Activeness based Propagation Probability Initializer for Finding Information Diffusion in Social Network

Ameya Mithagari[1] and Dr. Radha Shankarmani[2]

Sardar Patel Institute of Technology, Andheri, India
ameya.mithagari@spit.ac.in

Abstract. Information diffusion is the process of spreading information from one node to another over the network. To calculate the information diffusion coverage, it is important to assign propagation probability to every edge in the social network graph. Most popular models of information diffusion use Uniform Activation(UA) and Degree Weighted Activation(DWA) to calculate propagation probabilities. However, the results obtained by these methods are non-realistic. Therefore, we propose a new Activeness based Propagation Probability Initializer (APPI) model to obtain realistic information diffusion. This is achieved by assigning propagation probabilities based on activeness value inferred using topological node behavior. The experimental results shows that APPI provides balanced and meaningful information diffusion coverage when compared with UA and DWA.

Keywords: Diffusion · Cascade · Topological · Social Network .

1 Introduction

Online social networks are rapidly evolving these days, thereby increasing the market base for any new product. The process of spreading of information, ideas, views, product promotions, advertisements, etc. over a network is called Diffusion. For example, information about a product getting viral over the network can be termed as information diffusion in the network [1].

Initially, only a few people accept the product or an idea. As people observe that their neighbors have accepted the new product or an idea, they tend to accept it. As a result, a cascade of information is formed, thereby generating an indirect recommendation system. Hence information diffusion in the social network is studied by many researchers [2].

In this work, a novel model is proposed, named 'Activeness based Propagation Probability Initializer' for a temporal network. A temporal graph is a snapshot of the network taken at regular intervals of time and is represented as a graph. The time period can be monthly, quarterly, semi-annually or annually.

The paper is organized as follows. In the background section, we discuss various methods of information diffusion and the research gap in the algorithms to assign propagation probabilities. The next section consists of a proposed model for assigning propagation probabilities based on activeness. Experimental results and comparison with various methods are discussed in Section 4. Finally, the insights obtained by the research are summarized in the Conclusion.

2 Background

The process of acceptance of a product or an idea is modeled by the following diffusion models:
Liner Threshold model (LT): At every iteration, the node accepts the product or an idea only if the 'sum of the weights of connected edges to its neighbors who have already accepted the product or an idea' crosses a certain 'threshold (t)' [4].
Independent cascade model (ICM): At every iteration, The user (u) accepts the product or an idea with 'Propagation probability ($p_{\mathrm{u,v}}$), every time one of it's neighbor(v) accepts the product or an idea [3].

Propagation probability of an edge is the probability by which one node can affect the other node along the edge. Propagation probability must be assigned to every edge before starting the iteration using ICM. The propagation probability is assigned by the following two universally accepted models [5–7]:
Uniform Activation (UA): Same propagation probability is assigned to all the edges.
Degree Weighted Activation (DWA): Propagation probability assigned to edge(u, v) is the reciprocal of the number of neighbors of node v.

2.1 Research problem

Proper justification to assign propagation probability is not provided by both UA and DWA. Different results can be obtained for different propagation probabilities when UA is considered. In DWA, more the number of neighbors, lesser is the chance of the node accepting the product or an idea. For example, if there is only one neighbor of a node v, then the propagation probability of the edge(u, v) is calculated as 1. This means the node will surely accept the product or an idea. However, this conclusion is incorrect. To obtain realistic information diffusion, we propose a model that uses node topological behavior to initialize propagation probability.

3 Activeness based Propagation Probability Initializer (APPI)

The input to Activeness based Propagation Probability Initializer (APPI) are snapshot of same network taken at different time instances represented as Graph G with vertices V and edges E, $G(V,E)$, $G_1(V_1,E_1)$, $G_2(V_2,E_2)$ to $G_n(V_n,E_n)$. Following are the assumptions to deploy the model:

- Unweighted graphs are considered.
- Deleted accounts are represented as isolated nodes.

3.1 Activeness Value Finder

The two sequential snapshots, previous graph G_{i-1} and current graph G_i where i ranges from 1 to n are considered. Then the 'lists of neighbors' for each node in both the graphs are generated simultaneously. Comparing these lists per each node, below conditions are evaluated for inferring Activeness value of the node based on the node's topological behavior:

- *Increase in the number of neighbors:* Set Activeness value of the node to the total number of newly formed connections of the node.
- *No change in the number of neighbors:*
  – If elements in the list of neighbors changes, then Activeness value is set to the total number of changed connection.
  – If there is no change, Activeness value is set to zero.
- *Decrease in number of neighbors:* If the total number of lost connections is greater than 'average degree of the graph', then Activeness value is set to the total number of lost connections else Activeness value is set to zero because over the specified time range the node may be inactive then the loss of connections are due to other users.
- *New users in the network:* The Activeness value is exaggerated due to newly formed connections. Therefore, to compensate for that, the Activeness value is reduced by 75% and rounded off to the next whole number as they can be active nodes but the trust value won't be developed at the initial stages of the node in the network.

The change of neighbors may be negative or positive but the activeness value is always positive. The activeness value is always a whole number.

3.2 Propagation Probability Initializer

The propagation probability $p_{u,v}$ is assigned to every edge(u,v) in the current graph G_i. Activeness value of both the nodes u and v connected to an edge e is considered in assigning $p_{u,v}$. Below cases are considered to initialize the propagation probability $p_{u,v}$.

- *If the activeness of node u or node v* $\leq$ *1 :*
  The propagation probability $p_{u,v}$ for edge(u,v) is set to 0.01.
- *If the activeness of node u and node v* $>$ *1 :*
  The propagation probability $p_{u,v}$ for edge(u,v) is calculated as below:

$$p_{u,v} = (1 - \frac{1}{Activeness(u)}) * (1 - \frac{1}{Activeness(v)}) - \epsilon \tag{1}$$

  The propagation probability $p_{u,v}$ is directly proportional to activeness of both the nodes u and v. A node may be active or inactive and it may or may not spread the information further hence the propagation probability is 0.5. We define ϵ as the limiting factor ranging from 0.20 to 0.24 because even if we are assuming lower probability range by subtracting ϵ, higher probabilities are obtained only if both the nodes connected to the edge are highly active. Thereby generating an upper bound to the propagation probabilities. Therefore $p_{u,v}$ ranges from 0.01 and tends to 0.8 depending upon the ϵ value.

4 Experimental Results and Discussion

4.1 Implementation

Different snapshot graphs of the same network at different time interval are input to the model. At a time only two graphs are considered: the previous snapshot graph and the current snapshot graph. A page rank algorithm is applied to the current snapshot graph to obtain rank value. Top k nodes based on rank value are selected as the seed set. These are the influential nodes considered for starting the diffusion process using ICM. Before that, the propagation probabilities of edges in the current snapshot graph are assigned. We compare our model (APPI) with the following propagation probability assignment algorithms: Uniform Activation (UA) and Degree Weighted Activation (DWA).
Propagation Probability assignment: A node may be active or inactive and it may or may not spread the information further hence the propagation probability assigned to every edge using UA is 0.5 . According to DWA, reciprocal of the degree of the node is calculated and assigned as propagation probability to incoming edges of that node. The conditions discussed in section 3.1 and 3.2 are incorporated to form an APPI algorithm which assigns propagation probability to every edge in the current snapshot graph. The limiting factor(ϵ) is set to 0.2 .

The information diffusion coverage is considered for evaluation which is obtained by Independent Cascade Model (ICM) after assigning propagation probabilities to the edges in the current snapshot graph by each algorithm separately.

4.2 Results for Synthetic network

Dataset used are two randomly generated synthetic graphs. One of them is considered as previous snapshot graph with 10 nodes and 24 edges and the other as current snapshot graph with 10 nodes and 29 edges. The experiment is conducted according to the procedure in section 4.1 with the above synthetic dataset and seed set size of 2 nodes. The diffusion always starts from the influential nodes found by page rank algorithm which are node '4' and node '9' in current snapshot graph. Green nodes depicts the information diffusion. Results depicted in Fig. 1 shows over-estimation of diffusion when UA is used with ICM. Showcasing higher uncertainty of results. Signifying UA as more of an optimistic approach.

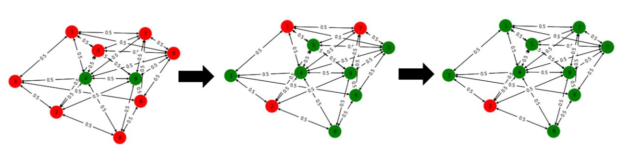

Fig. 1. Information diffusion in different iterations of ICM when UA is used to assign Propagation probabilities.

Results depicted in Fig. 2 shows under-estimation of information diffusion when DWA is used with ICM, signifying DWA as more of a conservative approach.

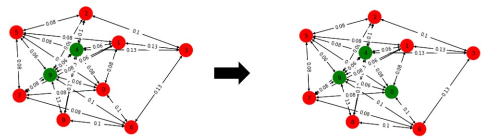

Fig. 2. Information diffusion in different iterations of ICM when DWA is used to assign Propagation probabilities.

Following are the Activeness values of nodes that are calculated using Activeness value finder discussed in section 3.1 : {'0': 2, '1': 3, '2': 0, '3': 0, '4': 5, '5': 3, '6': 3, '7': 5, '8':2, '9': 5}. The Propagation probabilities of edges are calculated using Propagation Probability Initializer discussed in section 3.2. For example, the activeness value of node '5' and node '9' is 3 and 5 respectively therefore, propagation probability $p_{5,9}$ of edge('5','9') is calculated using Equation (1) $p_{5,9} = (1-\frac{1}{3})*(1-\frac{1}{5})-0.2 = 0.33$. Taking activeness into consideration, we are obtaining de-escalated propagation probabilities for active nodes by subtracting ϵ i.e 0.2 . Hence the confidence of result is increased. Results depicted in Fig. 3 shows balanced and meaningful information diffusion when APPI is used with ICM.

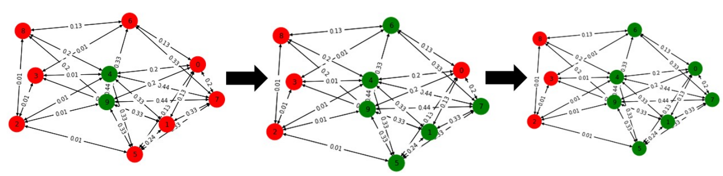

Fig. 3. Information diffusion in different iterations of ICM when our model(APPI) is used to assign Propagation probabilities.

4.3 Real-World network

Real world Facebook data-set archived in The Max Planck institute for software systems [8] is considered. We initially construct the graphs in GraphML format in an interval of 3 months using time-stamp of the activities. Number of nodes in 1^{st} Month, 4^{th} Month, 7^{th} Month, 10^{th} Month, 13^{th} Month, 16^{th} Month, 19^{th} Month and 22^{nd} Month snapshots of network are 9101, 12360, 15632, 20090, 25301, 30994 and 37769 respectively. Similar experiment is conducted according to the procedure in section 4.1 with the facebook dataset and 25 most influential nodes found using page rank as seed set. The results are depicted in Fig. 4.

The results using UA indicates over estimation of Information diffusion whereas the results using DWA indicates under estimation of information diffusion. Results using APPI (proposed model) shows a balanced and meaningful information diffusion in real-world network.

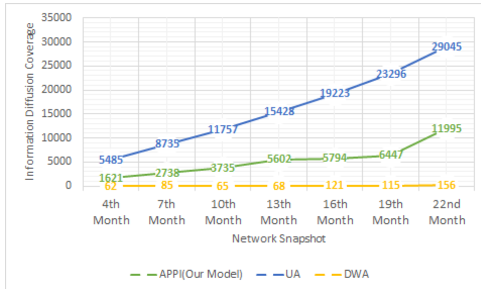

Fig. 4. Information Diffusion comparison results.

5 Conclusion and Future work

Independent Cascade Model (ICM) is generally used for testing information diffusion in Influence Maximization problem. However, the standard algorithms such as UA and DWA used with ICM does not provide justified and realistic results. This problem is efficiently solved using the proposed Activeness based Propagation Probability Initializer (APPI) model. Since this model calculates the propagation probability of every edge in the graph based on the activeness of its connected nodes, the generated results have good confidence. In a nutshell, the APPI model gives balanced and meaningful information diffusion based on topological node behavior which is not the case with earlier existing models. The future work is to calculate the confidence value and support of the propagation probabilities assigned by APPI.

References

1. J. Leskovec, L. A. Adamic, and B. A. Huberman.: The dynamics of viral marketing. In: ACM Transactions on the Web (TWEB), (2007)
2. E. M. Rogers.: Diffusion of Innovations. Free Press, fourth edition, New York (1995)
3. D. Kempe, J.M. Kleinberg, and E. Tardos.: Maximizing the spread of influence through a social network. In: Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining, (2003)
4. M. Granovetter.: Threshold models of collective behavior. In: American Journal of Sociology, (1978)
5. H. Zhuang, Y. Sun, J. Tang, J. Zhang and X. Sun.: Influence Maximization in Dynamic Social Networks. In: IEEE 13th International Conference on Data Mining, (2013)
6. G. Song, Y. Li, X. Chen, X. He and J. Tang.: Influential Node Tracking on Dynamic Social Network: An Interchange Greedy Approach. In: IEEE Transactions on Knowledge and Data Engineering, (2017)
7. W. Li, Q. Bai and M. Zhang.: SIMiner: A Stigmergy-based Model for Mining Influential Nodes in Dynamic Social Networks. In: IEEE Transactions on Big Data, (2018)
8. Social Computing Research Homepage, http://socialnetworks.mpi-sws.org. Last accessed 6 April 2019.