random choice of thresholds θ.

The LTM samples the value of v of each user v uniformly at random probability from [0,1]. In step 0, it sets the status of nodes in S as active and others as inactive.

Then, it updates the status of each user iteratively. In step *t*, all nodes that were active in step *t-1* remain active, and any user *v* that was inactive in step *t-1* switches to active. The influence spread of seed set *S* under the LT model (i.e., $\sigma(S)$) is the expected number of activated nodes when *S* is initially activated.

3.1.2 Independent Cascade Model (ICM)

In the independent cascade model (ICM), a probability *p(u,v)* is associated with each edge *e(u,v)*, whereas *u* and *v* are two nodes in the graph. *p(u,v)* is the probability of the ability that *u* succeeds in activating *v*. In this model, a node *v* is activated by each of its incoming neighbors independently by introducing an influence probability *p(u,v)* to each edge *e(u,v)*. This model's diffusion instance unfolds in discrete steps according to influence probabilities and seed sets S at time step 0. Each active node *u* at step *t* will activate each of its outgoing neighbor *v* that is inactive in step *t-1* with probability *p(u,v)*. The activation process can be considered as flipping a coin with head probability *p(u,v)*: if the result is heads, then *v* is activated; otherwise, *v* stays inactive. When there are no more nodes that can be activated, the diffusion instance ends. The expected number of activated nodes when S is used as the initial active node set and the above stochastic activation process is applied is the influence spread of seed set S under the ICM.

3.1.3 Epidemic Model

An epidemic model is a perfect tool for a simplified description of the diffusion strategy that contagious diseases follow in a population. As an epidemic spreads from an infected individual to another healthy one (i.e., non-infected before), the information can also spread from one individual to another through the same network that interconnects them. Epidemic models assume the existence of an implicit network between individuals (i.e., no explicit connections) and assume that exposure to infection (information being diffused) is enough to become infected (informed) and potentially transmit the infection to someone else. The underlying principles of those techniques are the basis of the models used in marketing for the prediction of new product adoption in the communities (Loucif 2016). One of the most popular model SIR or the model of Kermack and McKendrick (Cano 2020) is a mathematical approach created particularly for studying the plague disease that broke out in Bombay. This mathematical model is built upon a set of hypotheses, namely:

(a) In the population, all individuals are sensitive equally to the infection.
(b) The infection leads either to death or to a permanent immunity.
(c) When healthy and infected individuals are living together, there will be a number of healthy individuals who will become infected.

In this model, the individuals can be found in three different states:

Suspected (S) An individual is said susceptible, which means that he is very capable to be infected with the disease. Generally, infections can originate from outside of the population in which the disease spreads (e.g., by genetic mutation, contact with an animal, etc.). Denote $S(t)$ the number of individuals who may be infected with the disease at time t and $Sf(t)$ the current fraction of the population that is susceptible.

$$Sf(t) = S(t)/N \tag{1}$$

Infected (I) After an individual is affected by the disease, he becomes infectious, i.e., has the ability to infect other susceptible individuals in the population. Let $I(t)$ be the number of infected individuals at time t. $If(t)$ refers accordingly to the infected fraction of the population.

$$If(t) = I(t)/N \tag{2}$$

Recovered (R) It refers to the individuals who are either cured of the disease and acquired a full or partial immunity against the infection (can no longer be infected) or removed after being killed by the infection. Let $Rf(t)$ be the fraction of the healed (or withdrawn) population where $R(t)$ refers to its size.

$$Rf(t) = R(t)/N \tag{3}$$

The diffusion of the disease within the population is dynamic: the fractions of susceptible, infectious, and healed individuals evolve over time with respect to the contacts through which the disease passes from infected individuals to healthy ones.

It is worth mentioning that at every moment (Zafarani et al. 2014),

$$1 = Sf(t) + If(t) + Rf(t) \tag{4}$$

3.2 Identification of Influential Spreaders

A challenging issue in viral marketing is effectively identifying a set of influential users. By sending the advertising messages to this set, one can reach out to the largest area of the network. It is now much easier to identify the network's most influential spreaders (Bhat et al. 2020). In this part, we have selected the most recent work in this field.

The authors in (Okamoto et al. 2008) combine existing methods on calculating exact values and approximate values of closeness centrality and presented a new algorithm to rank the top-k nodes in the network with the highest closeness centrality.

In (Bae and Kim 2014), in an effort to better understand the spread of a node's influence in a network, the researchers developed a new measure called coreness centrality. For their experiment, they used unweighted undirected graph for both real and artificial networks. Their approach is based on the idea that a powerful spreader has more connections to the nodes that reside in the core of network. The K-shell indices of nodes neighbors could be good indicators of its spreading ability. To evaluate their proposed measure, they applied the (SIR) model for investigating an epidemic spreading process. They evaluated the performance of the ranking measures in 12 real networks with different sizes as shown in Table 1.

The study in (Basaras et al. 2013) introduced a new centrality measure; it is a combination of coreness and betweenness centrality. To evaluate their technique's accuracy, they compared it to K-shell decomposition and a baseline measure based solely on the node degree on a large number of complex networks. They used the susceptible-infected-recovered model for an infection originating from both a single spreader and multiple spreaders to investigate the spreading process.

The authors in (Zeng and Zhang 2013) proposed a mixed degree decomposition (MDD) procedure in which both the residual degree and the exhausted degree are considered. By simulating the epidemic spreading process on real networks (Dolphins, Jazz, NetSci, Email, HEP, PGP, TAP, Y2H, Power, Internet, *E. coli*, *C. elegans*, AstroPh), they used the K-shell to generate the influence.

The authors in (Liu et al. 2018) proposed a local h-index centrality (LH-index) method for identifying and ranking the top influential spreaders in networks by calculating the h-indices of the node. The new proposed local h-index (LH-index) method simultaneously considers two factors: the h-index value of the node itself and the h-index values of its neighbors. On the one hand, the h-index of one node indicates the direct influences exerted by its nearest influential neighbors. On the other hand, the h-index values of its neighbors indicate the two-hop indirect influences exerted by further influential neighbors. The performance of their method showed its superiority in both real-world and simulated networks. They adopted the SIR model to evaluate the real spreading ability of the ranking nodes.

The work presented in (Belfin and Bródka 2018) uses multiple measures of centrality to look for overlapping communities and combine them to find a suitable superior seed set. They used degree, eigenvector centrality, and clustering coefficient. The basic idea of the strategy is to find out a fraction of superior nodes of the input network, called superior seed set, around which local communities can be computed.

Finally, the authors of (Bhat et al. 2020) presented the Improved Hybrid Rank algorithm, which combines two centralities, namely, the extended neighborhood coreness centrality and the h-index centrality. For the simulation of their proposed method, they used the SIR (susceptible-infected-recovered) model for both undirected and directed real-world networks. They have tested their algorithm based on various performance matrices like Kendall-Tau's correlation coefficient, spreader's location diversity, and infected scale.

The comparison of identification of influential spreader models is listed in Table 1.

Table 1 The comparison of performance for related identification of influential spreader models

References	Contribution	Used measures	Diffusion model	Graph	Dataset
Okamoto et al. (2008)	New algorithms to rank the top-k vertices with the highest closeness centrality for	Combined existing methods on calculating exact values and approximate values of closeness centrality	/	Directed	/
Kitsak et al. (2010)	K-shell centrality	/	SIS/SIR	Directed	LiveJournal.com community The network of email contacts in the Computer Science Department of University College London The contact network of inpatients (CNI) collected from hospitals in Sweden The network of actors who have costarred in movies labelled by imdb.com as adult

				Unweighted/undirected	
Bae and Kim (2014)	A novel measure coreness centrality	K-shell indices	SIR	Unweighted/undirected	Zachary's karate club Lusseau's bottlenose dolphins Jazz musician's network *C. elegans* metabolic network Coauthorship network of scientists (NetScience) Email network of URV Communication network of blogs Western States Power Grid PGP network Collaboration networks of astrophysics (CA–AstroPh) and condensed matter physics (CA–CondMat) Enron email network
Basaras et al. (2013)	A new centrality measure U-power community index (U-PCI)	Coreness and betweenness	SIR	Undirected	CA-CondMat CA-AstroPh From the Stanford Network Analysis Project

(continued)

Table 1 (continued)

References	Contribution	Used measures	Diffusion model	Graph	Dataset
Zeng and Zhang (2013)	Mixed degree decomposition (MDD)	Amelioration of K-shell	SIR	Undirected/directed	Dolphins (friendship) Jazz (musical collaboration) Netsci (collaboration network of network scientists) Email (communication) HEP (collaboration network of high-energy physicists) PGP (an encrypted communication network) AstroPh (collaboration network of astrophysics scientists) CondMat (collaboration network of condensed matter scientists) Nonsocial networks are word (adjacency relation in English text) *E. coli* (metabolic) *C. elegans* (neural) TAP (yeast protein-protein binding network generated by tandem affinity purification experiments) Y2H (yeast protein-protein binding network generated using yeast two hybridization) Power (connections between power stations) Internet (router level)

Ma et al. (2016)	Gravity centrality index	/	SIR	Undirected/unweighted	Facebook (Slavko Žitnik's friendship network in Facebook)
					NetSci (collaboration network of network scientists)
					Email (email network of Universitat Rovira i Virgili, URV)
					TAP (yeast protein-protein binding network generated by tandem affinity purification experiments)
					Y2H (yeast protein-protein binding network generated using yeast two hybridization)
					Blogs (the communication relationships between owners of blogs on the MSN (Windows Live) Spaces website)
					Router (the router-level topology of the Internet)
					HEP (collaboration network of high-energy physicists)
					PGP (an encrypted communication network)

Table 1 (continued)

References	Contribution	Used measures	Diffusion model	Graph	Dataset
Ahajjam and Badir (2018)	Hybrid rank algorithm	New hybrid centrality measure	SIR	Undirected/unweighted	CondMat (condense matter physics) DBLP bibliography WikiVote The Epinions
Liu et al. (2018)	H-index centrality and extended H-index	H-indices	SIR	Undirected/unweighted	USAir: The network of the US air transportation Blogs: The network of the communication Relationships between owners of blogs on the MSN (Windows Live) Spaces website Email: The network of email interchanges between members of the Universitat Rovira i Virgili Power grid: The network of the power grid of the Western States in the USA
Belfin and Bródka (2018)	Overlapping-based spreader selection	Degree, clustering coefficient, eigenvector centrality	SIR	Undirected	Zachary's karate club Dolphins social network American College Football

Richardson and Domingos (2002)	Influence deep learning (IDL) model		/	Neural network learning	Sina Weibo Epinions WikiVote NetHEPT
Bhat et al. (2020)	Improved Hybrid Rank algorithm	Extended neighborhood coreness centrality H-index centrality	/	Directed/undirected	CondMat Brightkite PGP network HyperPhysics P2P Gnutella Football

3.3 Influence Maximization

In this section, we present influence maximization-related research in viral marketing, and we review the important available research progress.

Domingos and Richardson (2001) were among the first to model customers' network value, they used markov random field for modeling the influence between customers such as nodes representing the customers. After that, they extended their previous techniques, achieving a large reduction in computational cost, and applied them to data from a knowledge-sharing site. They founded optimal marketing plan, and they used continuously valued marketing actions and reduce computational cost (Richardson and Domingos 2002). The most cited papers on the matter, maximizing the spread of influence through a social network written by (Kempe et al. 2003), in which the random degree-based and distance centrality algorithms are used as baselines, which led to the development of the greedy algorithm for influence maximization. More generally, they developed an algorithm for selecting the optimal seed set S from nodes in the graph. They proved that the optimization problem is NP-hard under LTM and ICM, and they presented a greedy algorithm that guarantees that the influence spread is within $(1–1/e − \varepsilon)$ of the optimal influence spread, where e is the base of natural logarithm and ε depends on the accuracy of their Monte Carlo estimate of the influence spread given a seed set. A series of more efficient studies have been done, because the greedy algorithm is infeasible even for medium-sized networks of tens of thousands of nodes and edges (Wang and Street 2018).

The work presented in (Leskovec et al. 2007) exploited submodularity to create an algorithm that can handle large-scale problems, achieve near-optimal placements, and be 700 times faster than a simple greedy algorithm. They proposed the Cost-Effective Lazy Forward (CELF) algorithm based on a "lazy forward" optimization. The obtained solutions are guaranteed to achieve at least a fraction of $1/2(1–1/e)$ of the optimal solution. They evaluated their algorithm on several large-scale real-world problems, including a model of a water distribution network and real blog data.

Authors in (Goyal et al. 2011) introduced an algorithm called CELF++ that further optimizes CELF by exploiting submodularity property. By avoiding redundant re-calculations of CELF's marginal gains, the algorithm CELF++ has an advantage.

The authors in (Chen et al. 2009) studied the efficient influence maximization in social networks from two complementary directions. One is to improve the original greedy algorithm in (Kempe et al. 2003) and its improvement in (Leskovec et al. 2007) to further reduce running time for the greedy algorithm. After that, they proposed a new degree discount heuristic derived from the independent cascade model that improves influence. They also proposed an algorithm called new greedy algorithm for the influence maximization.

The study in (Chen et al. 2010a) proposed a new heuristic algorithm that is easily scalable to millions of nodes and edges in their experiments. An easy-to-

use tunable parameter allows users to balance the running time and the spread of the algorithm's influence in the general ICM. For the experiments, they used four real-world network and a synthetic dataset (NetHEPT, DBLP, Epinions, and Amazon).

The work in (Chen et al. 2010b) shows that to compute the expected influence spread for a given set is P-hard. But it can be expressed as a submodular monotone function of S, which can be used to guarantee the results using a simple greedy algorithm. As an added bonus, we now have the LDAG algorithm, the first-ever scalable heuristic algorithm designed specifically for LT model influence maximization. They firstly disclosed that the computation of influence in directed acyclic graphs (DAGs) can be done in linear time. Based on that, they created a local DAG for every node of the network and restricted the influence of the node in this local area. Gap-filling seed selection was used to update the nodes' incremental influence spread after the DAGs were built, along with an accelerated solution.

The authors in (Barbieri and Bonchi 2014) modeled the viral marketing process of product adoption based on social influence and the feature of the products. They proposed feature-aware propagation model F-TM, and they defined the influence maximization with viral product design (MAXINF-VPD) problem and the study of its properties under F-TM propagation model, using two real-world semantically rich datasets from the domain of social music consumption (Last.fm) and social movie consumption (Flixster).

Squillero and Burelli (2016) explored the problem of influence maximization using a genetic algorithm (GA), which makes use of simple genetic operators commonly found in discrete optimization. They evaluated the genetic algorithm on two large, real-world network datasets from the Stanford Network Analysis Platform (SNAP) repository.

The authors in (Wang and Street 2018) proposed a model in which they quantified influence and tracked its diffusion and aggregation. (MAT) multiple-path asynchronous threshold, for viral marketing on social network. The MAT model captures not only direct influence but also indirect influence passed along messengers and they developed an efficient heuristic IV-greedy to tackle the influence maximization problem. The experiments of the MAT and the IV-greedy were conducted on four real-life networks, and they illustrated an important performance in terms of influence spread and time efficiency.

Saxena and Saxena (2020) proposed an influence maximization model by combining a node connection and its actual past activity pattern. Firstly, they proposed a diffusion model, namely, HAC-Rank algorithm, for the selection of initial adopters. Furthermore, they proposed a new Hurst-based influence maximization for studying the influence spread of seed nodes, wherein the activation of a node depends upon its connections and the self-similarity trend shown by its past activity. The performance of the HAC-Rank has been evaluated under IC, and the HBIM diffusion model achieved an average influence spread of 20.3%. Under the proposed HBIM model, HAC-Rank achieved 49.8% average influence spread in comparison to other state-of-the-art algorithms.

The comparison of maximization influence models is listed in Table 2.

Table 2 The comparison of performance for related influence maximization models

References	Contribution	Diffusion model	Graph	Dataset
Domingos and Richardson (2001)	First algorithm for influence maximization	/	/	Knowledge-sharing site
Kempe et al. (2003)	Approximation algorithm	LTM ICM	Directed/undirected	Large collaboration network
Leskovec et al. (2007)	Efficient approximation algorithm CELE (Cost-Effective Lazy Forward)	ICM	Directed	Water sensor networks (BWSN) Blog data
Chen et al. (2009)	New degree discount heuristic	LTM /ICM	/	Collaboration graphs obtained from the online archival database arXiv.org
Chen et al. (2010b)	First scalable influence maximization problem	LTM	Directed	NetHEPT, DBLP, Epinions, Amazon
Chen et al. (2010a)	New heuristic algorithm that is easily scalable for millions of nodes	ICM	Directed/undirected	NetHEPT, DBLP, Epinions, Amazon
Goyal et al. (2011)	CELF++ algorithm (optimization of CELF algorithm)	ICM	Directed	NetHEPT, Net PYH

Barbieri and Bonchi (2014)	Novel feature-aware propagation model	LTM	Directed	Last.fm, Flixster
Squillero and Burelli (2016)	Influence maximization in social networks with genetic algorithms	ICM	Directed	SNAP repository
Wang and Street (2018)	Developed an effective and efficient heuristic tackle the influence maximization problem	MAT (multiple-path asynchronous threshold)	Directed/undirected	PGP, NetHEPT, WikiVote, C. *elegans*
McKay et al. (2019)	Newer solution for influence maximization using machine learning	Neural network	Neural network	DBLP
Saxena and Saxena (2020)	HAC-rank algorithm	HBIM (Hurst-based influence maximization)	Directed	UC Irvine messages, MathOverflow, Linux Kernel mailing list
Ko et al. (2020)	Inductive machine learning method called MONSTOR (Monte Carlo simulation)	ICM	Directed	Extended, WannaCry, Celebrity

4   Deep Learning Approach for Influence Analysis

Recent work in social network analysis and influence analysis has been applied using the deep learning (Najafabadi et al. 2015; Hayat et al. 2019; Gao et al. 2020; Wang et al. 2019; Keikha et al. 2020; Wu et al. 2019; Zhang et al. 2020). A fundamental step for social network analysis using the deep learning is to encode network data into a low-dimensional representation (Tan et al. 2019).

The work presented in (Luceri et al. 2019) studied the impact of social influence on offline dynamics to study human real-life behavior. They used the deep learning technique for modeling social influence and predicting human behavior on real-world activities. They proposed a social influence deep learning framework that combines deep learning with network science for modeling and forecasting social influence on real-life activities, their social influence deep learning (SIDL) framework based on DNNs.

Authors in (McKay et al. 2019) proposed a newer solution for the influence maximization problem using machine learning. Their objective was to create a deep learning model to solve the influence maximization problem for viral marketing in a faster, more efficient, and more current leading algorithm; they are comparing the result of their model named learning algorithm with three main algorithms: the random selector, sum of edge, and greedy algorithm. For their study they built an artificial neural network in order to test their model against the order to other algorithm following they used the real network (DBLP a computer science bibliography website) for obtained the real results on real data. For the comparison of their model and the main other model (algorithm) existed, they compare with two measures the time efficiency and influence spread in the network. Their model reduce the time running and maximize the number of nodes activated. Their model activates 25.46% of the network's threshold, the greedy, sum of edge and the best random algorithm activates more than 4% (2.98%, 2.87%, and 3.78%), these results of smaller network. The result of their model in the large scale network is 48.18%, sum of edge 26.42% and 32.58%. The result experiment proved that their model is more efficient in the total amount of influence spread and in time efficiency.

In (Tian et al. 2020) motivated by the application of viral marketing, first they proposed two topic-aware social influence propagation models based on IC and LT models. Second, they proposed a new graph-embedding network, called Diffusion2Vec, which can extract features for each user in social network automatically. It's also important to note that they came up with a method for calculating the influence of a candidate user based on their embeddings. Finally, they adopted an algorithm of reinforcement learning, called double DQN with prioritized experience replay to train models. For the experiments, they used real-world social network from Twitter.

5 Conclusion

Online social networks are popular services that have been studied heavily in recent years, as more and more people communicate with friends, colleagues, and family through different existing social networks. We found several interesting papers surveying aspects of social networks. In this chapter, we reviewed the major aspects of OSNs, namely, online social network, social network analysis, social influence analysis, and viral marketing, and we defined the problem space in the social influence analysis. We also surveyed the main existing models and algorithms for influence modeling, influence maximization, and influential spreaders for viral marketing. The main objective of this chapter was to summarize the most important algorithms and models of influence analysis for viral marketing for beginners in this area. As we have learned in this chapter, there are many new problems and challenges on social influence analysis in our future work; we aim at proposing a new model of influence spreading or a new method for the identification of influential nodes in online social network for viral marketing. We hope this chapter will be very useful in clarifying this exciting area of research and serve as a solid foundation for readers interested in this field.

References

Abd Al-Azim, N.A.R., Gharib, T.F., Afify, Y., Hamdy, M.: Influence propagation: interest groups and node ranking models. Phys. A: Statist. Mech. Appl. **124247** (2020)

Ahajjam, S., Badir, H.: Identification of influential spreaders in complex networks using hybrid rank algorithm. Sci. Rep. **8**(1), 1–10 (2018)

AlSuwaidan, L., Ykhlef, M.: Toward information diffusion model for viral marketing in business. Int. J. Adv. Comput. Sci. Appl. **7**(2), 637–646 (2016)

Al-Taie, M.Z., Kadry, S., Obasa, A.I.: Understanding expert finding systems: domains and techniques. Soc. Netw. Anal. Min. **8**(1), 57 (2018)

Arnaboldi, V., Conti, M., La Gala, M., Passarella, A., Pezzoni, F.: Information diffusion in OSNs: the impact of nodes' sociality. In: Proceedings of the 29th Annual ACM Symposium on Applied Computing, pp. 616–621 (2014, March)

Bae, J., Kim, S.: Identifying and ranking influential spreaders in complex networks by neighborhood coreness. Phys. A: Statist. Mech. Appl. **395**, 549–559 (2014)

Barbieri, N., Bonchi, F.: Influence maximization with viral product design. In: Proceedings of the 2014 SIAM International Conference on Data Mining, pp. 55–63. Society for Industrial and Applied Mathematics (2014, April)

Basaras, P., Katsaros, D., Tassiulas, L.: Detecting influential spreaders in complex, dynamic networks. Computer. **4**, 24–29 (2013)

Belfin, R.V., Bródka, P.: Overlapping community detection using superior seed set selection in social networks. Comput. Electr. Eng. **70**, 1074–1083 (2018)

Bhat, N., Aggarwal, N., Kumar, S.: Identification of influential spreaders in social networks using improved hybrid rank method. Procedia Computer Science. **171**, 662–671 (2020)

Bhattacharya, S., Gaurav, K., Ghosh, S.: Viral marketing on social networks: an epidemiological perspective. Phys. A: Statist. Mech. Appl. **525**, 478–490 (2019)

Boyd, D.M., Ellison, N.B.: Social network sites: definition, history, and scholarship. J. Comput.-Mediat. Commun. **13**(1), 210–230 (2007)

Cai, Q., Ma, L., Gong, M., Tian, D.: A survey on network community detection based on evolutionary computation. Int. J. Bio-Inspired Comput. **8**(2), 84–98 (2016)

Can, U., Alatas, B.: A new direction in social network analysis: online social network analysis problems and applications. Phys. A: Stat. Mech. Appl. **535**, 122372 (2019)

Cano, C.: The SIR Models, their applications, and Approximations of their Rates (2020)

Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 199–208 (2009, June)

Chen, W., Wang, C., Wang, Y.: Scalable influence maximization for prevalent viral marketing in large-scale social networks. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1029–1038 (2010a, July)

Chen, W., Yuan, Y., Zhang, L.: Scalable influence maximization in social networks under the linear threshold model. In: 2010 IEEE International Conference on Data Mining, pp. 88–97. IEEE (2010b, December)

Cheng, Y., Liu, J., Yu, X.: Online social trust reinforced personalized recommendation. Pers. Ubiquit. Comput. **20**(3), 457–467 (2016)

Dabaghi-Zarandi, F., Rafsanjani, M.K.: Community detection in social networks. In: Models and Theories in Social Systems, pp. 273–293. Springer, Cham (2019)

Daud, N.N., Ab Hamid, S.H., Saadoon, M., Sahran, F., Anuar, N.B.: Applications of link prediction in social networks: a review. J. Netw. Comput. Appl. **102716** (2020)

Dhamal, S., Prabuchandran, K.J., Narahari, Y.: Information diffusion in social networks in two phases. IEEE Trans. Netw. Sci. Eng. **3**(4), 197–210 (2016)

Domingos, P., Richardson, M.: Mining the network value of customers. In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 57–66 (2001, August)

Du, D.Z., Pardalos, P.M., Zhang, Z.: Nonlinear Combinatorial Optimization, vol. 147. Springer (2019)

Fu, X., Passarella, A., Quercia, D., Sala, A., Strufe, T.: Online social, networks. Comput. Commun. **73**, 163–166 (2016)

Gaeta, R.: A model of information diffusion in interconnected online social networks. ACM Transactions on the Web (TWEB). **12**(2), 1–21 (2018)

Gao, L., Zhou, B., Jia, Y., Tu, H., Wang, Y., Chen, C., Zhuang, H.: Deep learning for social network information Cascade analysis: a survey. In: 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC), pp. 89–97. IEEE (2020, July)

Girdhar, N., Minz, S., Bharadwaj, K.K.: Link prediction in signed social networks based on fuzzy computational model of trust and distrust. Soft. Comput. **23**(22), 12123–12138 (2019)

Goyal, A., Lu, W., Lakshmanan, L.V.: Celf++ optimizing the greedy algorithm for influence maximization in social networks. In: Proceedings of the 20th International Conference Companion on World Wide Web, pp. 47–48 (2011, March)

Guille, A., Hacid, H., Favre, C., Zighed, D.A.: Information diffusion in online social networks: a survey. ACM SIGMOD Rec. **42**(2), 17–28 (2013)

Gulati, A., Eirinaki, M.: Influence propagation for social graph-based recommendations. In: 2018 IEEE International Conference on Big Data (Big Data), pp. 2180–2189. IEEE (2018, December)

Hayat, M.K., Daud, A., Alshdadi, A.A., Banjar, A., Abbasi, R.A., Bao, Y., Dawood, H.: Towards deep learning prospects: insights for social media analytics. IEEE Access. **7**, 36958–36979 (2019)

Keikha, M.M., Rahgozar, M., Asadpour, M., Abdollahi, M.F.: Influence maximization across heterogeneous interconnected networks based on deep learning. Expert Syst. Appl. **140**, 112905 (2020)

Kemi, A.O.: Impact of social network on society: a case study of Abuja. Am. Sci. Res. J. Eng. Technol. Sci. (ASRJETS). **21**(1), 1–17 (2016)

Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137–146 (2003, August)

Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H.E., Makse, H.A.: Identification of influential spreaders in complex networks. Nat. Phys. **6**(11), 888–893 (2010)

Ko, J., Lee, K., Shin, K., Park, N.: MONSTOR: an inductive approach for estimating and maximizing influence over unseen social networks. arXiv preprint arXiv:2001.08853. (2020)

Kong, X., Gu, Z., Yin, L.: A unified information diffusion model for social networks. In: 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC), pp. 38–44. IEEE (2020, July)

Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N.: Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 420–429 (2007, August)

Li, D., Liu, J.: Modeling influence diffusion over signed social networks. IEEE Trans. Knowl. Data Eng. (2019)

Liu, Q., Zhu, Y.X., Jia, Y., Deng, L., Zhou, B., Zhu, J.X., Zou, P.: Leveraging local h-index to identify and rank influential spreaders in networks. Phys. A: Statist. Mech. Appl. **512**, 379–391 (2018)

Loucif, H.: *The Analysis of Social Influence in Social Media Networks* (Doctoral dissertation, Université de Bordj Bou Arréridj-Mohamed El Bachir El Ibrahimi) (2016)

Luceri, L., Braun, T., Giordano, S.: Analyzing and inferring human real-life behavior through online social networks with social influence deep learning. Appl. Netw. Sci. **4**(1), 34 (2019)

Ma, L.L., Ma, C., Zhang, H.F., Wang, B.H.: Identifying influential spreaders in complex networks based on gravity formula. Phys. A: Statist. Mech. Appl. **451**, 205–212 (2016)

McKay, D. B., Corse, J. A., & Gonsalves, M. S.: Deep Learning Method for Social Networks (2019)

Menta, V.P.T., Singh, P.K.: Efficient selection of influential nodes for viral marketing in social networks. In: 2017 IEEE, International Conference on Current Trends in Advanced Computing (ICCTAC), pp. 1–6. IEEE (2017, March)

More, J.S., Lingam, C.: A gradient-based methodology for optimizing time for influence diffusion in social networks. Soc. Netw. Anal. Min. **9**(1), 5 (2019)

Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R., Muharemagic, E.: Deep learning applications and challenges in big data analytics. J. Big Data. **2**(1), 1 (2015)

Okamoto, K., Chen, W., Li, X.Y.: Ranking of closeness centrality for large-scale social networks. In: International Workshop on Frontiers in Algorithmics, pp. 186–195. Springer, Berlin, Heidelberg (2008, June)

Ortiz-Gaona, R.M., Postigo-Boix, M., Melús-Moreno, J.L.: Extent prediction of the information and influence propagation in online social networks. Comput. Math. Organ. Theory. (2020)

Otte, E., Rousseau, R.: Social network analysis: a powerful strategy, also for the information sciences. J. Inf. Sci. **28**(6), 441–453 (2002)

Pan, T., Li, X., Kuhnle, A., Thai, M.T.: Influence diffusion in online social networks with propagation rate changes. IEEE Trans. Netw. Sci. Eng. (2020)

Peng, S., Wang, G., Xie, D.: Social influence analysis in social networking big data: opportunities and challenges. IEEE Netw. **31**(1), 11–17 (2016)

Pourhojjati-Sabet, M., Rabiee, A.: A soft recommender system for social networks. arXiv preprint arXiv:2001.02520. (2020)

Qiu, J., Tang, J., Ma, H., Dong, Y., Wang, K., & Tang, J.: Deepinf: social influence prediction with deep learning. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2110–2119) (2018, July)

Qiyao, W., Zhengmin, L., Yuehui, J., Shiduan, C., Tan, Y.: Ulm: a user-level model for emotion prediction in social networks. China Univ. Posts Telecommun. (2016)

Richardson, M., Domingos, P.: Mining knowledge-sharing sites for viral marketing. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 61–70 (2002, July)

Saxena, B., Saxena, V.: Influence maximization in social networks using Hurst exponent-based diffusion model. In: 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 167–171. IEEE (2020, January)

Singh, N., Malik, A., Maini, O., Rajput, G.: Identification of influence propagation metrics in social networks. In: 2019 International Conference on Automation, Computational and Technology Management (ICACTM), pp. 224–227. IEEE (2019, April)

Squillero, G., Burelli, P.: Applications of Evolutionary Computation: 19th European Conference, Evo Applications 2016, Porto, Portugal, March 30–April 1, 2016, Proceedings, Part I, vol. 9597. Springer (2016)

Sun, J., Tang, J.: A survey of models and algorithms for social influence analysis. In: Social Network Data Analytics, pp. 177–214. Springer, Boston, MA (2011)

Sun, Q., Li, Y., Hu, H., Cheng, S.: A model for competing information diffusion in social networks. IEEE Access. **7**, 67916–67922 (2019)

Talukder, A., Layek, M. A., & Hong, C. S.: A Novel Approach of Viral Marketing in Social Networks, 1265–1267 (2017)

Tan, Q., Liu, N., Hu, X.: Deep representation learning for social network analysis. Front. Big Data. **2**, 2 (2019)

Tian, S., Mo, S., Wang, L., Peng, Z.: Deep reinforcement learning-based approach to tackle topic-aware influence maximization. Data Sci. Eng. **111**, 1–11 (2020)

Toalombo, M., Wang, B., Xu, H., Xu, M.: A novel greedy fluid spread algorithm with equilibrium temperature for influence diffusion in social networks. IEEE Syst. J. (2020)

Towhidi, G., Sinha, A.P., Srite, M., Zhao, H.: Trust decision-making in online social communities: a network-based model. J. Comput. Inf. Syst., 1–11 (2020)

Wang, W., Street, W.N.: Modeling and maximizing influence diffusion in social networks for viral marketing. Appl. Netw. Sci. **3**(1), 6 (2018)

Wang, F., She, J., Ohyama, Y., Wu, M.: Deep-learning-based identification of influential spreaders in online social networks. In: IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society, vol. 1, pp. 6854–6858. IEEE (2019, October)

Wu, J., Sha, Y., Jiang, B., Tan, J.: DSINE: deep structural influence learning via network embedding. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 10065–10066 (2019, July)

Xu, W., Wu, W., Fan, L., Lu, Z., Du, D.Z.: Influence diffusion in social networks. In: Optimization in Science and Engineering, pp. 567–581. Springer, New York, NY (2014)

Yuan, S., Zhang, Y., Tang, J., Hall, W., Cabotà, J.B.: Expert finding in community question answering: a review. Artif. Intell. Rev. **53**(2), 843–874 (2020)

Zafarani, R., Abbasi, M.A., Liu, H.: Social Media Mining: An Introduction. Cambridge University Press (2014)

Zeng, A., Zhang, C.J.: Ranking spreaders by decomposing complex networks. Phys. Lett. A. **377**(14), 1031–1035 (2013)

Zhang, Y., Li, S., Yu, Z., Zhang, F., Lu, H.: A 2020 perspective on "predicting the influence of viral messages for VM campaigns on Weibo". Electron. Commer. Res. Appl. **40**, 100949 (2020)