Irregular Cellular Automata Based Diffusion Model for Influence Maximization

Mohammad Mehdi Daliri Khomami

Department of Electronic and Computer Engineering, Islamic Azad University, Qazvin branch, Qazvin, Iran
Daliri.mojtaba@gmail.com

Alireza Rezvanian

Department of Computer Engineering & IT, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran
a.rezvanian@aut.ac.ir

Negin Bagherpour

Department of mathematical Sciences, Sharif University of Technology, Tehran,Iran
nbagherpour@mehr.sharif.ir

Mohammad Reza Meybodi

Department of Computer Engineering & IT, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran
mmeybodi@aut.ac.ir

Abstract—**Due to great communication among users in social networks, a lot of attention is paid to the spreading of information. This issue is of a huge consideration in modern viral marketing either. So far, different models have been proposed in many of which active and inactive users are cooperating in the simple form. Since the influence of individuals in spreading of information happens differently in the real world, in this article we propose a multi-state model for information spread based on cellular automata. We used different states for the proposed model as well as various levels of influence from the beginning up to the end. As an evaluation, proposed model not only has been examined with standard data corresponding to different social networks, but also has been compared with different thresholds. The results of simulations show the superiority of proposed model in comparison with linear threshold model.**

Keywords—Social networks; information diffusion; Cellular Automata;linear threashold

I. Introduction

Recent years observe the increasing popularity of online social networks, where users are connected together by various social interactions. Online social networks enable appropriate information diffusion and marketing campaign by allowing ideas and behaviors to fast propagate along social connections in the word-of-mouth or viral marketing manners. Many companies have made attempts to promote and expansion their products through online social networks by starting campaigns similar to viral marketing [1] and utilizes the social influence of people to their friends [2] In deciding whether to adopt an innovation (such as a news or product), people will be influenced, explicitly or implicitly, by their social friends with different activation thresholds. For example, consider that a new software was devolved and promoted between potential users. The company can advertise it can advertise it by suggesting capabilities to some users in network. Let consider the new software is absorbing. After trying the new software, the earlier users would recommend it to their friends through sharing the new computer software through social networks; their friends would probably use software and share the recommendation product to their friends in the same way. This procedure is continued until the new computer software diffusion in social network starting with small number of initial users. In the network concept this process called diffusion and is an important part of what we call viral marketing or influence maximization.

Prosper of viral marketing depends on the social influence, which has been studied in many domains [1].

Influence maximization is fundamental problem for viral marketing. In [3] *Kempe et al* first formulated influence maximization as a discrete optimization problem. Given a social network with users as nodes, edge weights reflecting influence between users, and threshold number k, influence maximization goals is to find *k* nodes in the network, such that by activating these nodes, the expected spread of the influence can be maximized. For this aim two commonly-used influence spread models are suggested: the independent cascade (IC) and linear threshold (LT) models. In [4], it is proved that the influence maximization under both IC and LT model is NP-hard. Kempe et al a greedy algorithm which iteratively selects the nodes with maximum marginal gain and adds to seed set until to obtained k seed nodes. We note that the computing exact marginal gain with both LT and IC model is NP-hard. In order to estimate the marginal gain used Monte Carlo simulation and reaches a 1-1/e approximate coefficient of the optimal solution for influence maximization problem. In [5] the best approximation guarantee one can hope for is proved.

There are also many recent scholars target at addressing influence maximization problem. In [4] proposed a greedy algorithm to solve the influence maximization problem. But the proposed algorithm suffers from two aspects. First one is the need to assess many candidate nodes before selecting a new seed in each iteration. The second issue is about the calculation of influence spread of any seed set which relies on Monte-Carlo simulations [5] and is time consuming. Moreover, to improve the performance of the greedy algorithm presented in [4], it is noticed that not all remaining nodes need to be evaluated in each iteration and proposed Cost-Effective Lazy Forward (CLEF) algorithm. Hence, the

experimental results show that CLEF leads to appropriate speed up in term of selecting seed node.

In [6], a Degree Discount algorithm is proposed which estimate the expected number of additional vertices influenced by adding node v in the seed set based on v's one-hop neighborhoods. In addition assume that the influence probability is identical on all edge. In [7], proposed an approximation algorithm, LDAG with the aim of finding maximum influence set based Linear Threshold (LT) models. In LDAG, it was proved that computing influence spreading a DAG has linear time complexity, and a heuristic on local DAG construction provides further reduction in the execution time. Java et al. [8] used the basic Linear Threshold Model proposed by Kempe et al. in order to select an influential set of bloggers to maximize the spread of information on the blogosphere. Kimura and Saito in [9] proposed the shortest-path based on the influence cascade model and provided efficient algorithms for computing the spread of influence under this model. All algorithms introduced in the literature depend heavily on the diffusion model. In other words, the reduction of time in diffusion model is the best way to enhance the accuracy and scalability of the algorithms.

Another challenge in diffusion model is how to simulate the diffusion process in more realistic manner in social networks. The primary model was devoted for the Bass model which is used to simulate the adoptions of innovation. In this paper, we make a contribution in two aspects: how to resolve the diffusion model so as to simulate more realistic diffusion process and second how to accelerate the diffusion process in scalable viral marketing applications.

The rest of paper is organized as follows: Section II dedicated to brief introduction of Cellular automata. Then, our proposed model is described in Section III. Next, our experimental results are reported in Section IV. Finally the paper is concluded in Section V.

II. CELLULAR AUTOMATA

Cellular automata (CA) are a model for investigate the behavior of complex systems and have many applications in image processing, sensor networks and complex system. CA is a non-linear dynamical system in which space and time are discrete. It is called cellular because it is made up of cells like points in a lattice or like squares of checker boards, and it is called automata because it follows a simple rule in order to solve complicated problem [10]. The simple components act together to produce complicated patterns of behavior. Cellular automata perform complex computations with a high degree of efficiency, robustness and parallelism. They are especially suitable for modeling natural systems that can be described as massive collections of simple objects interacting locally with each other [11, 12] Informally, a d-dimensional CA consists of an infinite d-dimensional lattice of equal cells. Each cell can assume a state from a finite set of states. The cells update their states synchronously on discrete steps according to a local rule. The new state of each cell depends on the previous states of a set of cells, including the cell itself,

and constitutes its neighborhood [12]. The state of all cells in the lattice is described by a configuration. A configuration can be described as the state of the whole lattice. The rule and the initial configuration of the CA specify the evolution of CA that tells how each configuration is changed in one step. Figure 1 Demonstrate the two dimensional cellular automata.
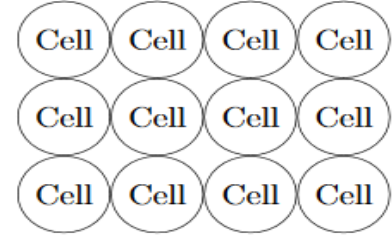

Figure 1. A 2-dimential cellular automata

In this paper, we used irregular cellular automata as a directed graph, in which every vertex consists of one cell of ICA. In the following we described proposed algorithm based on irregular cellular automata.

III. THE PROPOSED MODEL

The proposed model is composed of four states, which is corresponding to impact of nodes in the network. The essential characteristic of the proposed model is different activation level of people in the process of diffusion in the network. In figure 2 we present a new model in which shows the diffusion process in realistic manner. As you can see in Figure 2, the proposed model consists of four states of this model: inactive, active, follower, and adopt. The idea behind of the proposed model is that different people can have different level of influence in the world. Let consider that a corporation, which promotes its product free discount among a group of people as a seed nodes. As a result, it expects these people advertise the product based on their own experiences. The important thing in this case is the impact of the product on the customers. There are some people who acquire great knowledge about the product; however, sometimes they are not convinced enough to by the product. This group of people is called active users. On the other hand, there is another group of people who have the information but need the more information or encouragement for buying the product. This kind of people in the proposed model called followers. On the third case, we have people who have the tendency will buy the product, we name this type of people adapted ones.
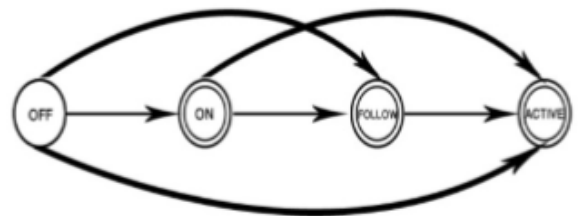

Fig 2. The proposed diffusion model with final state

Given a directed and weighted network G=(V,E) where V, E represents a set of nodes and edges in the networks. For each

edge (u,v) there is corresponding value of influence representing the probability that user v would be affected by user u. The network diffusion model inputs an initial set of users as seeds and outputs the final number of influenced users in the entire networks. The CA based diffusion model inputs an initial set of users as seeds set and outputs the final number of affected users as an output in the network.

The cellular automata based network diffusion model can be describe as a 6-tuple where S is the set of all states which is shown in figure 2. Here, $S = \{\alpha, \beta, \gamma, \delta\}$ in which $0 < \alpha < \beta < \gamma < \delta < 1$ meaning that if a node is chosen between 0 and β, it remains inactive, if it is greater than α and less than γ it is activated, if it is grather than β and less than δ it is follower and otherwise it is adopted. As the time goes on, the influence process is continued based on discrete time in the networks and nodes changing their states. Moreover, $A = S^n$ is defined as the union of all states of nodes in time n and initially is equal to zero for all nodes. We note that $A_0 = \{0, 0, 0, ..., 0\}$ for all nodes. When $A_t = A_{t-1}$ the model will stop, since no more changes of the state occur. N stands for the set of all cells, also nodes in the social network in our proposed model. For each selected node, we define its in-neighbors in the form of a set. For instance, in the set of input nodes for v_1, the set of v'_1 in-neighbors $\Psi(v_1)=\{v_2, v_3, v_9...\}$. N is the set of all nodes' in-neighbors $N=\{\Psi(v_1), \Psi(v_2), \Psi(v_3) ...\}$. *Sum* is the total number of active nodes at some time t when $A_t=A_{t-1}$. R is the local rule of the proposed cellular automaton based diffusion model and is defined as bellows:

$$S_{t+1} = \sum_{j \in \Psi(v_i)} (S_t(j) F_t(v_{i,j}) - T(\Delta t)) \qquad (2)$$

As can be noticing in (2) of state update, cell i's state can be updated based on the influence received by the directed neighbors. $F_{i,j}$ is used for the influence that node j apply on node i. Whenever the influence that a particular node received from its' in-neighbors exceeds the pre-define threshold, the node's state is changed and remains until all of its out-neighbors changing their states. Moreover, $T(\Delta t)$ in which introduced in Formula 3 is presented is a non-increasing function that to simulate the influence would be decrease as time goes on. In other words, a user would most probably accept the recommendation from it friends at the time recommend information is received. Otherwise, the acceptance of the recommendation information is only lower should no other new friends influence.

$$T(\Delta t) = F(1 - \frac{1}{e^{-\Delta t}}) \qquad (3)$$

Where F is the value of influence in networks. We note that, according to the Formula 3 the reduction of influence decreases with the time following the decreasing factor.

Based on the proposed diffusion model, we present the Cellular Automata based algorithm for network diffusion in the social networks. The pseudo code of the algorithm is presented in Figure 3.

Algorithm1: **Cellular Automata Based Algorithm for information diffusion**

Input
    G=(V,E) be the directed and weighted network
    N=|V| // be the number of vertices
    K=|S| // be the number of Seed Set

Output
    Number of activated nodes according to proposed model

Pseudo-Code:
    Let t be number of iteration and initial set to 1;
    Let set A be the state of nodes during the execution of model alongside of algorithm and initial set to off
    Begin Algorithm
    For i=1 to n **Do**
      For j=1 to n **Do**
        If((v_i, v_j))$\in$ **E**)
        $\Psi(v_j) \leftarrow v_i$
        End If
      End For
    End for
    Repeat
      If (v_j state is equal Off or v_j. State is equal On or v_j. State
      is equal Follow) **Then**
      Automata v_j compute their state according to Formula (2)
      If(v_j state is changed) **Then**
        v_j state and A_t are updated
      End If
      End If
      t $\leftarrow$ t+1
    Until A_{t-1} Equal **to** A_t
    End Algorithm

Figure 3: The Pseudo code for our Proposed Algorithm

IV. EXPERIMENTS

In order to establish the performance of the proposed algorithm, we conducted experiments on four types of real world directed networks, including squeak foundation [13], Robots [14], Wiki-Vote [15] and P2P-Gnutella [16]. All of these networks are commonly used benchmarks in simulating information diffusion in the networks. In addition, we assume that all datasets are weighted and directed for influence maximization, in which the weight of the edge (u, v) is 1/deg(v), where deg(v) is the in-degree of node v. Table 1 shows the characteristics of the networks that were used in the experiments.

Table 1. Real networks used in simulations.

Network	Nodes	Edge	Description
Squeak foundation	697	5009	Web community site with different levels
Robots	1698	3492	Share Robotics Software with squeakfoundation
Wiki-Vote	7115	103689	free encyclopedia community
Gnutella	8717	31525	A P2P sharing networks

In order to specific the activation threshold in different states, we used a constant value for different states in the proposed model. For this goal, the values 0.5, 0.8 and 1 are considered for states On, Follower and Adopted respectively. In addition, for each of the four datasets, we select different seed size by varying the parameter k which is presented as a percentage of number of selected nodes in the network randomly. The results is reported in terms of mean and standard deviation $\mu \pm \delta$ with the average of 1000 runs for both the proposed model and the LT model. We note that time was recorded in microsecond when new active nodes appeared. All experiments with the algorithm were launched on a system with a hardware configuration of Intel® Core i5 2.67 GHz and 4 GB RAM.

To evaluate the performance of the proposed algorithm, we conducted a number of experiments in terms of coverage rate and execution time on four real networks. As can be seen from the results of the two models, both the proposed diffusion model and the LT model's active nodes increased with time. Moreover, the performance of LT model is better in the small activated nodes while in all dataset we observe that by increasing the initial activated nodes at the end of diffusion process is increased due to involving diffusion effects and changing the activity level of influence in the networks. In addition, by increasing the initial seed size, because more active nodes can be trigger, the more cells would be put in the queue in the model, hence we observe that the running time of the proposed algorithm is increased. The best result for each networks are highlighted in boldface.

Table 3. The obtained Activated nodes by Algorithm 1 for different initial seed set in terms of the average and standard deviation values for Squeak foundation network

	MVICA		LTM	
Seed Size	Time	Coverage Rate	Time	Coverage Rate
35	2.7±0.4	**58.3±9.7**	2.7±0.6	56.3±7.9
70	3.03±0.4	**138.1±34.6**	2.9±0.4	112.2±19.2
105	4.1±0.3	**220.8±70.0**	3.3±0.5	202.1±25.5
140	5.5±0.6	**314.4±69.5**	5.2±0.4	270.7±79.1
175	6.4±0.5	**349.2±76.8**	5.4±0.6	345.2±84.7

Table 4. The obtained Activated nodes by Algorithm 1 for different initial seed set in terms of the average and standard deviation values for Robot network

	MVICA		LTM	
Seed Size	Time	Coverage Rate	Time	Coverage Rate
85	2.7±0.6	**125.1±4.6**	2.8±0.6	108.5±6.9
170	3.3±0.4	**264.2±4.9**	3.3±0.4	261.3±6.3
255	3.3±0.4	**416.8±7.6**	3.1±0.7	351±5.7
340	3.4±0.5	**523.4±9.5**	3.5±0.5	507.9±7.1
425	3.5±0.5	**652.7±7.8**	3.7±0.4	621.4±7.2

Table 5. The obtained Activated nodes by Algorithm 1 for different initial seed set in terms of the average and standard deviation values for Wiki-vote network

	MVICA		LTM	
Seed Size	Time	Coverage Rate	Time	Coverage Rate
305	2.7±0.2	309.3±2.4	2.1±0.3	**310.5±2.0**
611	2.07±0.2	620.9±3.5	2.17±0.3	**621.8±3.0**
915	2.13±0.3	934.3±3.4	2.27±0.4	**938.2±14.3**
1224	2.23±0.4	**1256.0±24.4**	2.4±0.5	1254±4.3
1530	2.38±0.4	**1574.7±8.4**	2.47±0.5	1572.3±8.4

Table 6. The obtained Activated nodes by Algorithm 1 for different initial seed set in terms of the average and standard deviation values for Gnutella network

	MVICA		LTM	
Seed Size	Time	Coverage Rate	Time	Coverage Rate
436	7.2±1.3	**980.6±5.2**	6.97±2.2	961.4±8.0
872	7.08±1.2	**2165.3±6.2**	9.97±2.0	2108.4±8.2
1308	8.5±0.9	**3626.6±9.2**	12.7±2.4	3505.6±14.3
1717	8.4±0.8	5153.7±9.4	14.1±0.5	**6620.0±8.1**
2126	8.1±0.9	6768.4±5.2	18.7±3.06	**8305.3±10.4**

In this experiment we goal to compare the convergence behavior of the proposed model and LT model. In order to perform this experiment, we plot influence spread versus number of iterations. The results of this experiments for each of the four networks are given in figure 4 and 5 separately. Moreover, for the sake of clarity, we only presented the reported results are the average taken over 30 runs for the both proposed algorithms and LT models. According to the results in figure 3 and 4, we may conclude that in initial time steps, the influence spread increases sharply; however, as close as the convergence is the trend becomes smoother.
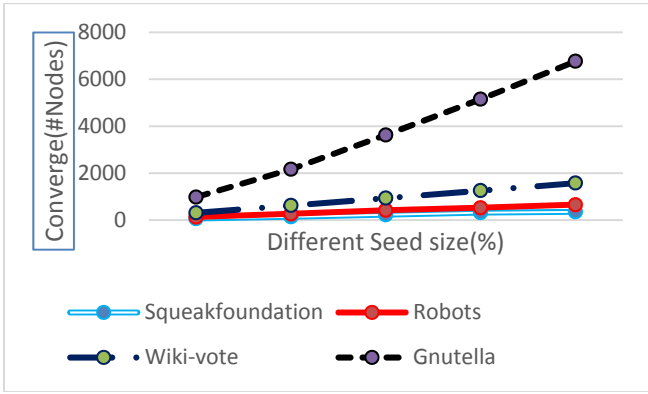
Figure 4: Convergence behavior of the proposed model in terms of the influence spread with different initial active nodes.
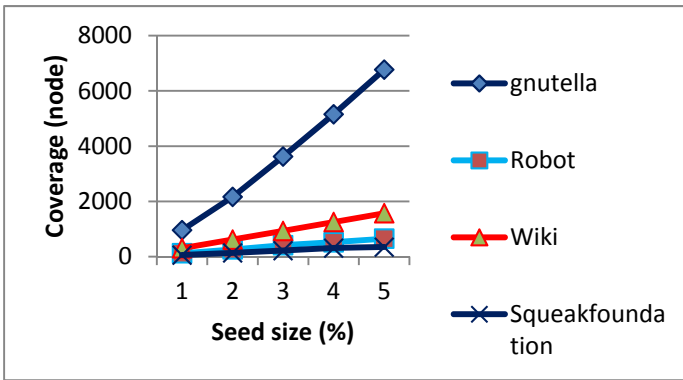

Figure 5: Convergence behavior of the proposed model versus iteration number

The proposed algorithm obtains greater spread influence in some of the cases including Wikivotes and Robots. We note that in all cases the proposed algorithm is very close to the best obtained performance.

V. CONCLUSION

Information diffusion on social networks is considered as one of the most significant challenge on viral marketing. One of the fundamental problems in information diffusion is presenting a realistic model which simulates the diffusion in real world networks. It is even an essential topic for analyzing social networks. In this paper a realistic diffusion model based on irregular cellular automata have been proposed. On the proposed model, different states are considered to show the different level of users that are influenced and the influence threshold is being changed based on states of each neighbors design. Different experiments have also been performed on standard social networks to show the superiority of the proposed model. From the experiments results obtained upon four real life social network datasets, it can be concluded that our model has enhanced the Coverage rate and execution time in the real world networks compared to the LT model. Other advantages of the proposed model includes

its closeness to real world, scalability and high speed of convergence.

VI. REFERENCES

[1] S. Banerjee, H. Al-Qaheri, and A. E. Hassanien, "Mining Social Networks for Viral Marketing Using Fuzzy Logic," in *2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation*, 2010, pp. 24-28.

[2] S.-J. Salvy, D. Jarrin, R. Paluch, N. Irfan, and P. Pliner, "Effects of social influence on eating in couples, friends and strangers," *Appetite,* vol. 49, pp. 92-99, 7// 2007.

[3] D. Kempe, J. Kleinberg, #201, and v. Tardos, "Maximizing the spread of influence through a social network," presented at the Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, D.C., 2003.

[4] A. Goyal, W. Lu, and L. V. S. Lakshmanan, "CELF++: optimizing the greedy algorithm for influence maximization in social networks," presented at the Proceedings of the 20th international conference companion on World wide web, Hyderabad, India, 2011.

[5] H. Nguyen and R. Zheng, "On Budgeted Influence Maximization in Social Networks," *IEEE Journal on Selected Areas in Communications,* vol. 31, pp. 1084-1094, 2013.

[6] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," presented at the Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, DC, USA, 2010.

[7] W. Chen, Y. Yuan, and L. Zhang, "Scalable Influence Maximization in Social Networks under the Linear Threshold Model," in *2010 IEEE International Conference on Data Mining*, 2010, pp. 88-97.

[8] A. Java, P. Kolari, T. Finin, and T. Oates, "Modeling the spread of influence on the blogosphere," in *Proceedings of the 15th international world wide web conference*, 2006, pp. 22-26.

[9] H. Kim and E. Yoneki, "Influential Neighbours Selection for Information Diffusion in Online Social Networks," in *2012 21st International Conference on Computer Communications and Networks (ICCCN)*, 2012, pp. 1-7.

[10] E. Fredkin, "An informational process based on reversible universal cellular automata," *Physica D: Nonlinear Phenomena,* vol. 45, pp. 254-270, 9/2/ 1990.

[11] M. Mitchell, "Computation in cellular automata: A selected review," *Nonstandard Computation,* pp. 95-140, 1996.

[12]   N. H. Packard and S. Wolfram, "Two-dimensional cellular automata," *Journal of Statistical physics,* vol. 38, pp. 901-946, 1985.

[13]   E. R. Murphy-Hill and A. P. Black, "Traits: experience with a language feature," presented at the Companion to the 19th annual ACM SIGPLAN conference on Object-oriented programming systems, languages, and applications, Vancouver, BC, CANADA, 2004.

[14]   R. A. Brooks and M. J. Mataric, "Real robots, real learning problems," in *Robot learning*, ed: Springer, 1993, pp. 193-213.

[15]   N. Ruan, R. Jin, and Y. Huang, "Distance preserving graph simplification," in *2011 IEEE 11th International Conference on Data Mining*, 2011, pp. 1200-1205.

[16]   K. Aberer, M. Punceva, M. Hauswirth, and R. Schmidt, "Improving data access in p2p systems," *IEEE Internet Computing,* vol. 6, p. 58, 2002.