A Novel Approach to Information Spreading Models for Social Networks

Burcu Sayin, Serap Şahin
Department of Computer Engineering, Izmir Institute of Technology, Izmir, Turkey
Email: burcusayin@iyte.edu.tr, serapsahin@iyte.edu.tr

Abstract— **Analyzing and modelling the spreading of any information through a social network (SN) is an important issue in social network analysis. The proposed solutions for this issue do not only help with observing the information diffusion, but also serve as a valuable resource for predicting the characteristics of the network, developing network-specific advertising, etc. Up-to-date approaches include probabilistic analysis of information spreading and the information cascade models. In this paper, we propose a hybrid model, which considers an information spreading model, and combines it with cascades and social behavior analysis. We propose a new hybrid usage approach to represent a real-world modelling for the information spreading process.**

Keywords-social network analysis; information spreading; information cascades.

I. INTRODUCTION

Information spreading on social networks (SNs) is getting more popular in social network analysis. Thanks to the developing technology, information has become quickly accessible, especially via SNs. This situation creates new domains on SNs, such as advertising, marketing, etc. Hence, it is important to have an information spreading model for predicting the effect of the information on SNs.

In the literature, there are many models that either support or modify the Susceptible – Infected – Removed (*SIR*) model [1], [2], [3] or adopt it to new approaches. We selected some of the most current ones and presented them in Section 3. However, it is hard to find a model that matches real-life scenarios because SNs are dynamic platforms and SN users act with their emotions. Therefore, models should also represent SN users' real behaviors. Developing such a model serves as a solution to problems in many areas, like security. For example, in the case of any malicious information existing in the network, we can predict its spreading area and pattern. In this way, we can take precautions against a possible crisis. These are the reasons for which we would like to propose a real-world information spreading model.

In this paper, we present out work in progress as well as our opinion. The existing models do not provide a complete solution to reflect real SN user's decisions for spreading the information. We point out the deficient points on proposed studies and propose an alternative hybrid model.

The paper is organized as follows. Section 2 proposes some of the current requirements for a realistic information spreading model. Section 3 gives an overall explanation of the basis information spreading model "*SIR*" in the literature and its current applications, with some modifications and new approaches. Section 4 includes our proposed hybrid solution for a real-world modelling of information spreading. We conclude the paper and outline our future research directions in Section 5.

II. IDENTIFIED REQUIREMENTS OF A REALISTIC INFORMATION SPREADING MODEL

As SNs have a dynamic characteristic, it is hard to model the spreading of information only with a probabilistic approach. Current information spreading models should focus on a user-specific approach and consider SN users' behavioral effects, because SN users shape the diffusion of any information on the network.

Based on our research, we noted some requirements, as listed below, for a realistic information spreading model with behavioral analysis:

- **Popularity of the source:** Popularity level of the information source affects SN users' decisions on whether to spread that information or not. This factor can also be related with the trustworthiness and credibility of the source.
- **Strength of relations among users:** Strong relations/features (such as similar political views, education, gender, etc.) among SN users cause information to spread faster.
- **Content of the information:** If a SN user gives importance to the content of any information, he/she is more likely to spread it.
- **Personal interests:** If a SN user's interests are related to the information, he/she becomes more likely to spread it.
- **Privacy preferences:** Privacy preferences of a SN user in his/her profile have an impact on information spreading. For example, if the user is conservative about his/her privacy, then he/she is more likely to abstain from spreading information.

Each mentioned requirement affects the information spreading on SNs with a probability. However, the value of this probability may vary according to each user. Hence, developing models for information spreading should take into account these real-world conditions.

III. INFORMATION SPREADING MODELS

In SNs, information spreads via posts from one user to another. This spreading continues until it loses actuality and attractiveness to users. In literature, researchers proved that the information spreading process and epidemics resemble each other [4]. Hence, the *SIR* model reflects epidemics. We present this model in the following section and then provide an overview of up-to-date information spreading models.

A. *SIR Model*

The *SIR* Model is based on epidemics. Epidemics spread for a time and then lose their effect. The size of the area affected by the epidemics depends on population size. It is obvious that the probability of the disease spreading in a crowded area is higher than in a deserted area. Hence, population size is an important determinant in the spreading process of epidemics [5]. Similarly, in SNs, a post spreads quickly if the owner of the post has lots of connections.

In epidemics, time evolution of a disease is managed by a threshold; information spreading also has a threshold theorem. The Threshold Theorem of Kermack–McKendrick [1], [2] defines the evolution process of epidemics. This theorem models the population with three types: Susceptible (*S*), Infected (*I*), and Removed (*R*), which constitutes the *SIR* model. Each variable/state refers to the number of people in the related group. Susceptible ones are ignorant, which means that they are not yet infected but have a potential to be an Infected. Infected ones have the disease and they can infect the Susceptible ones. Removed ones have recovered from the disease and stopped the spreading process.

The theorem consists of two differential equations which define transitions between *S*, *I* and *R* states [1], [2]:

- Transition from Susceptible state to Infected state,
- Transition from Infected state to Removed state.

A critical point here is to model the transition from the Infected state to Removed. Researchers first proposed a counter value (*ctr*) to control this process [4], [6]. The main idea behind this value is that it counts the number of users who became infected, and it stops spreading when this number reaches to the *ctr* value. The value is determined before the spreading process and is valid for each user in the network. Unsurprisingly, if we choose a big value for *ctr*, information reaches a bigger portion of the network, but more rounds are required to complete the spreading process. Eventually, *ctr* controls the termination of the spreading process and the size of spreading area in a network.

In addition, information can be observed in two different ways within a spreading process; (i) static information and (ii) dynamic information. During the whole spreading process, if the information does not have any revision, it is accepted as static information. In the contrary case, it becomes dynamic information [7]. When we consider this concept in SNs, users may revise someone's post and publish it as a new post, so information in SNs is dynamic. Hence, we need some modifications to the *SIR* model, or we need new approaches to model the current information spreading processes.

B. *Current Information Spreading Models*

Although most current studies consider the *SIR* model as a baseline and modify it according to today's requirements, some of them also propose new approaches, such as cascades. Information cascades allow us to predict how well the information will spread. This section first summarizes the studies that focus on the adaptation of the *SIR* model and then describes an information spreading model with cascades. We do not provide an algorithmic comparison between the current studies, because our aim is not to find an algorithm which provides the best performance. Instead, we focus on how well the proposed algorithms represent real-world circumstances. In this study, we consider the effectiveness of the algorithms as the achievement level of realistic modelling.

Bao et al. [8] criticizes the *SIR* model in terms of the idea behind the Infected state. They propose that an infected user does not have to believe/accept the information; may also oppose that information. Hence, they divided the infected state into two distinct ones: (i) positive infected (supports the information) and (ii) negative infected (opposes the information). They named this model Susceptible – Positive Infected – Negative Infected – Removed (*SPNR*). According to the *SPNR* model, when an ignorant user receives information from a positive/negative spreader, becomes a new positive/negative spreader with a probability value [8]. In the same way, there is a probability that a positive spreader may affect a negative spreader or vice versa. If a positive/negative spreader gets the information from a stifler (removed user), becomes a stifler, also with a probability. They define the transition from a spreader state to a removed one with a spreading threshold. This model is an enhanced version of *SIR* that takes into account users' decisions to believing the information. However, they only use this decision as a probabilistic value; they do not consider the mentioned requirements.

Serrano et al. [9] consider that a SN user may have a first impression about an information before being infected by other users. In this model, they modified *SIR* and proposed the following four states: (i) neutral (initial state), (ii) infected (believe the information), (iii) vaccinated (believe the anti-information before being infected) and (iv) cured (believe the anti-information after being infected). According to this model, all users are initially neutral. Then, they assign some of them as infected. Infected ones start to infect their neutral neighbors with a given probability. To simulate cured or vaccinated ones, they define a time as delay. At that time, a randomly selected infected user starts to spread anti-information, which says the opposite of the original information in the network. Hence, they try to cure or vaccinate their neighbors with a probability of accepting or denying (probAcceptDeny). Finally, cured and vaccinated ones try to cure or vaccinate their neighbors with the value of probAcceptDeny. This model uses an agent-based modelling so that it can reflect the real world better than *SIR*. However, it is still insufficient to be applied to SNs today.

Cordasco et al. [10] consider the infected state of the *SIR* model from a different aspect. They propose that a user may not immediately start spreading just after it is infected; they define a new state for this situation: "aware". They claim that there should be a threshold that controls the transition from being aware to start spreading. This model resembles the Susceptible – Exposed – Infected – Removed (*SEIR*) epidemic model [11], which differs from *SIR* model with the additional "Exposed" state. This state contains people who had contact with an infected user but have not yet started to infect other people. Similarly, Cordasco et al. [10] propose three states: (i) ignorant, (ii) aware and (iii) spreading. As usual, all users are ignorant at the beginning. When an ignorant user takes information from a spreader, it becomes aware. To be a spreading one, any aware user should take the information from more than a pre-defined number (threshold value) of spreading users. This model has no state for removed, but they define a termination rule in the original paper [10]. Although this model considers the transition process from being aware to start spreading, the model is not sufficient to represent today's information spreading process.

Tong. et al. [12] describes an information cascade model in SNs. First, they provide an extensive study on cascade scales, the scope of the cascade subgraphs, and topological attribute of spread tree. Then, based on the evaluation results, they analyze the spread of the user's decisions for city-wide activities. Decisions include "want to take part in the activity" and "be interested in the activity". This study introduces three mechanisms to use for making a decision:

- **Equal probability:** A user has an equal probability to make any of two decisions.
- **Similarity of users:** Similarity of users is the criteria to make a decision for any user.
- **Popularity of users:** Popularity of users affects users' decision.

Experimental results show that the popularity of users is an important criterion for information spreading. Although this study evaluates some user-specific parameters, such as popularity of the information source and similarity of users to model information spreading on SNs, it does not satisfy all the requirements proposed in Section 2 and it does not consider an epidemic approach.

When we examine the existing studies, we notice that most of them consider a behavioral effect/model on the information spreading process and so modify the *SIR* model to represent this effect in some way. However, they do not propose the factors that affect this behavioral model; they just take it as a probabilistic value for the behavior decision. Indeed, there are many factors (some of which were proposed in Section 2 as the requirements) that affect SN users' behavior on the information spreading process, and they are completely interrelated. Hence, we need to consider them all as a complete impact on users' decisions for the spreading process.

IV. A Hybrid Information Spreading Model

We propose to develop a hybrid model, which considers the models of Bao et al. [8] and Cordasco et al. [10] but modifies their threshold idea by using information cascade characteristics to meet requirements mentioned in Section 2. Novelty comes from using such a hybrid model, which combines epidemic models with up-to-date approaches like cascades to represent behavioral effect on SNs. Considering this effect together with the proposed requirements is so important to observe information spreading on SNs because it affects users' decision on spreading the specific information or not. By using this approach, we can make more realistic transitions between different states of our model. In addition, we will use the idea of Bao et al. [8] regarding the infected state. We will also divide our infected state into two: positive infected and negative infected, because there is a probability that a user will reject an information.

Figure 1 shows the state transitions of our model. The proposed model includes the following properties:

- There will be five states: (i) ignorant (user is not aware of the information), (ii) aware (user is aware of the information but he/she has not started to spread it), (iii) positive infected (user believes the information and spreads it) (iv) negative infected (user opposes the information and tries to convince other people in this way) and (v) removed (user stops spreading).

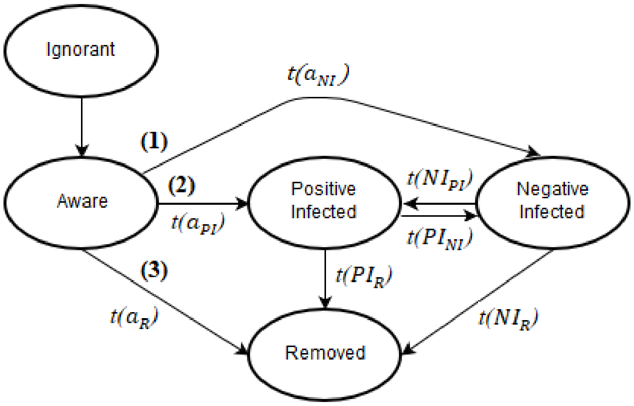

Figure 1. A Hybrid Information Spreading Model

- Initially, we assume that all users are ignorant. Then, some of them are selected as positive infected and some as negative infected. This selection may be important for some domains. For example, if we are working in the advertising or marketing sectors, it is important to reach more users in a short time. Hence, the selection process of initial positive/negative infected users should be performed according to the topology of the network. After this selection, information starts to be spread in the network.

- If an ignorant user takes the information from a positive/negative infected user, he/she becomes aware. After being an aware, the user passes to one of three possible states: (1) he/she may believe that the information is true and pass to a positive infected state via function "$t(a_{PI})$: transition from aware state to positive infected state", (2) he/she may refuse the information and decide to infect others negatively by passing to a negative infected state via function "$t(a_{NI})$: transition from aware state to negative infected state", and (3) an aware user may prefer not to infect any other user either positively or negatively. In this case, that user may pass directly to the removed state via transition function $t(a_R)$.
- After being a positive/negative infected, there may be a transition between those two infected states, which can be controlled with "$t(NI_{PI})$: transition function from negative infected state to positive infected state" or "$t(PI_{NI})$: transition function from positive infected state to negative infected state." Alternatively, they may pass to removed state via "$t(NI_R)$: transition function from negative infected state to removed state" and "$t(PI_R)$: transition function from positive infected state to removed state".
- All transition functions will have a threshold value, including the effect of cascading mechanisms and behavioral effects to meet all five requirements. Hence, these functions will depend on a user-specific approach. We will first analyze each user based on the requirements and define user-specific behavioral effect values for them (training phase). This means that each SN user will have a behavioral impact value and this value will be used in the information spreading process. Hence, users will make a decision based on this impact value for any transition in our model.

Consequently, we will base our hybrid model on the modified version of the *SIR* model and generate a new formulation by also using information cascades and users' social behavior analysis. To verify our model, we will implement both referred models [8], [10] and our proposed model in a real SN dataset to observe effectiveness, and then we will compare them in terms of success and failure rates on real-world modelling.

V. CONCLUSION

In this paper, we discussed the main information spreading model *SIR* and the current modifications of it. We also emphasized that information cascades are important to adjust information spreading models to SNs to create more realistic structures. Hence, we are working on developing a hybrid information spreading model which can meet the presented requirements and dynamism of SNs. Because users' decisions on spreading any information also depend on social behavioral factors, we will include behavioral analysis of SN users in our model. What we expect from this research is that anyone will be able to use our model to predict the spreading area and pattern of an information so that they can measure the effect of it on SNs. Additionally, this model can be used for interaction analysis among SN users.

As this paper proposes our preliminary work, we roughly provided our model. We will continue with the formulation, validation, and simulation phases of the model.

REFERENCES

[1] W. O. Kermack, and A. G. McKendrick, "A Contribution to the Mathematical Theory of Epidemics", Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, Volume 115, pp. 700-721, 1927. DOI: 10.1098/rspa.1927.0118.

[2] F. Brauer, "The Kermack–McKendrick epidemic model revisited", Mathematical Biosciences, Volume 198, Issue 2, pp. 119-131, ISSN 0025-5564, 2005. DOI: http://dx.doi.org/10.1016/j.mbs.2005.07.006.

[3] C. Nowzari, V. M. Preciado, and G. J. Pappas, "Analysis and Control of Epidemics: A Survey of Spreading Processes on Complex Networks," IEEE Control Systems, Volume 36, Issue 1, pp. 26-46, 2016. DOI:10.1109/MCS.2015.2495000.

[4] A. Demers et al., "Epidemic algorithms for replicated database maintenance", In Proceedings of the sixth annual ACM Symposium on Principles of distributed computing (PODC '87), Fred B. Schneider (Ed.). ACM, New York, NY, USA, pp. 1-12, 1987. DOI=10.1145/41840.41841 http://doi.acm.org/10.1145/41840.41841.

[5] D. J. Daley and D. G. Kendall, "Stochastic Rumours", IMA Journal of Applied Mathematics (Institute of Mathematics and Its Applications) 1, pp. 42-55, 1965. DOI: 10.1093/imamat/1.1.42.

[6] N. T. J. Bailey, "The Mathematical Theory of Infectious Diseases and its Applications", Hafner Press, Second Edition, 1975.

[7] Y. Zhang, S. Zhou, Z. Zhang, J. Guan, and S. Zhou, "Rumor Evolution in Social Networks". Physical Review E, vol. 87, no. 3, Article ID 032133, 2013.

[8] Y. Bao, C. Yi, Y. Xue, and Y. Dong, "A new rumor propagation model and control strategy on social networks", In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '13). ACM, New York, NY, USA, pp. 1472-1473, 2013. DOI:http://dx.doi.org/10.1145/2492517.2492599.

[9] E. Serrano, C. Á. Iglesias, and M. Garijo, "A Novel Agent-Based Rumor Spreading Model in Twitter", In Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion). ACM, New York, NY, USA, pp. 811-814, 2015. DOI: http://dx.doi.org/10.1145/2740908.2742466.

[10] G. Cordasco, L. Gargano, A. A. Rescigno, and U. Vaccaro, "Brief Announcement: Active Information Spread in Networks", In Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing (PODC '16). ACM, New York, NY, USA, pp. 435-437, 2016. DOI: http://dx.doi.org/10.1145/2933057.2933069.

[11] J. Zhang, J. Li, and Z. Ma, "Global Dynamics of an SEIR Epidemic Model with Immigration of Different Compartments", Acta Mathematica Scientia, Volume 26, Issue 3, pp. 551-567, ISSN 0252-9602, 2006. DOI: http://dx.doi.org/10.1016/S0252-9602(06)60081-7.

[12] C. Tong, W. He, J. Niu, and Z. Xie, “A novel information cascade model in online social networks”, Physica A: Statistical Mechanics and its Applications, Volume 444, pp. 297-310, 2015. DOI:10.1016/j.physa.2015.10.026.