IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 7, 2017, accepted October 10, 2017, date of publication October 17, 2017, date of current version December 5, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2764001

Information Diffusion Predictive Model Using Radiation Transfer

LULWAH ALSUWAIDAN [1,2], (Member, IEEE), AND MOURAD YKHLEF [1]
[1]Department of Information Systems, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia
[2]Department of Information Management, College of Computer and Information Sciences, Al-Imam Mohammad ibn Saud Islamic University, Riyadh 11432, Saudi Arabia

Corresponding author: Lulwah AlSuwaidan (lnsuwaidan@imamu.edu.sa)

This work was supported by the Deanship of Scientific Research and Research Center of College of Computer and Information Sciences, King Saud University.

ABSTRACT The impact of online social networks on information exchange between humans has revealed the need to study the mechanisms of information diffusion. Multiple prior works have considered empirical studies and introduced new diffusion models to understand the dynamics of the diffusion process. However, the complexity of network structures and user interactions make it challenging to model the diffusion mechanisms of online social networks and to accurately predict diffusion. In this paper, we propose an information diffusion prediction model based on a physical radiation energy transfer mechanism. The aim of this model is to predict the diffusion graph of a certain contagion throughout an interest-based community. This non-parametric model can accommodate the dynamicity of online social networks because it can receive different input diffusion parameters at different diffusion contagions. With our RADiation DIFFusion (RADDIFF) model, we precisely capture the information diffusion process from both temporal and spatial dimensions and measure the level of influence initiated by certain influencers for each diffusion process. To our knowledge, this model is the first in this domain that exploits the prediction of information networks based on a physical radiation mechanism. We conduct an extensive analysis using an experiment that includes two well-known prediction diffusion models, the linear influence model (LIM) and NETINF. The results show that RADDIFF effectively outperforms both the LIM and NETINF in terms of accuracy and the quality of forecast.

INDEX TERMS Radiation, information diffusion, online social networks, RADDIFF.

I. INTRODUCTION

Online social networks have increased in popularity since 2010 and have become an essential type of media in spreading and exchanging information. This exchange of information through online social networks has affected the communication between entities, particularly in real life. Online social networks place no limit on the content that can be shared between users. Photos, videos, text, and links are some of the main content exchanged. This capability makes online social networks a key alternative to traditional sharing media channels. In this context, integrating computer science and mathematical modeling opens new possibilities to quantitatively and systematically understand the information diffusion process of online social networks.

The majority of prior information diffusion models on online social networks have concentrated on understanding the network structure and user interaction [1], [2]. Another direction has considered empirical prospectives for information diffusion [3]. This type of modeling analyzes the properties of diffusion and how they are manifested between network entities. It is only recently that researchers have begun to systematically examine the effects of applying mathematical models to predict information diffusion over time [4]–[7]. In previous studies, diffusion has been formulated according to physical or natural phenomena to reflect how diffusion in real life can be transformed into computer-based modeling.

This article proposes a novel radiative diffusion model to mathematically formulate the information diffusion of online social networks. The proposed model correlates the properties between both temporal and spatial patterns of information diffusion and radiation energy transfer in physical space and time. The purpose of the radiative diffusion model is to predict and measure the influence density of the diffusion at time t and distance x of the source s. RADiation DIFFusion (RADDIFF) modeling is non-parametric and can thus accommodate the dynamicity of online social networks by adapting different input parameters depending on the contagion. The problem of studying the prediction problem is crucial for various applications, such as viral marketing and opinion spreading [8]–[10].

The rest of the article is organized as follows: Section 2 presents a brief literature review on related works. Section 3 introduces the flow of the proposed diffusion model and its mathematical formulation. Section 4 contains validation measures to quantify the models effectiveness, which also includes the model evaluation on a real dataset collected from Twitter. The model is compared to other well-known diffusion models, the linear influence model (LIM) and NETINF, which were produced by Stanford University. Finally, we conclude the article and outline future works.

II. RELATED WORKS

Although the information diffusion problem has been an active research topic for many years, it has proven to be a challenging task relating to online social networks. Specifically, when diffusion models incorporate spatial, topical, and temporal aspects, the problem becomes more complex. The early adaption of information diffusion has stimulated the spread of ideas and influence as epidemiological models through the consideration of active and inactive nodes, where active nodes can spread contagions such as information, diseases, or influence throughout the network.

The information diffusion models proposed to date can be categorized into predictable and influence maximization models. Predicting the diffusion of information in social networks has been a research problem for many applications, including tribe leaders detection, social influence and viral marketing. In recent years, the majority of diffusion models have been based primarily on Independent Cascade (IC) [3], [11]–[16] or Linear Threshold models [17]–[22], which are originally introduced in epidemiology and social studies [23]. Several attempts have been devoted to predict information diffusion of temporal and structural patterns [24], [25]. Inferring links of diffusion was performed early by Adamic *et al.* [26] and Bakshy *et al.* [27], who formulated the links of diffusion using classification and machine learning. Tracking diffusion paths and inferring contagion network propagation have also been studied by Gomez-Rodriguez *et al.* [28] who proposed the NETINF algorithm. Other studies have concentrated on solving the inference problem of networks, as with CONNIES [29] and NETRATE [30], which are inference algorithms that address network sparsity. The philosophy of diffusion was taken in another direction by Yang and Leskovec [31], who considered the global influence of diffusion rather than the direct node influence. The T-BaSIC model presented by Guille and Hacid [19] can predict the temporal dynamics of diffusion in social networks. This approach is based on machine learning techniques and the inference of time-dependent diffusion probabilities from a multidimensional analysis of individual behaviors. A recent diffusion prospective has focused on stimulating physical phenomena, such as water motion or heat transfer, as a diffusion model, which has been introduced by different researchers. Diffusion logistic models based on heat diffusion and Flacks law were applied by Wang *et al.* [4], [5]. Dynamic controls have also been considered as a solution for information diffusion in online social networks. Hu *et al.* [6], [7] introduced a hydro-IDP information diffusion model that models the evolution of a fluid based on a set of partial differential equations.

In contrast, influence maximization models incorporate link analysis and network structure analysis. Data mining, statistical modeling, and empirical approaches have been used as a measurement and analysis techniques in information modeling [32]. Researchers in the field have begun to examine the effects of choosing optimal initial influential nodes [33]. Chaudhury *et al.* [34] proposed the degree-based scaling method, which aims to increase the active set of nodes that reaches an optimal point in the least amount of time. Then, they used the maximum spanning tree to find the k-influence nodes. Both algorithms ensured optimal seeding regardless of the amount of time consumed. Abadi and Khayyambashi [35] discussed the problem of influence maximization with viral marketing and introduced a new algorithm aimed at selecting the expert and leader in social networks based on spatiality and knowledge. A study by Zhou *et al.* [36] addressed the influence maximization problem by integrating greedy algorithms and mining the top influences. These authors proposed GAUP to mine the most influential nodes in the network. Wang *et al.* [37] proposed an independent cascade-based model for influence maximization (IMIC-OC) to determine positive influence.

All of the aforementioned studies have focused on maximizing influence by concentrating on the social network structure. The dynamics of online social networks has been an active research topic addressed by multiple studies [38]. Zhuang *et al.* [39] concentrated on maximizing the influence in dynamic social networks and proposed an influence maximization algorithm called Maximum Gap Probing (MaxG). Kim *et al.* [40] introduced a decentralized influential maximization problem by influencing k-neighbors rather than randomly selected users in the network. They showed that users with higher propagation rate neighbors were more suitable for spreading than those with a high number of neighbors. The chosen methods for selecting the best neighbors were Random, Degree, Propagation-weight and Hybrid selection. The Hybrid method was considered to be the best selection method, as it provided the most influence maximization regardless of the number of k-nodes. Wu *et al.* [41] introduced an independent cascade model with accepted probability (ICMAP) to describe cooperative influence spreading in a social network. This model employs an improved greedy algorithm to maximize the approximation of the cooperative influence spread. Models of natural phenomena, such as heat diffusion, were adopted in influence maximization problems by Yang *et al.* [42]. They introduced a targets heat

TABLE 1. Summary of diffusion models categorized into prediction and influence maximization models.

	Approach	Pattern		Math Modeling		Model Type	
		Temporal	Structural	Parametric	Non-parametric	Physical	Non-physical
Predictive	LT [14-19]		Yes	Yes			Yes
	IC [3, 8-13]		Yes	Yes			Yes
	LIM [28]	Yes			Yes		Yes
	CONNIES [26]		Yes		Yes		Yes
	NETRATE [27]	Yes			Yes		Yes
	T-BaSIC [16]	Yes		Yes			Yes
	NETINF [25]		Yes	Yes		Yes	
	PDE [4, 5]	Yes	Yes	Yes		Yes	
	Hydro-IDP [6, 7]	Yes	Yes		Yes	Yes	
Influence Max.	degree-based scaling method [31]		Yes		Yes		Yes
	GAUP [33]		Yes	Yes			Yes
	IMIC-OC [34]		Yes	Yes			Yes
	MaxG [36]	Yes			Yes		Yes
	targets heat greedy algorithm [39]	Yes			Yes	Yes	
	ICMAP [38]	Yes		Yes			Yes

greedy algorithm based on an analysis of the laws of heat propagation.

In Table 1, we show a summary that compares popular diffusion models. The summary is categorized into two categories: predictive models and influence maximization models. The comparison is based on reviewing the diffusion model in the literature with respect to the model pattern, mathematical modeling mode, and model type, either physical or non-physical. In terms of the model type, it is clear that the majority of the existing models are non-physical and are not based on any physical phenomenon. Most of the recent models consider both temporal and structural patterns when introducing a new predictive diffusion model, while influence maximization models focus on temporal patterns only, and the earlier models consider network structure. Parameter setting modes and mathematical modeling, on the other hand, affect the overall performance of the model. Setting the model as a non-parametric mode keeps the model more effective and dynamic. In summary, there have been limited attempts to simulate a physical phenomenon using a diffusion model. In addition, it is more effective to consider non-parametric mathematical modeling for diffusion. Considering both temporal and structural patterns when introducing a new diffusion model ensures the models robustness and accuracy.

III. INFORMATION DIFFUSION PREDICTION

We have presented different issues related to information diffusion. In this section, we address the aforementioned issues by introducing a novel radiative information diffusion model directed at online social networks. The model concentrates on predicting the diffusion graph for certain information initiated from source s and measuring the influence density over time. This section will cover two parts. First, we present a description of the RADDIFF model that contains fundamental parts of the proposed model. Second, we present a detailed mathematical representation of the RADDIFF model that correlates the physical radiation phenomena into an online social network information diffusion model.

A. RADDIFF MODEL

The RADDIFF model (Figure 1) describes the flow of diffusion throughout its components. The first and foremost component is the interest-based community detection, where the community is selected based on chosen interest keywords, such as a business social networks community. Then, the seed extraction process on the selected community is implemented to determine the most influential nodes that initiate the diffusion. Identifying influence spreaders has been an active research topic for many years. Since the performance and simplicity in identifying influential spreaders are critical objectives of the RADDIFF model, we have reviewed the most dominant influence spreader identification algorithms in the literature [43]–[45]. We have taken into consideration many aspects when choosing an influencer identification algorithm. First, the algorithm should have the ability to handle weighted graphs since the communities are weighted. Second, the algorithm must consider reciprocal links between entities since social links may be unidirectional and bidirectional. Third, node ranking is another important aspect that should be considered in the influence spreader algorithm. PageRank [46], [47] is one of the most promising influence spreader algorithm candidates because its effectiveness has been proven on weighted and directed graphs. In addition, this model guarantees the aforementioned conditions that should exist in the chosen algorithm. PageRank has an endorsement

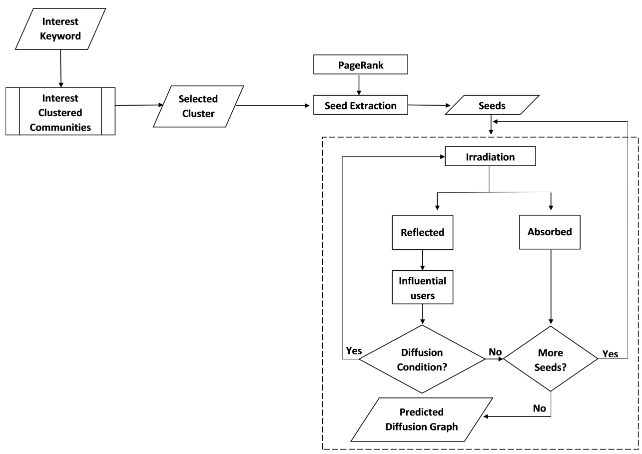


FIGURE 1. The RADiation DIFFusion (RADDIFF) Model.

property in which a page is important if it is cited by other important pages. This property is also essential in terms of online social networks if the node has more in-links that reflect its importance among other nodes in the community.

After identifying the seeds by PageRank, the diffusion component starts, a detailed description of this component will be provided in the next subsection. Once the seed initiates the information diffusion message, this message is considered an irradiation to other nodes within the community. There are two possibilities for the node receiving the irradiation: either reflect or absorb the message. If the message is absorbed, then the diffusion will terminate, and the model will check for more diffusion seeds. Conversely, if the message is reflected, then the message will be transferred to other nodes in the community, and the nodes will become influenced. Then, there is a diffusion component that checks for other seeds that can diffuse the message between its connections. If there are no more seeds, then the predicted diffusion graph is presented.

B. MATHEMATICAL MODELING OF RADDIFF MODEL

In this section, we formally introduce the mathematical form of the RADDIFF model. The discussion of this model is restricted to the setting of online social networks in which we treat the nodes of online social networks as sources/receivers of radiation and links between nodes as their interaction. The modeling focuses on simulating the information spreading as a radiation transfer problem. In this section, we physically define the radiation transfer. Then, we define the information diffusion representation of the RADDIFF model. Next, we illustrate the RADDIFF model by example.

1) RADIATIVE TRANSFER DEFINITION

Radiative transfer is the propagation of electromagnetic waves between objects surfaces. Generally, the transfer is a transmission process between surfaces as illustrated in Figure 2.

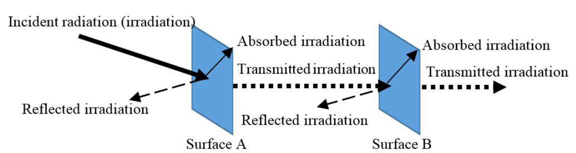


FIGURE 2. Radiation transfer process between different surfaces.

For simplicity, we consider two surfaces, A and B. The radiation impinges on surface A when it is received from a source such as the Sun. The irradiation on surface A can be divided into three cases: absorbed, reflected, and transmitted. The transmitted radiation is transferred from surface A to surface B. Part of the irradiation is reflected by the surface and might be absorbed by the surrounding surfaces. Radiation emitted by the surface originates from the thermal energy of matter bounded by the surface, and the rate at which energy is released per unit area (W/m^2) is termed the surface emissive power E [15]. The upper limit to the emissive power is prescribed by the Stefan-Boltzmann law as follows:

$$E_b = \sigma T_s^4 \tag{1}$$

where T_s is the absolute temperature (K) of the surface and the Stefan-Boltzmann constant $\sigma = 5.67 \times 10^8 W/w^2 \cdot K^4$. Such a surface is called an ideal radiator or blackbody.

The heat flux emitted by a real surface is less than that of a blackbody at the same temperature [53]. To compute this flux, we must consider the radiative property of the surface, termed the emissivity ε. Its values range within $0 \leq \varepsilon \leq 1$, and this property provides a measure of how efficiently a surface emits energy relative to a blackbody. Then, the heat flux is given by:

$$E = T_s^4 \tag{2}$$

Real surfaces have emissive powers, E, which are somewhat less than that obtained theoretically by Boltzman. Here, the emissivity, ε is:

$$\varepsilon = E/E_b \tag{3}$$

The radiation may originate from a special source, such as the sun, or from other surfaces to which the surface of interest is exposed. Irrespective of the source, s, the rate at which all such radiation is incident on a unit area of the surface characterizes the irradiation G.

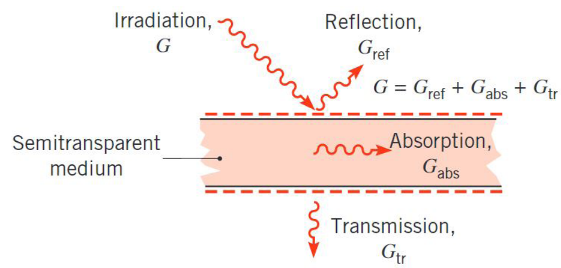


FIGURE 3. Radiation incedent on semetransparent.

When the radiation is incident upon a semitransparent medium, portions of the irradiation may be reflected, absorbed, and transmitted, as illustrated in Figure 3. In the case of reflected irradiation, the reflectivity ρ is the fraction of the irradiation that is reflected. In the case of absorption, the absorptivity α is the fraction of the irradiation that is absorbed, and the transmissivity τ is the fraction of the irradiation that is transmitted. Because all of the irradiation must be reflected, absorbed, or transmitted, it follows that

$$\rho + \alpha + \tau = 1 \tag{4}$$

A medium that experiences no transmission $\tau = 0$ is opaque, as shown in Figure 4, in which case:

$$\rho + \alpha = 1 \tag{5}$$

With this understanding of the partitioning of the irradiation into reflected, absorbed, and transmitted components, two additional and useful radiation fluxes can be defined. The radiosity, $J(W/m^2)$, of a surface accounts for all radiant

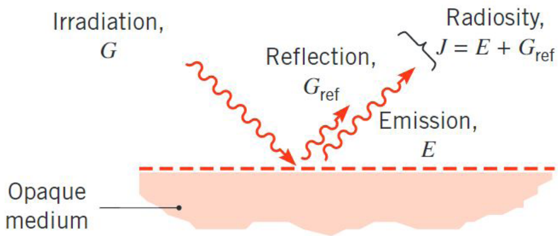


FIGURE 4. **Radiation incedent on opaque.**

energy leaving the surface. For an opaque surface, this parameter includes emission and the reflected portion of the irradiation, as illustrated in Figure 4, and is therefore expressed as:

$$J = E + G_{ref} = E + \rho G \tag{6}$$

where G_{ref} is the power reflected by the surface that received the irradiation.

One more important component is the view factor, which represents the geometrical features of the radiation exchange problem. The view factor is defined as follows:

Definition 1: The view factor F_{ij} is the fraction of the radiation leaving surface i that is intercepted by surface j.

$$A_i F_{ij} = A_j F_{ji} \tag{7}$$

This expression, termed the reciprocity relation, is useful in determining one view factor from knowledge of the other. The net rate exchange between surfaces is as follows:

$$q = J \cdot A \cdot F \tag{8}$$

where J is defined in Equation 6, A represents the surface area, F is the view factor.

2) INFORMATION DIFFUSION FORMULATION OF RADDIFF MODEL

The information diffusion on online social network shares similar properties with radiation transfer. The radiation propagates the electromagnetic waves, where diffusion in online social networks propagates a contagion between nodes. To formulate the problem, we begin with Equation 6. We assume G_{ref} for the source node is zero because no emissions are received from other nodes. Then,

$$J = E \tag{9}$$

Otherwise, if the node is not the source, then irradiation is received as G_{ref} that consists of the original irradiation G. Thus, the received information from the predecessor, along with the surface reflectivity constant ρ, reflects the ability or the curiosity to retransmit the received information. We assume that each surface has its emissive power E. We consider that each node has the power to transfer the information and reflect emissions. For example, let us consider that node X has received information from the predecessor node Y. We assume that E is the information that node X would add to the information received from node Y. However G_{ref} is the capacity to retransmit the same information to its connections. Then, the net emission power is shown in Equation 6.

To describe element E in Equation 9, we return to Equation 2. Elements T and ε shown in Equation 2 can be defined in the information diffusion problem as follows.

$$\varepsilon = \frac{no.\ posts\ by\ user\ per\ day}{Maximum\ post\ limit\ day} \tag{10}$$

where "maximum post limit per day" is the maximum posts allowable for each user by the social network platform. For example, Twitter allows 2400 tweets as a maximum limit of tweets per day.

T in the physical radiation model is the surface temperature, for which high temperatures mean that the surface is capable of diffusing at a higher rate than another surface with less temperature. Therefore, T in information diffusion reflects the number of links connected to the node, which represents the connections in the interaction graph. When the node has a higher number of connections, it means that the node has a higher importance among other nodes in the network and can diffuse information better than other nodes. Emissivity ε, in contrast, is the ratio of the number of posts that users' post per day to the number of posts that the blackbody posts. The purpose of this computation is to measure how active a user is in the network.

Similarly, the majority of online social networks platforms have considered the number of followers for each registered user. Therefore, we can map A, which is the surface area in the physical radiation model, to the number of node followers because it represents the coverage limit. In the physical representation of the radiation model, the surface area affects the quality of diffusion where a bigger area covers more objects in the space. In the context of information diffusion, when the number of followers is high, then the diffusion coverage is also high. Using definition 1, F is the fraction of the radiation propagating from a node to other nodes. We assume that the interaction graph is weighted, and these weights represent the interaction volume and that they reflect the distance between the source node and its connection. The view factor is given as Equation 11:

$$F_{(Source \to i)} = \frac{W_{(Source \to i)}}{\sum_{i=1}^{n} W_{(Source \to i)}} \tag{11}$$

where $W_{(Source \to i)}$ is the weight on edge from *Source* to i, n is the number of links connected to the *Source* node.

Since a node in a social network is connected to many nodes, we define $q_{i \to n}$ as $q_{source} = \sum_{i=1}^{n} J_{source} A_{source} F_{source \to i}$ where $\sum_{i=1}^{n} F_{source \to i} = 1$. In the case of information diffusion, we treat each vertex in the network as an object that has the ability to be influenced (absorptivity α) by the information received and then reflect (reflectivity ρ) it to another vertex. The definition of these factors are as follow (see Equations 12 and 13):

$$\rho = \frac{no.\ reposts}{total\ no.\ of\ posts} \tag{12}$$
$$\alpha = 1 - \rho \tag{13}$$

Certain assumptions are required to effectively compute the proposed RADDIFF model. The first assumption is that the previous equations hold if and only if the received node has a lower edge count than the sender. Second, based on the interaction graph, we assume that nodes will receive the message since it shares the same interest and is within the community range.

3) ILLUSTRATIVE EXAMPLE

In this section, we present an illustrative example that describes the RADDIFF model steps. The example considers a graph of connected nodes on the Twitter online social network. Let us suppose that the source or influencer takes the message and posts it in his tweet list. The radiosity of all the radiant energy leaving the surface is given by Equation 6 and G will be

$$J = E + G_{ref} = E + \rho G \tag{14}$$
$$G = \frac{q_{source \to i}}{A_i} \tag{15}$$

To compute the radiosity J, which represents the diffusion, we apply the Equation 9 where $J = E$. Since the source has no irradiation received from other nodes, then G_{ref} for the source node is zero. The RADDIFF model is applicable for different online social networks, but our concentration in this example is on Twitter. We refer to our definition of irradiation as $q_{source} = \sum_{i=1}^{n} J_{source} A_{source} F_{source \to i}$ where $\sum_{i=1}^{n} F_{source \to i} = 1$. The parameter definition based on Twitter social characteristics are as follows:

$$T = number\ of\ edges\ in\ interaction\ graph$$
$$A = number\ of\ node\ followers.$$
$$F_{source \to i} = \frac{weight\ on\ edge}{total\ weights\ for\ all\ connected\ edges}$$

Reflectivity ρ

$$\rho = \frac{no.\ retweets\ done\ by\ user}{no.\ tweets\ of\ user\ timeline}$$

Absorptivity α

$$\alpha = 1 - \rho$$

Emissivity ε

It is the number of tweets per day that nodes post relatively to the blackbody to measure the degree of activity of the user.

$$\varepsilon = \frac{no.\ tweets\ posted\ by\ user\ per\ day}{2400}$$

Blackbody

Every real-world problem requires a definition of the ideal material or blackbody. In the problem of diffusion in online social networks, the blackbody node has the higher number of tweets per day. Let us assume that the blackbody of Twitter has a limit of 2,400 tweets per day.

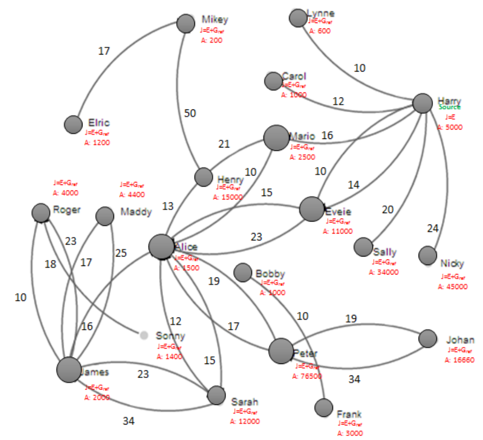


FIGURE 5. Graph representing connected twitter nodes with their followers and links weights.

Figure 5 illustrates a graph representing nodes in social network with their interaction connections. This Figure consists of 19 nodes connected, as shown in the graph, and the number associated with each link denotes the weight between two nodes. The weight indicates the interactions between nodes and is computed as described by Falkowski *et al.* [54].

To conduct the example, we first refer back to Equation 8 to compute the net exchange. Let us assume that Harry is the source node. He has 5,000 followers. Harry also has 7 edges representing his interactions with other nodes. Thus, Harry will have $A = 5000$ and $T = 7$. Let us assume that he can post 1,000 tweets per day, where the maximum tweets per day for each user is 2,400. Therefore $\varepsilon = \frac{1000}{2400} = 0.42$.

$$\begin{aligned} q_{Harry \to Mario} &= J_{Harry} A_{Harry} F_{Harry \to Mario} \\ &= 0.42 \times 5.67 \times 10^{-8} \times 7^4 \\ &\quad \times 5000 \times \frac{16}{96} = 0.047647845 \end{aligned}$$

Harrys other connections have no subsequent nodes. Thus, when Mario receives Harrys message, then

$$J = E + G_{ref} = E + \rho G$$
$$G_n = q_{sn}/A_n$$

Mario has $A = 2500$ followers, and the number of edges connected to him in the interaction graph is $T = 3$. Let assume Mario can post 600 tweets per day, thus $\varepsilon = \frac{600}{2400} = 0.25$. To compute the reflectivity ρ we must know the number of retweets that user performs relative to his total tweets posts. Assume that Mario has a total of 100 tweets and that 40 of them are retweets:

$$
\begin{aligned}
G_{ref} &= \rho G_{Mario} = \frac{40}{100} \times \frac{q_{Harry \to Mario}}{A_{Mario}} \\
&= 0.4 \times \frac{0.047647845}{2500} = 7.62366 \times 10^{-6} \\
E &= \varepsilon \sigma T_s^4 = 0.25 \times 5.76 \times 10^{-8} \times 3^4 = 1.14818 \times 10^{-6} \\
J &= E + G_{ref} \\
J &= 1.14818 \times 10^{-6} + 7.62366 \times 10^{-6} = 8.75329 \times 10^{-12} \\
q_{Mario \to Henery} &= J_{Mario} A_{Mario} F_{Mario \to Henery} \\
&= 8.75329 \times 10^{-12} \times 2500 \times \frac{21}{47} = 9.777611 \times 10^{-9}
\end{aligned}
$$

Henery receives the message to be diffused from Mario, then

$$
\begin{aligned}
J &= E + G_{ref} = E + \rho G \\
G_n &= q_{sn}/A_n
\end{aligned}
$$

Henery has $A = 15000$ and $T = 3$. Let assume Henery can post 900 tweets per day; thus $\varepsilon = \frac{900}{2400} = 0.375$. Henery has 970 total tweets, 300 as retweets. Then, G_{ref} is calculated as follow:

$$
\begin{aligned}
G_{ref} &= \rho G_{Hanery} = \frac{300}{970} \times \frac{q_{Mario \to Henery}}{A_{Henery}} \\
&= 0.31 \times \frac{9.777611 \times 10^{-9}}{15000} = 2.02071 \times 10^{-13} \\
E &= \varepsilon \sigma T_s^4 = 0.375 \times 5.76 \times 10^{-8} \times 3^4 = 1.722 \times 10^{-5} \\
J &= E + G_{ref} \\
J &= 1.14818 \times 10^{-6} + 2.02071 \times 10^{-13} = 3.5 \times 10^{-18} \\
q_{Henery \to Mikey} &= J_{Henery} A_{Henery} F_{Henery \to Mikey} \\
&= 3.5 \times 10^{-18} \times 15000 \times \frac{50}{84} = 9 \times 10^{-15}
\end{aligned}
$$

Subsequently, the message will reaches Mikey from Henery, Mikey has $A = 200$ and $T = 2$. Let assume Mikey averages 30 tweets per day; thus $\varepsilon = \frac{30}{2400} = 0.0125$. Henery has 670 tweets in total, of which 176 are retweets. Then, G_{ref} is calculated as follow:

$$
\begin{aligned}
G_{ref} &= \rho G_{Mikey} = \frac{176}{670} \times \frac{q_{Henery \to Mikey}}{A_{Mikey}} \\
&= 0.26 \times \frac{9. \times 10^{-15}}{200} = 1.17 \times 10^{-17} \\
E &= \varepsilon \sigma T_s^4 = 0.0125 \times 5.76 \times 10^{-8} \times 2^4 = 1 \times 10^{-8} \\
J &= E + G_{ref} = 1 \times 10^{-8} \\
q_{Miky \to Elric} &= J_{Miky} A_{Miky} F_{Miky \to Elric} \\
&= 1 \times 10^{-8} \times 200 \times \frac{17}{17} = 2 \times 10^{-6}
\end{aligned}
$$

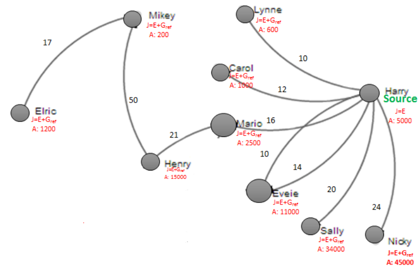


FIGURE 6. **The diffusion graph initiated by Harry.**

The diffusion that was initiated by Harry is complete once it reaches Elric. The diffusion RADDIFF model traverses the graph and guarantees the diffusion condition. Therefore, according to the example, the predicted diffusion graph of a message that begins from Harry is as shown in Figure 6. The proposed diffusion model has the feature of different influence spreaders, but due to the space limit, we focus our example on one influence spreader. We assume that Alice is another influence spreader. Then, the computation is similar to that for Harry.

IV. EXPERIMENTS

In this section, we will evaluate the RADDIFF model against state-of-art models, LIM and NETINF. The evaluation is done on a real dataset collected from Twitter; thus a description of dataset and its graph structure are presented next. Experiment setup, configuration, and results' analysis are discussed in the following.

A. DATASET DESCRIPTION

The dataset used in this work was collected using the Twitter Search API. The dataset has different features, including users common interests. This important feature renders the community detection more accurate because it finds the crucial relative community. To infer users interests, we have retrieved an expert list for each user and applied list-based methodology as proposed by Bhattacharya *et al.* [48]. The similarity between users requires knowing friends interests. Then, we mapped the links between each topic, including the number of friends who shared an interest in the same topic. Another feature is the graph connection links, which are based on the interaction between entities. To model how two nodes interact with each other to build the network structure, we applied the distance function that is based on how often two nodes interact with each other, as proposed by Falkowski *et al.* [54]. The dataset is considered a real dataset because it is assembled using real Twitter users with their tweets and expert lists. The dataset size contains 9,081 Twitter users with 17,573 edges. Even though the size is relatively small, the overall dataset contains more than 5 GB of data. This large size is due to the problem of interest clustering, which uses friends interests and expert lists, along with the number of a users friends who share common interests, to infer user interest. After detecting the correct community that

$$Interact(v,w) = \left\{ \begin{array}{cc} 0 & \text{if } v = w \\ min(NoI_{v,w}, NoI_{w,v})^{-1} & \text{if } NoI_{v,w} > 1 \wedge NoI_{w,v} > 1 \\ 1 & \text{Otherwise} \end{array} \right\} \tag{16}$$

matches the diffusion message, the RADDIFF model starts to predict the diffusion graph of that community. In addition, the advantage of the RADDIFF model is that with a relatively small number of users, the diffusion can reach millions of users, as will be shown in the analysis and results.

B. GRAPH STRUCTURE OF DATASET

The RADDIFF model is a graph-based model that is applied to a connected graph. The connections in the graph consider node interests and interactions. The interactions between two nodes are computed by including their interactions between each other. To model how two nodes interact with each other and build the network structure, we applied the distance function that is based on how often two nodes or users interact with each other. This method proposed by Falkowski *et al.* [54] is shown in Equation 16, as shown at the top of this page, where $NoI_{v.w}$ and $NoI_{w.v}$ are the number of interactions between nodes v and w initiated by v and w, respectively.

The interaction graph is normalized by the number of reciprocal interactions. The philosophy of the interaction graph is that two nodes with a high number of reciprocal interactions are closer and that the distance between two nodes that do not interact is 1. We model the graph based on interactions rather than structural graphs since we consider that node interactions with each other reflect the real interactions and how often two nodes connect. Since nodes interact with other nodes that share common interests, this approach meets the RADDIFF model aim for selecting the matched community with marketing keywords. In this context, the RADDIFF model applies a novel interest-based inference method using a node expert list, as proposed by Bhattacharya *et al.* [48]. This method was chosen because it achieves better results than that of the traditional topic-modeling approaches such as LDA [49], [50] and topical recommendations [51]. In addition, Bhattacharya *et al.* proposed social annotation to infer topical expertise [52], [53]. The list-based method has been used to infer the nodes interests, depending on the experts that the node follows. Mapping between the node and interests is recorded in the nodes interest vector. The vector is initiated if the node participates with at least three experts of interest i. The vector is also sorted according to the number of experts on a specific topic. When the interests have been associated with each node, the detecting community processes can be computed.

To detect suitable communities for diffusion, we have applied a density-based clustering algorithm termed IntClus [55]. The IntClus is selected because of its simplicity and effectiveness in identifying core, hub, and outlier nodes. This approach is proven capable in online social networks and dense networks. The process of detecting the community starts by receiving the interaction graph along with the parameters ε and μ and then stating each nodes vector interest. ε states interest-based similarity threshold for a given vertex in a network. Vertex becomes a core vertex if it has a minimum of μ neighbors with an interest-based similarity that exceeds the threshold ε. Therefore, μ is the minimal number of vertices to form a cluster with a core vertex. Then, clusters are formulated from core vertices, and parameters ε and μ determine the clustering of networks. Once the above steps are completed, IntClus will take the interaction graph and the parameters ε and μ. Afterwards, IntClus will be processed to identify whether the node is core, hub, or an outlier to show the resulting clustered graph. Each cluster is associated with multiple interests that represent the nodes' interest within it. This approach facilitates the process of matching the diffusion message keywords with the suitable community.

C. BASELINE METHODS

We compare the performance of RADDIFF with that of two well-known diffusion algorithms. First, the LIM was proposed by Yang and Leskovec [31]. This model shares a similar objective with the RADDIFF model of predicting diffusion based on the past infected nodes. In addition, it models the new infected nodes as a function of time based on past infections by associating each node with an influence function. The influence in the LIM is measured as the summation of influences of previously infected nodes. The LIM considers a node infected if it mentions the information. Therefore, the model defines the volume $V(t)$ as the number of nodes that mention the information at time t.

The second method is NETINF, proposed by Gomez-Rodriguez [56], which infers network diffusion and influence. This method assumes a static underlying network but considers the time of infection by a contagion; it is formulated as a probabilistic model of contagion that diffuses as a directed tree. It constructs a directed network G^* with multiple contagions spreading over it. Then, the contagion traverses the network to form a set of triples $(u, v, t_v)_c$, which means that contagion c has reached node v at time t_v by spreading from node u by propagating over the edge (u, v). The contagion inference is based on the probability that contagion c in a particular tree cascades that tree pattern, where the tree determines which nodes infect other nodes. In addition, the method includes the definition of the probability that the cascade c occurs in the network. Then, the estimation process occurs to infer the underlying cascade.

TABLE 2. Forecast error measures over RADDIFF, LIM, and NETINF on Twitter tweets.

Diffusion Model	Tweet	MAE	NMAE	RMSE	NRMSE	MAPE
RADDIFF	T1	14.37	0.33	14.72	0.37	18.2%
	T2	20.5	0.44	24.30	0.53	25.03%
	T3	11.6	0.37	13.6	0.43	14.43%
LIM	T1	17.4	0.44	26.2	0.49	18.54%
	T2	23.4	0.41	29.1	0.62	26.02%
	T3	24.9	0.43	17.6	0.45	19.74%
NETINF	T1	18.56	0.53	27.03	0.47	24.34%
	T2	35.6	0.54	27.5	0.78	27.3%
	T3	23.35	0.67	18.8	0.88	25.6%

Note: The unit for accuracy error computation are in million

D. EXPERIMENT CONFIGURATION

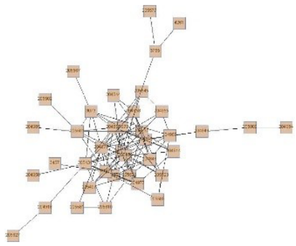
The diffusion between nodes in the network can be viewed as radiation reflected between network entities. The density of the contagion is formally represented as irradiation q that is reflected between network nodes. Since the objective of the RADDIFF model is to predict the diffusion over time, we evaluate this diffusion as a time series prediction task. The prediction in RADDIFF observes the seed nodes at time t and aims to predict the influence of contagion c at a future time $t + 1$. To measure the accuracy of the forecast, we have focused on quantifying the error between the forecasted and eventual values. Multiple accuracy forecast methods are available, including the Mean Absolute Error (MAE) [57], Normalized Mean Absolute Error (NMAE) [58], Root Mean Squared Error (RMSE) [57], Normalized Root Mean Squared Error (NRMSE) [59], and Mean Absolute Percentage Error (MAPE) [59]–[61]. In addition, cross validation was employed to evaluate the RADDIFF model [62]. This validation model estimates the accuracy of the predictive model, which is the aim of the proposed model. A 10-fold cross validation is adopted for validation, and we divide the contagion into 10 folds, where 9 are used for estimation and the remaining 1 is for testing and evaluation. Then, we predict the level of contagion at $t + 1$, which measures the difference between the actual and predicted values.

The mathematical modeling of the RADDIFF model is non-parametric, and the prediction is processed by including the current state data that are observed at each stage. As described earlier, this approach requires identifying the influence spreaders only and other parameters determined dynamically based on the data for each influencer. For testing purpose, we assume the diffusion guarantees that the next node has fewer connection links than its predecessor. The purpose of this condition is to assure the termination of the diffusion after a period of time. As the RADDIFF model predicts the future diffusion graph based on a time series prediction problem, we evaluate this model with the result shown by the LIM. The purpose of this evaluation is to reach an accurate estimation of model prediction performance and to measure the level of diffusion observations for node influence. The evaluation includes other measures for comparison, including precision, recall, and F-measure, which have been used extensively in measuring the accuracy of any approach that is based on finding the variation between actual and predicted values [63].

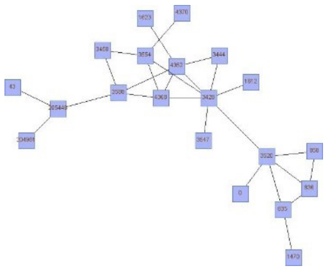
E. DIFFUSION INFLUENCE RESULTS AND ANALYSIS

We compare the RADDIFF model with both NETINF and LIM. We tested three different tweet categories, namely, the news, the media, and the world. These tweets were used for all diffusion methods to measure the forecast accuracy. According to the forecast shown by these tweets, we compute the forecast error for three models using the MAE, Normalized MAE (NMAE), RMSE, Normalized RMSE (NRMSE) and MAPE. Table 2 shows the relative error fits for the different prediction diffusion models. As shown in Table 2, the RADDIFF model provides results superior to that of other prediction diffusion models. In T1, which represents the first tweet tested, the RADDIFF model outputs 18.2% in MAPE; in T2 it outputs 25.3%, and in T3 it outputs 14.43%. The best value for LIM among three tweets is shown to be 18.54%, which is almost the same value resulted from the RADDIFF. The minimum percentage error value showed by NETINF, 24.34%, is almost double the value given by RADDIFF. In terms of NMAE, the worst value of 0.67 is produced by NETINF. Since NMSE measures the quality of an estimator, this metric determines the forecast between the estimator and what is estimated. Therefore, a large difference between NMAE and NMSE indicates an inconsistent error size. In the experiment, we observe that the difference is relatively reasonable in the case of RADDIFF. In contrast, LIM and NETINF have NMSE values that are double the MAE. The test comparison has done on three models with the dataset which has unit of million.

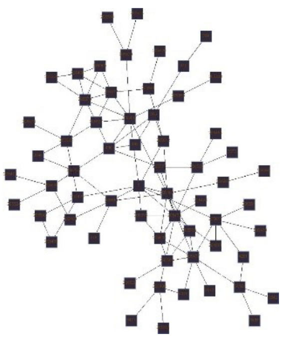
The resulted diffusion graphs of the RADDIFF model for each tweet entered are shown in Figure 7. As discussed in section III-A and III-B, the model predicts the diffusion pathway of certain information within a range of interested community. The model identifies the appropriate community for the information that being diffused in the network, and then determines the influencers to initiate the diffusion. For example, Figure 7 (a) predicts the information diffusion pathway that is targeting news community. The message diffusion starts from the center of the graph representing the influencer seeds, and then the diffusion proceed further to other nodes



(a)



(b)



(c)

FIGURE 7. Predictive RADDIFF Diffusion graph of three tweets concerning (a) news community, (b) media community and (c) world community. These graph are resulted to give an approximated prospective of information diffusion in online social networks.

TABLE 3. Accuracy measures over RADDIFF, LIM, and NETINF on Twitter tweets.

Diffusion Model	Tweets	Precision	F-measure
RADDIFF	T1	0.72	0.4
	T2	0.78	0.2
	T3	0.8	0.03
LIM	T1	0.63	0.4
	T2	0.57	0.09
	T3	0.4	0.13
NETINF	T1	0.58	0.25
	T2	0.67	0.23
	T3	0.7	0.1

till the leave nodes. The benefit behind this graphs is to provide an initial prospective of the diffusion before it takes place in reality.

Table 3 shows the RADDIFF precision values for T1, T2 and T3 as 0.72, 0.78, and 0.8, respectively. The corresponding F value results are 0.4, 0.2, and 0.03. LIM and NETINF, on the other hand, have resulted in low level of accuracy than RADDIFF. All F values of LIM and NETINF are lower than RADDIFF except the first tweet where RADDIFF and LIM reached the same value. This indicates that RADDIFF model has higher accuracy, ensures targeting the most relevant community, and save the time and cost for diffusion while hitting the most influencer nodes.

Figure 7(a), Figure 7(b), and Figure 7(c) indicate that the RADDIFF model concentrates on the interactions between top influencers because they can easily diffuse the contagion to their connections. Moreover, the resulted diffusion graph is relatively small, but it consists of the most crucial influencers that consume the minimum required time for diffusion. This condition affects the performance of diffusion, which consumes the minimum required cost and time to diffuse the specific contagion. In addition to time, the RADDIFF density outperforms LIM and NETINF since its coverage exceeds 6 million users because we consider the users who have received the tweets in their timeline as influenced users.

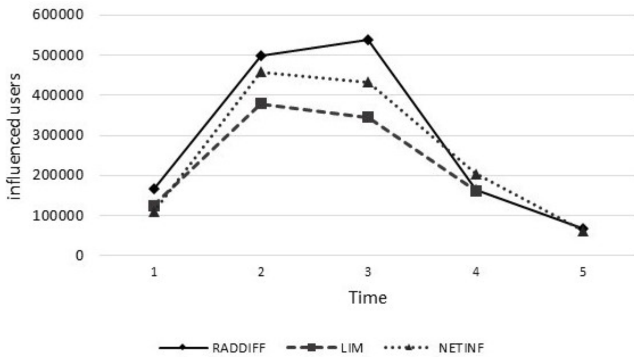


FIGURE 8. Number of influenced users of news community.

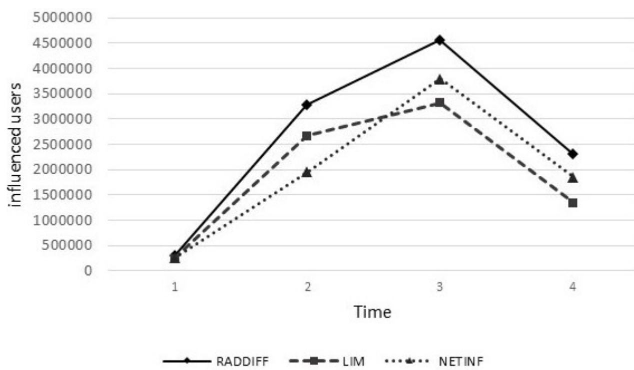


FIGURE 9. Number of influenced users of media community.

Figure 8 shows the distributions of the influenced users with the time of the news community. Five time series are shown, representing the time required for the message to be propagated in the network. Figure 9 illustrates the diffusion of the contagion in the community of users that concerns the media. It takes less time than the news community because it shows 4 time series. Finally, Figure 10 represents time levels of diffusion within users tweeting about world issues. It requires 7 time series for diffusion to reach the maximum influence.

In this context, the density of diffusion (Figure 8, 9, and 10) starts low, and over time, the tweets propagate faster while the density is increased at the peak time period. Then, after

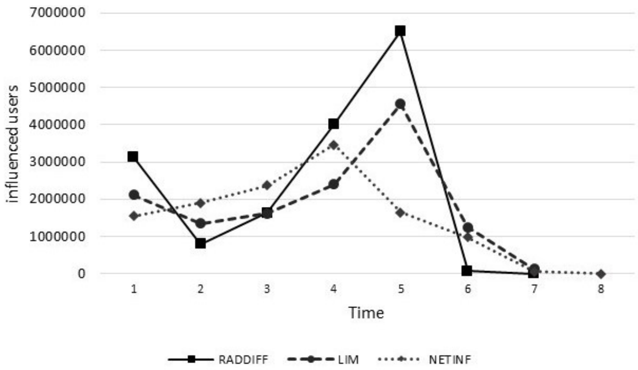


FIGURE 10. **Number of influenced users of world community.**

the peak period tweets, the topics life drops suddenly. This decrease reflects a normality of diffusion where the natural diffusion of viruses initiates with limited infectious and, over time, reaches its peak and is more infectious. Subsequently, people become immunized from the virus, which prevents further spread.

Based on the experimental results shown in Table 2, we have collected observations regarding the RADDIFF model. It is clear that the RADDIFF model can reach levels of diffusion with realistic goodness of fit that results from substantial benefits over the non-parametric features. Moreover, we also observe that RADDIFF gives superior results for modeling the adoption of interested communities in online social networks than for modeling the adoption of complete and non-clustered networks. In terms of the time required to propagate a certain message, the model consumes the least possible time relative to other well-known prediction diffusion models. Therefore, the model is applicable for large-scaled networks, considering its temporal and structural patterns.

V. CONCLUSION

Information diffusion prediction modeling has been studied from various perspectives, including marketing, business, medical, and education. A new computer-based information diffusion model for online social networks is required. In this context, we have presented the RADDIFF model to predict information diffusion in online social networks. This model aims at predicting information diffusion graphs that can show a preliminary prospective of any campaign in online social networks. In this matter, organizations that seek to diffuse their campaign throughout online social networks can use the RADDIFF model to predict their campaign process before it is launched. Moreover, this type of modeling can minimize the cost of marketing via the inappropriate influencers that do not match the campaigns audience target. The validation and evaluation measurements have concentrated on measuring the model accuracy and the density of influenced users. The experimental results show that the model achieves high predictive accuracy. With respect to density, the information diffusion starts narrow and reaches the peak time within average times. This finding indicates the distance that the model reaches throughout the network. Future planned work includes enhancing the overall model performance to include datasets with different social media platforms while also optimizing the model scalability.

REFERENCES

[1] T. Kawamoto and N. Hatano, "Viral spreading of daily information in online social networks," *Phys. A, Stat. Mech. Appl.*, vol. 406, pp. 34–41, Jul. 2014.

[2] A. Anderson, D. Huttenlocher, J. Kleinberg, J. Leskovec, and M. Tiwari, "Global diffusion via cascading invitations: Structure, growth, and homophily," presented at the 24th Int. Conf. World Wide Web, Florence, Italy, 2015.

[3] K. Lerman and R. Ghosh, "Information contagion: An empirical study of the spread of news on Digg and Twitter social networks," in *Proc. 4th ICWSM*, vol. 10. 2010, pp. 90–97.

[4] F. Wang, H. Wang, and K. Xu, "Diffusive logistic model towards predicting information diffusion in online social networks," in *Proc. ICDCS Workshop*, Washington, DC, USA, 2012, pp. 133–139.

[5] F. Wang, H. Wang, K. Xu, J. Wu, and X. Jia, "Characterizing information diffusion in online social networks with linear diffusive model," in *Proc. IEEE 33rd Int. Conf. Distrib. Comput. Syst.*, Jul. 2013, pp. 307–316.

[6] Y. Hu, R. J. Song, and M. Chen, "Modeling for information diffusion in online social networks via hydrodynamics," *IEEE Access*, vol. 5, pp. 128–135, 2017.

[7] Y. Hu and M. Chen, "Information diffusion prediction in mobile social networks with hydrodynamic model," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–5.

[8] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: A survey," *ACM Sigmod Rec.*, vol. 42, no. 2, pp. 17–28, 2013.

[9] J. Niu and L. Wang, "Structural properties and generative model of non-giant connected components in social networks," *Sci. China Inf. Sci.*, vol. 59, p. 123101, Dec. 2016. [Online]. Available: https://doi.org/10.1007/s11432-015-0790-x

[10] H. S. Rodrigues and M. J. Fonseca, "Can information be spread as a virus? Viral marketing as epidemiological model," *Math. Methods Appl. Sci.*, vol. 39, no. 16, pp. 4780–4786, 2016.

[11] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," *Marketing Lett.*, vol. 12, no. 3, pp. 211–223, 2001.

[12] N. Alrajebah, "Investigating the structural characteristics of cascades on Tumblr," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, 2015, pp. 910–917.

[13] M. Farajtabar, M. Gomez-Rodriguez, M. Zamani, N. Du, H. Zha, and L. Song, "Back to the past: Source identification in diffusion networks from partially observed cascades," in *Proc. AISTATS*, 2015, pp. 232–240.

[14] C. Tong, W. He, J. Niu, and Z. Xie, "A novel information cascade model in online social networks," *Phys. A, Stat. Mech. Appl.*, vol. 444, pp. 297–310, Feb. 2016.

[15] J. Kim, J. Kim, O. Kwon, O. Kwon, D. H. Lee, and D. H. Lee, "Social influence of hubs in information cascade processes," *Manage. Decision*, vol. 55, no. 4, pp. 730–744, 2017.

[16] Q. Wang, Y. Jin, T. Yang, and S. Cheng, "An emotion-based independent cascade model for sentiment spreading," *Knowl.-Based Syst.*, vol. 116, pp. 86–93, Jan. 2017.

[17] M. Granovetter, "Threshold models of collective behavior," *Amer. J. Sociol.*, vol. 83, no. 6, pp. 1420–1443, 1978.

[18] A. Nematzadeh, E. Ferrara, A. Flammini, and Y.-Y. Ahn, "Optimal network modularity for information diffusion," *Phys. Rev. Lett.*, vol. 113, no. 8, p. 088701, 2014.

[19] A. Guille and H. Hacid, "A predictive model for the temporal dynamics of information diffusion in online social networks," in *Proc. WWW Companion*, Lyon, France, 2012, pp. 1145–1152

[20] J. T. Khim, V. Jog, and P.-L. Loh, "Computing and maximizing influence in linear threshold and triggering models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 4538–4546.

[21] E. Yoo, W. Rand, M. Eftekhar, and E. Rabinovich, "Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises," *J. Oper. Manage.*, vol. 45, pp. 123–133, Jul. 2016.

[22] S. Peng, A. Yang, L. Cao, S. Yu, and D. Xie, "Social influence modeling using information theory in mobile social networks," *Inf. Sci.*, vol. 379, pp. 146–159, Feb. 2017.

[23] C. Lagnier, L. Denoyer, E. Gaussier, and P. Gallinari, "Predicting information diffusion in social networks using content and user's profiles," in *Proc. 35th Eur. Conf. Inf. Retr. Res. (ECIR)*, 2013, pp. 74–85.

[24] J. Dhar, A. Jain, and V. K. Gupta, "A mathematical model of news propagation on online social network and a control strategy for rumor spreading," *Soc. Netw. Anal. Mining*, vol. 6, p. 57, Dec. 2016.

[25] Y. Hashimoto and H. Ohsaki, "On estimating message diffusion dynamics in epidemic broadcasting," in *Proc. IEEE 39th Annu. Comput. Softw. Appl. Conf. (COMPSAC)*, Jul. 2015, pp. 133–138.

[26] L. A. Adamic, T. M. Lento, E. Adar, and P. C. Ng, "Information evolution in social networks," in *Proc. 6th ACM Int. Conf. Web Search Data Mining*, 2016, pp. 473–482.

[27] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," in *Proc. 21st Int. Conf. World Wide Web*, 2012, pp. 519–528.

[28] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," *ACM Trans. Knowl. Discovery from Data*, vol. 5, no. 4, pp. 1–37, Feb. 2012.

[29] S. A. Myers and J. Leskovec, "On the convexity of latent social network inference," presented at the 23rd Int. Conf. Neural Inf. Process. Syst., Vancouver, BC, Canada, 2010.

[30] M. Gomez-Rodriguez, D. Balduzzi, and B. Schölkopf, "Uncovering the temporal dynamics of diffusion networks," in *Proc. ICML*, Washington, DC, USA, 2011, pp. 561–568.

[31] J. Yang and J. Leskovec, "Modeling information diffusion in implicit networks," presented at the IEEE Int. Conf. Data Mining, 2010.

[32] A. L. P. Chen, R. Zafarani, and M. A. Abbasi, *Social Media Mining: An Introduction*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[33] T.-T. Quach and J. D. Wendt, "A diffusion model for maximizing influence spread in large networks," in *Proc. Int. Conf. Soc. Inform.*, 2016, pp. 110–124.

[34] A. Chaudhury, P. Basuchowdhuri, and S. Majumdar, "Spread of Information in a Social Network Using Influential Nodes," in *Proc. 16th Pacific-Asia Conf. Adv. Knowl. Discovery Data Mining*, 2012, pp. 121–132.

[35] N. S. N. Abadi and M. R. Khayyambashi, "Influence maximization in viral marketing with expert and influential leader discovery approach," in *Proc. 8th Int. Conf. e-Commerce Develop. Countries, Focus e-Trust (ECDC)*, 2014, pp. 1–8.

[36] J. Zhou, Y. Zhang, and J. Cheng, "Preference-based mining of top-*K* influential nodes in social networks," *Future Generat. Comput. Syst.*, vol. 31, pp. 40–47, Feb. 2014.

[37] Q. Wang, Y. Jin, Z. Lin, S. Cheng, and T. Yang, "Influence maximization in social networks under an independent cascade-based model," *Phys. A, Stat. Mech. Appl.*, vol. 444, pp. 20–34, Feb. 2016.

[38] G. Tong, W. Wu, S. Tang, and D. Z. Du, "Adaptive influence maximization in dynamic social networks," *IEEE/ACM Trans. Netw.*, vol. 25, no. 1, pp. 112–125, Feb. 2017.

[39] H. Zhuang, Y. Sun, J. Tang, J. Zhang, and X. Sun, "Influence maximization in dynamic social networks," in *Proc. IEEE 13th Int. Conf. Data Mining (ICDM)*, Dallas, TX, USA, Dec. 2013, pp. 1313–1318.

[40] H. Kim, K. Beznosov, and E. Yoneki, "Finding influential neighbors to maximize information diffusion in twitter," in *Proc. WWW Companion*, Seoul, South Korea, 2014, pp. 701–706

[41] H. Wu, Z. Zhang, K. Yue, B. Zhang, and W. Liu, "Maximizing the cooperative influence spread in a social network oriented to viral marketing," in *Proc. Asia–Pacific Web Conf.*, 2016, pp. 3–15.

[42] S. Yang, X. Wang, and C. Liu, "A heat diffusion model based algorithm for personalized influence maximization in social network," *J. High Speed Netw.*, vol. 23, no. 2, pp. 137–147, 2017.

[43] Q. Li, T. Zhou, L. Lü, and D. Chen, "Identifying influential spreaders by weighted LeaderRank," *Phys. A, Stat. Mech. Appl.*, vol. 404, pp. 47–55, Jul. 2014.

[44] Y. Xia, X. Ren, Z. Peng, J. Zhang, and L. She, "Effectively identifying the influential spreaders in large-scale social networks," *Multimedia Tools Appl.*, vol. 75, no. 15, pp. 8829–8841, 2016.

[45] V. V. Le, H. T. Nguyen, V. Snasel, and T. T. Dao, "Identify influential spreaders in online social networks based on social meta path and PageRank," in *Proc. Int. Conf. Comput. Social Netw.*, 2016, pp. 51–61.

[46] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the Web," Stanford InfoLab, Stanford Univ., Stanford, CA, USA, Tech. Rep. SIDL-WP-1999-0120, 1999.

[47] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," presented at the 7th Int. Conf. World Wide Web, vol. 7. Brisbane, Qld., Australia, 1998.

[48] P. Bhattacharya, M. B. Zafar, N. Ganguly, S. Ghosh, and K. P. Gummadi, "Inferring user interests in the Twitter social network," presented at the 8th ACM Conf. Recommender Syst., Silicon Valley, CA, USA, 2014.

[49] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning, "Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora," presented at the Conf. Empirical Methods Natural Lang. Process., vol. 1. Singapore, 2009.

[50] M. Michelson and S. A. Macskassy, "Discovering users' topics of interest on twitter: A first look," presented at the 4th Workshop Anal. Noisy Unstructured Text Data, Toronto, ON, Canada, 2010.

[51] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi, "Cognos: Crowdsourcing search for topic experts in microblogs," presented at the 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Portland, OR, USA, 2012.

[52] N. K. Sharma, S. Ghosh, F. Benevenuto, N. Ganguly, and K. Gummadi, "Inferring who-is-who in the Twitter social network," presented at the ACM Workshop Workshop Online Soc. Netw., Helsinki, Finland, 2012.

[53] U. Wickström, "Heat transfer by radiation," in *Temperature Calculation in Fire Safety Engineering*, Cham, Switzerland: Springer, 2016, pp. 65–87.

[54] T. Falkowski, A. Barth, and M. Spiliopoulou, "DENGRAPH: A density-based community detection algorithm," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell.*, Nov. 2007, pp. 112–115.

[55] L. AlSuwaidan and M. Ykhlef, "Interest-Based Clustering Approach for Social Networks," *Arabian J. Sci. Eng.*, pp. 1–13, Aug. 2017.

[56] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," in *Proc. KDD*, New York, NY, USA, 2010, pp. 1019–1028.

[57] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature," *Geosci. Model Develop.*, vol. 7, no. 3, pp. 1247–1250, 2014.

[58] W. Ma, D. Zheng, Z. Zhang, and B. Chen, "Bias-compensated normalized least mean absolute deviation algorithm with noisy input," in *Proc. 20th Int. Conf. Inf. Fusion (Fusion)*, Jul. 2017, pp. 1–15.

[59] J. S. Armstrong and F. Collopy, "Error measures for generalizing about forecasting methods: Empirical comparisons," *Int. J. Forecasting*, vol. 8, pp. 69–80, Jun. 1992.

[60] C. Tofallis, "A better measure of relative prediction accuracy for model selection and model estimation," *J. Oper. Res. Soc.*, vol. 66, no. 8, pp. 1352–1362, 2015.

[61] H. K. Sul, A. R. Dennis, and L. I. Yuan, "Trading on twitter: Using social media sentiment to predict stock returns," *Decision Sci.*, vol. 48, no. 3, pp. 454–488, 2017.

[62] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," in *Encyclopedia of Database Systems*. New York, NY, USA: Springer, 2009, pp. 532–538.

[63] K. M. Ting, "Precision and recall," in *Encyclopedia of Machine Learning*. New York, NY, USA: Springer, 2011, p. 781.

LULWAH ALSUWAIDAN received the Master's degree in information systems from the Department of Information Systems, College of Computer and Information Sciences, King Saud University, where she is currently pursuing the Ph.D. degree in information systems. She is currently a Lecturer with the Department of Information Management, College of Computer and Information Sciences, Al-Imam Mohammad bin Saud Islamic University. During her graduate studies, she has had the opportunity to participate in various conferences and has published various journal articles. She received a certificate of completing an English program from Florida State University, U.S. Her main research interests lie in the field of data mining, information diffusion, and social analysis.

MOURAD YKHLEF received the B.Eng. degree in computer science from Constantine University, Algeria, the M.Sc. degree was in artificial intelligence from University Paris 13, France, and the Ph.D. degree in computer science from University Bordeaux 1, France. He is currently a Professor with the Department of Information Systems, College of Computer and Information Sciences, King Saud University, Saudi Arabia. His main research interests include data mining, data warehouse, XML and bio-inspired computing.