A Novel Approach to Information Spreading Models for Social Networks

Burcu Sayin
Department of Computer Engineering
Izmir Institute of Technology
Izmir, Turkey
burcusayin@iyte.edu.tr

Serap Şahin
Department of Computer Engineering
Izmir Institute of Technology
Izmir, Turkey
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. Proposed solutions for this issue do not only help 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, combines it with cascades and social behavior analysis. We propose a new hybrid usage approach to represent a real-world modelling for information spreading process.**

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

I. INTRODUCTION

Information spreading on social networks is getting more popular in social network analysis. Thanks to the developing technology, information has become quickly accessible especially via social networks (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 an information on SNs.

In the literature, there are many models either supports/modifies the base model SIR (Susceptible – Infected - Removed) or adopts it to new approaches. We selected some of the most current ones and propose them in the following section. However, it is hard to find a model that matches with the real-life scenario because SNs are dynamic platforms and SN users act with their feelings. Therefore, models should also represent SN users' real behaviors. Developing such a model also serves solution to problems in many areas like security. For example; if we have such a model, we can know the spreading pattern of an information. Hence, in case of a malicious information existing in the network, we can predict the spreading area and pattern of it. In this way, we can take precaution for a possible crisis. These are the reasons that we would like to propose a real-world information spreading model.

This paper involves the contribution which we are currently working on and shows our point of view. The existing models do not provide a complete solution to reflect real SN user's decisions for information spreading. We point out the deficient points on proposed studies and propose an alternative hybrid model.

The rest of the paper is organized as follows. We give an overall explanation about the basis information spreading model "SIR" in the literature and then show current applications of it with some modifications or new approaches in Section 2. Section 3 includes our proposed hybrid solution for a real-world model of information spreading. We conclude the paper and propose our research direction in Section 4.

II. INFORMATION SPREADING MODELS

In SNs, information spreads via posts from one user to another. This spreading continues until it loses actuality and attraction from users. In literature, researchers proved that information spreading process and epidemics resemble each other [1]. Hence, SIR model reflects epidemics. We give this model in the following section and then provide an overview about up-to-date information spreading models.

A. SIR Model

SIR Model is based on epidemics. Epidemics spread for a time and then lose its effect; information spreading also has the same behavior but it has a threshold theorem. The theorem models the population with three types; susceptible (s), infected (i), and removed (r) which constitutes SIR model. "Susceptible" ones are ignorant. After a susceptible one gets an information from an infected one, it becomes "Infected". "Removed" means it stopped spreading process after a while. The time passed for the transition from "Infected" state to "Removed" state is variable. It is defined by some thresholds or termination rules.

In epidemics, time evolution of a disease is managed by a threshold which is defined by the number of susceptible, infected, and removed rates respectively. It is obvious that the probability of spreading a disease in a crowded area is bigger than spreading in an environment consists few people. Hence, population size is an important effect in the spreading process of epidemics [2]. Similarly, in SNs, a post spreads so fast if owner of the post has lots of connections.

In literature, epidemics consists of two models: simple epidemics and complex epidemics [1]. Simple epidemic~~s~~ spreads based on the log of the population size. If the population size is n, then we can measure the spreading of epidemics with $\log n$. In this model, nodes can be either susceptible or infected. On the other hand, in complex epidemics, nodes can be susceptible, infected, or removed. A critic point here is to model the transition from infected state to removed. Researchers first proposed a counter value (ctr) to control this process [1, 3]. The main idea behind this value is that it counts the number of nodes that an infected node affected and it stops spreading when this number reaches to ctr value. The value is determined before spreading process and it is valid for each node in the network. Unsurprisingly, if we choose a big value for ctr, information reaches a bigger portion of network but it requires more rounds to complete spreading process. Eventually, ctr controls the termination of the spreading process and the size of spreading area in a network.

What is more, complex epidemics models come with two sub-models: complex epidemics with static information and complex epidemics with dynamic information. During the whole spreading process, if the information does not have any revision, it belongs to static sub-model. In the contrary case, it becomes a dynamic information and refers to the dynamic sub-model [4]. When we consider this concept in SNs, users may revise someone's post and publish as a new post so information in SNs is dynamic; complex epidemics with dynamic information describes this case in a best way.

B. *Current Information Spreading Models*

Although most of current studies consider 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 provide us to predict how well an information will spread. This section first gives the studies that focus on the adaptation of SIR model and then shows an information spreading model with cascades.

Bao et al. [5] criticize SIR model in terms of the idea behind infected state. They propose that any node which becomes an infected does not have to believe/accept the information; they may also oppose it. 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 as SPNR. According to SPNR model, when an ignorant node takes an information from a positive/negative spreader, then it becomes a new positive/negative spreader with a probability value [5]. In a 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 node), it becomes a stifler also with a probability. They define the transition from a spreader state to removed one with a spreading threshold. Although this model is an enhanced version of SIR model in case of the scope, it does not represent users' behavioral effect on the spreading process. Hence, it is not a complete realistic approach to use in SNs today.

Serrano et al. [6] considers that a node may have a first impression about an information before infected by other nodes. That is why, they modified SIR model with the following four states: (i) neutral (initial state), (ii) infected (believe the rumor), (iii) vaccinated (believe the anti-rumor before being infected) and (iv) cured (believe the anti-rumor after being infected). According to this model, all nodes are ignorant at initial step. Then, they assign some of them as infected. Infected nodes start to infect their ignorant neighbors with a given probability. To simulate cured or vaccinated ones, they define a time as delay and at that time a randomly selected infected nodes start to spread an anti-information which says the opposite of 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 but still it has deficiencies to apply in a SN because it does not include any evaluation of users' behavioral model; it only uses a probabilistic approach to decide.

Behavioral characteristics affect the selected spreading probability value(s) for these models. The decision on probability parameter has critical importance to precisely predict spreading path of an information. The value of probability can change according to communicating members, their common interest area, content of the message etc.

Cordasco et al. [7] consider the infected state of SIR model from a different aspect. They propose that any node 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. Their model consists three states: (i) ignorant, (ii) aware and (iii) spreading. Similar to other models, all nodes are ignorant at the beginning. When an ignorant node takes an information from a spreader, it becomes aware. To be a spreading node, any aware node should take the information from more than a pre-defined number (threshold value) of spreading node. This model has no state for removed but they define a termination rule in the original paper. This model can be easily adapted to SNs if transition process from aware to spreading state also considers users' social behavior analysis.

Tong. et al. [8] shows an information cascade model in social networks. First, they provide an enhance study on the scale of the cascades, scope of the cascade subgraphs and topological attribute of spread tree. Then, based on the evaluation results, they analyze the spread of 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 taking a decision:

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

Experiment results of this study shows that popularity of nodes is an important criterion for information spreading. Although this study evaluates some user-specific parameters to model SNs, it does not use an epidemic approach. We would like to take epidemic model as a baseline and improve it to adapt todays' social network structures with a hybrid model.

III. A Hybrid Information Spreading Model

We propose to develop a hybrid model which considers the models of Bao et al. [5] and Cordasco et al. [7] but modifies their threshold theories by using information cascade characteristics like popularity of users or effect of strong features among users such as gender, education etc. Novelty comes from using such a hybrid model which will also be supported by strong features. Strong features define how well two users are connected so we can infer how strong friendship relation exists between two friends on SN. Those features are so important to observe information spreading together with a cascade model because they affect the behavior of users for deciding whether to spread a specific information or not. By using this approach, we can make more realistic transitions between different states. What is more, we will use the idea of Bao et al. [5] regarding to infected state. Because there is a probability for a user to reject an information, we will also divide our infected state into two: positive infected and negative infected.

Fig. 1 shows the state transitions of our model. The proposed model includes 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 (opposes the information and tries to convince other nodes in this way) and (v) removed (user stops spreading).

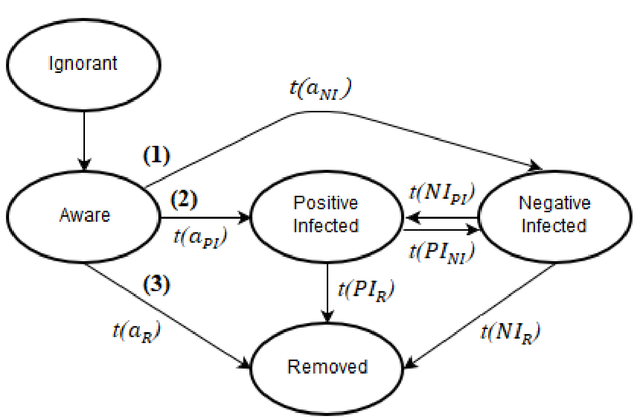

***Figure 1**. A Hybrid Information Spreading Model*

- Initially, we assume 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 advertising or marketing domain it is important to reach most number of users in a short time. Hence, the selection process of initial positive/negative infected users should be performed according to the topology of 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. At this point, there are three different states that an aware may pass: (1) he/she may believe that the information is true and pass to 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 negative infected state via function "$t(a_{NI})$: transition from aware state to negative infected state", and (3) an aware node may prefer not to infect any other node either positively or negatively. In this case, that node may pass directly to the removed state via transition function $t(a_R)$. Functions $t(a_{PI})$ and $t(a_{NI})$ includes the effect of cascading mechanisms, strong features, social behavior analysis, content analysis of the information etc. 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". Besides, 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". Transition functions to pass removed state will use a threshold value which based on the threshold theorem of SIR model but modified with some additional parameters of today's dynamic structure such as users' attention on the content of information, source of the information etc. Hence, this threshold will be associated with the parameters of transition function and it will be user-specific.

Consequently, we will base our hybrid model on the modified version of basis SIR model and generate a new formulation by also using users' social behavior analysis and content analysis of the information. To verify our model, we will implement both referred models [5, 7] and our proposed model in a real SN dataset to observe the results and then we will compare the success and failure rates of these three models.

IV. 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 with today's dynamics. Because users' decisions on spreading any information depend also on social behavioral factors, we will include behavioral analysis on 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 SN. Additionally, this model can be used for interaction analysis among SN users.

REFERENCES

[1] Alan Demers, Dan Greene, Carl Hauser, Wes Irish, John Larson, Scott Shenker, Howard Sturgis, Dan Swinehart, and Doug Terry. 1987. 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, 1-12. DOI=10.1145/41840.41841 http://doi.acm.org/10.1145/41840.41841.

[2] D. J. DALEY AND D. G. KENDALL (1965), Stochastic Rumours, IMA J Appl Math (1965) 1 (1): 42-55. doi: 10.1093/imamat/1.1.42

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

[4] Yichao Zhang, Shi Zhou, Zhongzhi Zhang, Jihong Guan and Shuigeng Zhou. 2013. Rumor Evolution in Social Networks. Physical Review E, vol. 87, no. 3, Article ID 032133, 2013.

[5] Yuanyuan Bao, Chengqi Yi, Yibo Xue, and Yingfei Dong. 2013. 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, 1472-1473. DOI=http://dx.doi.org/10.1145/2492517.2492599.

[6] Emilio Serrano, Carlos Ángel Iglesias, and Mercedes Garijo. 2015. 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, 811-814. DOI: http://dx.doi.org/10.1145/2740908.2742466.

[7] Gennaro Cordasco, Luisa Gargano, Adele A. Rescigno, and Ugo Vaccaro. 2016. 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, 435-437. DOI: http://dx.doi.org/10.1145/2933057.2933069.

[8] Chao Tong, Wenbo He, Jianwei Niu and Zhongyu Xie. 2015. A novel information cascade model in online social networks. Physica A: Statistical Mechanics and its Applications, Volume 444, p. 297-310. DOI:10.1016/j.physa.2015.10.026.