Providing a Model to Explain the Communication Impact on the Way of Information Dissemination in the Social Networks

Sharmin Rezaiy, Rassoul Roustaei
Department of computer, Islamic Azad University, Malayer Branch, Malayer, Iran
Corresponding Author Email: rezaiyuni@gmail.com

***Abstract*—The online social networks have attracted a wide range of users by providing different services and goals. The way in which the social network users are modeling and disseminate the information in these social networks would be so important, considering the impact of communication, and due to the structure of social relationships among users, while there are many kinds of information in these social networks. Many different models of dissemination have been proposed so far, while they have been used in a simple mode as active and inactive users, by considering that the way in which people being affected by information dissemination in the real world would take place step by step. Each node of the graph has been equipped with a learning automat by using an algorithm based on Cellular Learning Automat in this article, and the the graph nodes as well as the relationships between the whole graph, and the reward and fining process of learning automat actions would apply based on the neighborhood relationships. Some tests have been arranged on social networking standards in order to evaluate the efficiency of suggested algorithm, and it has been compared to various parameters. The Simulation results indicate the relative superiority of the proposed model in comparison to the model of cellular automat dissemination based on the nature of the learning of cellular learning automata and based on the receive feedback from the environment too.**

***Index Terms*—Cellular Learning Automata, Dissemination development, Dissemination model, Social Network.**

I. NTRODUCTION

THE development of web based social networks in recent years has led to providing the valuable collections of knowledge and information broadly, so it has caused to see many efforts in applying these rich information collections. That's why transferring information based on communication and the impact of individuals on each other in these social networks would be important and valuable.

The communication between users in social networks would get more attention in the studies which have done in this area, and they tend to explore complex networks in recent decade which are checking the network structure and its impact on the events in the network. Some properties and features like the willingness of users to share information on social networks, get the researcher's attention. The expand optimization has been suggested for the first time by Domingos and Richardson with some models which some people would be selected at first in them, and behavior (or ads) spreads between them, then the community is given up so that the behavior or the idea of selected people to be developed based on their impact on their friends in the community [1] Then Domingos and the colleagues [2] checked the Distribution Marketing on the data mining aspect. They put forward the basis of algorithm's subject for such social processes as the maximization of dissemination by the aim of designing Viral Marketing Strategies. Kempe and the colleagues [3] considered this issue in several extensive study models in social networks analysis. They present provable approximate guarantee for efficient algorithms at first.

Sharmin Rezaiy, The Computer Group of Islamic Azad University, Malayer Branch, Malayer, Iran (e-mail rezaiyuni@gmail.com)
Rassoul Roustaei, The Computer Group of Islamic Azad University, Malayer Branch, Malayer, Iran (e-mail Rassoulroustaei@yahoo.com)

https: IJCSWN.com
ISSN 2519-0814

They formulated this issue and proved that the subject of optimizing the selection of the most influential knot, would be an NP_hard type, and they suggested a greedy algorithm with approximate coefficient of 1- 1/e. The adopted method by Kempe and his colleagues had a high a lot of calculation details. Leskovec and the colleagues [4] have suggested the CELF algorithm by using the sub-module feature to solve this problem. The CELF algorithm decreases the number of times which an algorithm is going to run the dissemination model, but Chen Wei and the colleagues [5] showed that selecting a small collection of the most influential nodes in a social network would take a lot of time. They started to study the most penetration in Linear threshold model, and it would be one of the most important models which is the behavior in social networks has been formulated there. They showed that the exact calculations of penetration in networks in the Linear threshold model would be is generally the "P_ hard" type, and it would be near to the examined issues related to the maximum penetration which is explained by Kempe and the colleagues and it would be faster than the greedy algorithm which was proposed by them. Forest Stonedahl and his colleagues [6] used the physical criteria of the network to arrange initial nodes. They have used a genetic algorithm for searching and selecting the optimal combination of physical properties of the network. The results show that a simple strategy is close to the optimal mode, but a more diverse strategy is working better in experimental networks. They also found a relationship between the desirable seed's budget for a network and the inequality of distribution degree. Yin and the colleagues [7] have suggested a fantastic algorithm based on the genetic algorithm for selecting the first nodes.

Selecting a fixed number of primary users between the entire population with the aim of maximizing profits as a discrete approximation problem would be open for a long time. The issues have been attacked successfully by smart implementing of some algorithms like PSO, DE, GA in this article and they reduced the network's scalability greatly through analysis and data sampling according to the feature of power rule and based on the great search space. Eftekhar analyzed the dissemination on social networks by using the theory of games, and he helped to the speed and accuracy of responds as well as increasing the emission wave by creating coherent groups on the network. So he offered a way in which he could classify the people inside the network by using the ideas of Rock algorithm and their compatibility for directional networks using the information of primary raw groups [8]. Maadi and Tawali made a comparison in their studies between independent dissemination, threshold size, and evolutionary threshold size for dissemination of information on social networks and viral marketing. They define the parameters of each one completely, and comprised them together [9]. After that, Daliri and Khmami and some other colleagues used the irregular cellular automata's idea to present their model. They considered different modes from dissemination to user acceptance as a condition of irregular cellular automata cells. Their simulation results indicate the relative superiority of the proposed model in comparison to the linear threshold release model [10]. A new model has been presented in this article for dissemination which is based on cellular learning automata while different values of the impact threshold would be considered for each user to get closer to the reality. In fact, the purpose of this article is to provide a realistic model on dissemination process at first, so that it can simulate the dissemination process in the real world, and then make results from accelerating the dissemination process. A dissemination model is suggested using cellular learning automata which all nodes of the network would be considered as a cell in this mode. Also, all the cells would change their state based on the designed rules in the specified states. The proposed model works in adaptation to the dissemination process in the real world.

II. Cellular Learning Automata

Cellular Learning Automata is composed of the components which are in the neighborhood of each other, and the behavior of these components would be defined based on the behavior of neighbors and their experiences and neighbors. We need to know these two models of learning automata and cellular automata initially, to get familiar with this model and after recognizing these two models then we will put forward the Cellular Learning Automata which is a combination of two above mentioned models.

A. Cellular Automata

Cellular automata would be simple discrete systems which can reveal their complicated computing and behavior by simple and local rules. Locality means that the cells which are in a neighborhood would be so effective in determining the new value of each cell, while the distant cells have no effects. Each cell has a set of states for itself who decides at any moment which state it needs to go according to the state of itself or the neighbors. The rules of changing the state in cellular automata would be constant throughout the work and they will not change. One of the problems of using the cellular automata is designing the rules to do what we want. The space and time are going in a discrete way, and the automata is homogeneous and the updating operation is performed synchronously, while updating the states can also take place in an un-synchronous form; it means we start to update from one corner of the network and go to the end, while all cells are updated according to the previous state of the neighbors in the synchronization mode. [11]

https: IJCSWN.com
ISSN 2519-0814

B. Learning Automata

Learning automation is an agent which is designed to be located in a possible and uncertain environment. Every Learning Automata has a vector of possibilities, which shows the possibility of doing each action. Each action which is performed by the automata, would be evaluated as a chosen action by a probable environment and the result of the evaluation is given to the automata in the form of a positive or negative signal and the automata would be influenced by this response for choosing the next action. The purpose is that the automata can choose its best action from its own actions, while the best action is the one which maximizes the possibility of receiving bonuses from the environment. [11]

A single learning automata does not have much efficiency, but if there are a lot of learning automata in neighboring and collaborating together they can solve difficult problems. As we said before, it is difficult to design constant rules for cellular automata, and it would be so difficult to imagine cellular automata behavior without the simulation. Combining cellular automat with learning automat in cellular learning automat can solve this problem to some extent.

The space of problem would be as a directed graph while each vertex in this graph (each person in the network), is considered as a cellular learning automat cell. Each learning automata selects an action from its set of actions in the cellular learning automat at any moment. This selection can be based on previous observations or it can be random selection. The selected action is rewarded or fined according to actions selected by neighbor cells and even based on the existing rule in the cellular learning automat.

The automat would correct its behavior and the internal structure of the automation is synchronized according to whether selected act is rewarded or fined. Every automat in the cellular learning automat selects again an action from its set of actions after updating. The selection process of an action and giving a bonus or a fine is continued until the system reaches a steady state or a predefined benchmark is established. The updating of the structure of the automata in the cellular learning automat would be carried out by the learning algorithm.

III. THE PROPOSED MODEL

Finding the most influential set of nudes would be one of the most important questions which has been put forward on social networks. So the problem is defined such as finding the collection by having a constant value which finally effects on the network's nodes so that most people of that collection adopt a new strategy. The proposed model consists of four conditions which indicates the level of impact on nodes in the network. Actually the main approach of this model is the different effects of individuals on each other. There are four conditions in this model which are defined respectively as: Inactive, Active, Follower, and Adapted. The main idea of this model is that one person may has a different impact facing to those who have already been affected.

In fact, they are a group of people who will know the necessary information, but their awareness and willingness are not enough to be encouraged about dissemination, so these type of people would be considered as active members in the model. There are other people who have the necessary information while they need more stimulation to be encouraged about dissemination. There are some other people who would be encouraged about dissemination as soon as they get any information while they are considered as adapted type of people.

The proposed multiple model with the positions of each node would be an oriented graph of G = (V, E), in which V represents the collection of all nodes in the social network, and E stands for all Vertex. There is a probability value for each edge of (u, v), which indicates the user's influence of the u user on the user v. The proposed dissemination model receives a number of primary users as the input, and active users at the end of the dissemination process on the network would be its output.

The S collection would be the collection of all provided cellular states. The M collection would be considered as:

M = {0 - 0·5, 0·5 - 0·6, 0·6 - 0·8, 0·8 - 1} is

So if the threshold value of a cell while running the algorithm is between 0 - 0·5, this node would be inactive. If it is 0·5 and less than 0·6 it means that this node is activated. If it is between 0.6 and 0.8, it would be the follower node and if it is between 0.8 to 1, it would be considered as an adapted node. The nodes accept more and more offers from their neighbors passing the time, while the impacting process is continuing too. We define the input nodes collection, and on the other hand as it was said before, it needs to have a module as learning, and receiving the rewards or fines from the environment in the cellular learning automate. Initially, a number of nodes are considered as active, which have the values more than 0.5, and they are in active status. A function must be considered too which run the rewarding with the appropriate action. So two numbers have been considered for each active node which shows the effect of the node and its neighbors respectively. The sum of these values is always equal to 1, and if an operation is performed which leads to any changes in the state of a node, this weight would be increases for the nodes which are causing this state. In other words, if an action leads to any changes of a node state, the impact coefficients of those thing made these changes will increase, while this increase is equal to 0.05.

IV. THE PROPOSED ALGORITHM

The collection of primary active nodes would be the input of algorithm, and the number of nodes which are activated at the end of the algorithm would be the output of algorithm. We define the time variable equal to zero, at the initial valuation stage and we set a collection as the collection of nodes status when we run the algorithm. In fact, this change of situation would be included in this collection. Initially a number of nodes are considered as randomly activated nodes. Each active node is affected by two factors to change the state; one is the internal factor, which represents the current state of the node, and the other one is external factor, which shows the status of the neighbors of the node. At first, 50% of the cell's own status and 50% of the average position of the neighboring nodes would be affected, and then we need to update the status of all active nodes. It is examined that which factor between the two internal or external factors has the greatest impact on changing the state of the node. In the next step, the cell learning process begins, by considering that the individual or his / her neighbors have had a greater impact on the cell status changes into the better form, and by giving a weighted average percentage of that factor, the dissemination is taking place. The algorithm continues until none of nodes change their status, because if the status wouldn't change, it means the network is stable and balanced.

ALGORITHM I
PROPOSED CELLULAR LEARNING AUTOMATA ALGORITHM

```
clear;
close all;
%% Parameters and Initialization
t= cputime;
N = 10000;            % Number of Population
EoT = 10;             % Effect of time
index = 18;           % Index of Target Person
nSelcted = 10;
M = [0, 0.5, 0.6, 0.8, 1];
per = 0.6;
Nit =100;
Nseed = 1074;

% Getting Raw Dataset from Input
socDataset = xlsread ('soc-delicious.xlsx');
socDataset = socDataset(1:N,:);
trustDataset = xlsread ('edgeEpinion.xlsx');

% Trancate Trust Dataset
uniquePersons = unique(socDataset(:));
trustDataset = trustDataset(ismember( trustDataset(:,1), uniquePersons ),: );
trustDataset = trustDataset(ismember( trustDataset(:,2), uniquePersons ),: );

%Calculating Number of Population
numberOfpopulation = length(uniquePersons);
numberOfrows = length(socDataset);
k = randi(numberOfpopulation,1,Nseed);
state = zeros(1,numberOfpopulation);

state(k) = (M(5)-M(4)).*rand(1,length(k))+M(4);

stateNumber = State(state, M);

stateF = find(stateNumber>1);

% Create an empty Adjanceny Matrix of people relations
Adjancency =zeros(numberOfpopulation,numberOfpopulation);

%% Main
it=1;
flag = false;
cof = 0.5.*ones(2,length(state));
itt=1;
while (it<Nit)
    stateNew = state;
    for i=1:length(stateF)
        friends = find(Adjancency(i,:)==1);
        activeFriends = find(stateNumber(friends)>1);

if(cof(1,stateF(i))*state(i)+cof(2,stateF(i))*mean(state(activeFriends))>state(i))

            stateNew(i) =
state(i)+10*(cof(1,stateF(i))*state(i)+cof(2,stateF(i))*mean(state(find(Adjancenc
y(i,:)==1))));

        end

    end

        stateNumberNew = State(stateNew, M);

        difference = stateNumberNew - stateNumber;

        %award
        cof(1,find(difference)) = cof(1,find(difference))+per;
        cof(2,find(difference)) = cof(2,find(difference))-per;

         if ((length((difference)==0))==length(difference))

            flag = true;
         end

stateNumber = stateNumberNew;
        state = stateNew;
        stateF = find(stateNumber>1);

    it=it+1
end
OFF = length(find(stateNumber==1))
ON = length(find(stateNumber==2))
FOLLOW = length(find(stateNumber==3))
ACTIVE = length(find(stateNumber==4))
```

V. TESTS

The standard data collection for social networks has been used for simulation in this study. So the first data collection is selected from the Epinions social networking data collection. It is a kind of social service website that helps people's shopping. In fact, it is a platform for checking the consumer's ideas which allows to the users to comment, recommend, review, and evaluate the products which they have bought, or they want to buy all over the world. The second data collection is considered from the Delicious social network which is a social bookmarking service (bookmarking the web pages) to search, share, and save the web bookmarks. In the end of 2008, Delicious has announced that over 180 million links are being bookmarked and more than 5.3 million users subscribed there.

TABLE I
PARAMETERS RELATED TO THE VALIDATION OF SUGGESTED METHOD

Scenario Name	Data Collection	The number of Statistical Society
First scenario	Epinions	2732
Second scenario		4297
Third Scenario	Delicious	2732
Fourth Scenario		4297

The aim is to examine the effect of the number of initial nodes on a social network in the first test, so the introduced data has been used for this test. The number of initial nodes are considered randomly from 5% to 25% of the nodes in each network as the initial nodes of the disseminator. The inactive nodes on the network are influenced by the early nodes of the neighborhood over the time and the network's spreading would expand in the proposed dissemination model. This extension continues until there would be a change in network status. The results of the dissemination for the number of first nodes with the statistical population of 2732 in Epinions social network have been reported, and the time which is required for dissemination and the rate of dissemination (covering rate) is provided according to the number of variables of the initial nodes.

TABLE II
The results of comparisons between cellular models and learning cellular models in the Epinions social network

	Cellular		Learning Cellular	
Number of Primary Nodes	Dissemination Time	Coverage Rate	Dissemination Time	Coverage Rate
136	599	10%	432	27%
273	543	20%	444	42%
409	567	29%	433	29%
546	588	37%	476	38%
683	587	45%	490	47%

TABLE III
The results of comparisons between cellular models and learning cellular models in the Epinions social network

	Cellular		Learning Cellular	
Number of Primary Nodes	Dissemination Time	Coverage Rate	Dissemination Time	Coverage Rate
214	601	8%	570	13%
429	610	20%	572	25%
644	612	29%	581	37%
859	602	38%	589	49%
1074	603	44%	586	60%

TABLE IV
The results of comparisons between cellular models and learning cellular models in the Delicious social network

	Cellular		Learning Cellular	
Number of Primary Nodes	Dissemination Time	Coverage Rate	Dissemination Time	Coverage Rate
136	710	10%	695	27%
273	710	20%	692	42%
409	707	29%	693	29%
546	702	39%	692	38%
683	712	44%	699	47%

TABLE V
The results of comparisons between cellular models and learning cellular models in the Delicious social network

	Cellular		Learning Cellular	
Number of Primary Nodes	Dissemination Time	Coverage Rate	Dissemination Time	Coverage Rate
214	740	10%	721	29%
429	766	20%	732	20%
644	752	30%	749	29%
859	732	39%	731	40%
1074	748	49%	736	49%

The charts of activated nodes have been shown in figure respectively and by the percentage of initial nodes for different networks for nodes which are activated under the cellular automat model and the proposed model of cellular learning automat

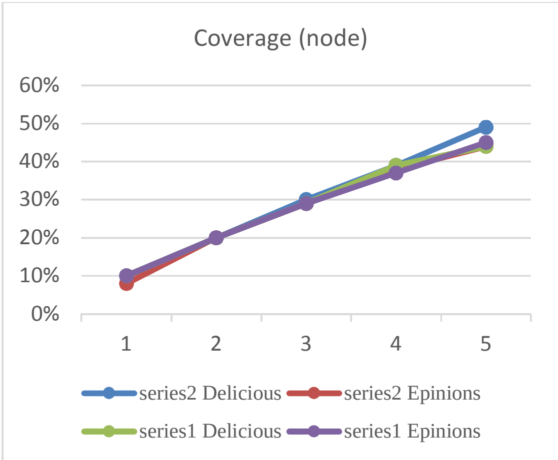

Fig 1. The comparison of the coverage rate of dissemination by percentage of initial activated nodes (5 to 25%) for the cellular automat model

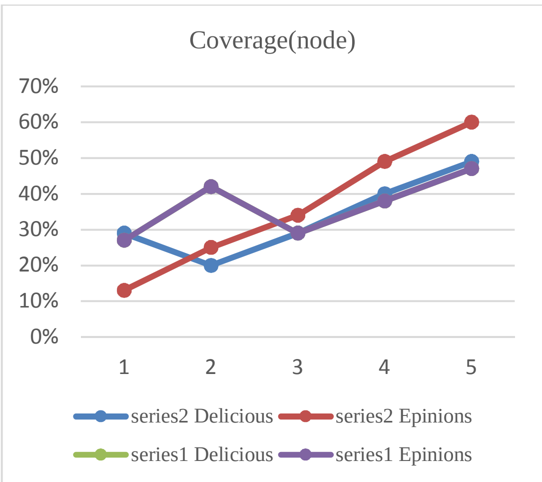

Fig 2. The comparison of the coverage rate of dissemination by percentage of initial activated nodes (5 to 25%) for the cellular learning automat model

https: IJCSWN.com
ISSN 2519-0814

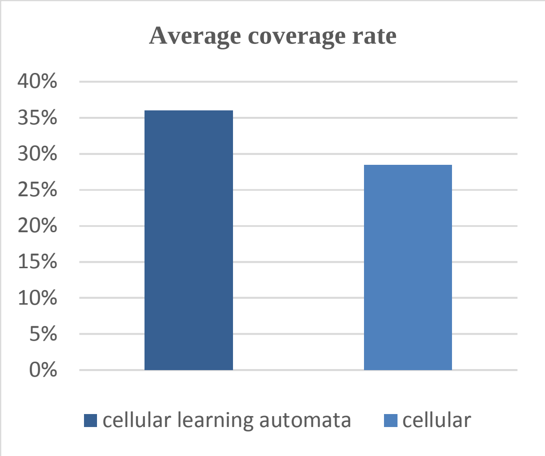

Fig 3. The total average coverage rate on 4 testes

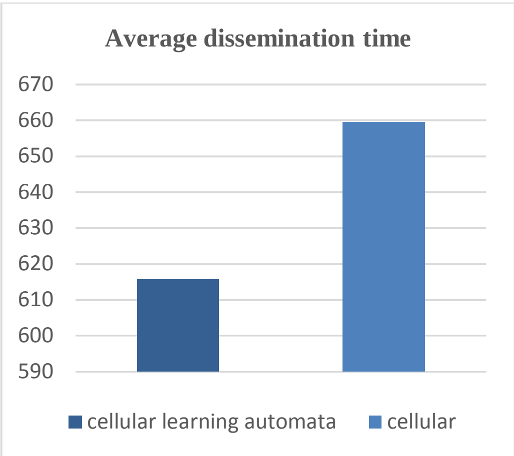

Fig 4. The total average dissemination time on 4 testes

According to the results presented in the tables I-V and shapes 1-4, different results have been obtained for different networks with different features such as the number of initial active nodes. We can see a higher coverage rate than cellular automat in the results of the proposed model and in all cases, except when the number of high initial nodes have been used and when the network is very empty. As it is seen in table, by increasing the number of active nodes the dissemination also increases, but it has less amount in comparison to the cellular automat for the time of dissemination. The simulation of the information dissemination process in the proposed method would be closer to the dissemination process in reality, in comparison to the cellular automat method.

VI. Conclusion

Dissemination is one of the most important issues on the social networks. Providing a near-reality model in this field would be so necessary issue for analyzing the social networks. We made the issue more comprehensive by increasing the speed of dissemination in the network in this study, and we offered the dissemination technique based on cellular automata. Some states of which the users have been influenced in each cell and some rules for changing the mode of each cell have designed in the proposed model. The main aim of this study was to consider different percentages of the statistical society as active users, and to examine the rate of dissemination and the time of dissemination in social networks. There were many tests which were conducted for the proposed model on standard social networks, indicating a relative improvement of the proposed model. The importance of the number of initial nodes in the results would be evident in the performed tests. Also we can point out to the state of being close to the real world, the scalability and high speed of the algorithm from the features of the proposed model.

The results show that while the proposed model sometimes may not provide better responses than cellular automata models, but generally it has had a very good performance in terms of coverage rate and dissemination time.

References

[1] P. Domingos and M. Richardson. Mining the network value of customers. In Proc. 7th International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, 2001, pp. 57–66

[2] M. Richardson and P. Domingos. Mining Knowledge-Sharing Sites for Viral Merketing. In Proc. 8th International Conference on Knowledge Discovery and Data Mining, 2002, pp. 61–70

[3] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence in a social network. In Proc. 9th International Conference on Knowledge Discovery and Data Mining, Washington DC, USA, 2003, pp.137–146.

[4] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. S. Glance. "Cost-effective outbreak detection in networks". In Proceedings of the 13th ACMSIGKDD Conference on Knowledge Discovery and Data Mining, pages 420–429, 2007.

[5] W.Chen, C.Wang, Y.Wang," Scalable influence maximization for prevalent viral marketing in largescale social networks", Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, July 25-28, 2010

[6] F.Stonedahl, W. Rand, U.Wilensky," Evolving viral marketing strategies", Proceedings of GECCO'10 Proceedings of the 12th annual conference on Genetic and evolutionary

[7] Y.Gui-sheng, W.Ji-jie, D. Hong bin, L.Jia , "Intelligent Viral Marketing Algorithm over Online Social Network," Networking and Distributed Computing (ICNDC, pp.319-323, 21-24 Sept. 2011

[8] Eftekhar, M. (2009 - 2010). Dissemination Analysis in Social Networks Based on Games Theory, Master's thesis for Computer Science, Computer Engineering Faculty, Sharif University of Technology, 83 pages

[9] Daliri Khumami, Mohammad Mehdi; Alireza Rezvanian and Mohammadreza Meybodi, 2014; Distribution model based on irregular cellular automata in multiple ways.

https: IJCSWN.com
ISSN 2519-0814

[10] Meybodi, Mohammad Reza, Beigi, Hamid, Taherkhani, Masoud, 2000, Cellular Learning Automata, 6th Iran's Computer Association Conference, pages 153-163

[11] Maadi, Moein and Reza Tavali, 2014, Information dissemination in social networks and viral marketing: A new e-commerce strategy, the first national computer conference, information and communication technology, Marvdasht, Marvdasht Islamic Azad University