Dynamic Interactions by Strong Influencers in Social Networks Using Opinion Propagation

Saloni Laddha
NIT Hamirpur
Himachal Pradesh
saloni1427@gmail.com

Shrisha Rao
IIIT-Bangalore
Bangalore
shrao@ieee.org

ABSTRACT

Social influence is an essential aspect of human social interaction. Influence propagation modeling has been extensively studied and applied in various research fields, such as the maximization of product adoption, the spread of infectious diseases, etc. The primary sources of information in any social network are the influential nodes that propagate this information further. Our study gives a method to identify the strong influencers in a network and dynamically predict the links/interactions made by them during information propagation. Previously, many efforts have been made to solve the link prediction problem in social networks such as the Adamic Adar index, based on the degree of common connecting neighbors; however, they disregard the fact that social networks can be dynamic over time—any network dataset collected by a snapshot at a particular instant of time may be incomplete and not reflect all the previous links formed. This may in turn result in false predictions. Our approach considers the current dynamics of the social network where information propagated by an agent leads to the prediction of new links with other nodes. Taking opinion propagation as an essential feature, we test our algorithm on Stanford's Facebook network dataset and compare the results with the previous Adamic Adar index. The results obtained cover all the links predicted by Adamic Adar, as well as some new links formed dynamically. This is indicative that considering dynamic link formation is more efficient and helps address the problem of the incompleteness of network datasets reflecting the state at a time instant.

KEYWORDS

Social influence analysis; linear cascade model; strong influencers; link prediction; Adamic Adar index

1 INTRODUCTION

A social network is a social structure and network representation of virtual community i.e. a set of social agents and a set of ties between these agents. A social network is mostly visualized as a graph where the nodes represent agents (individuals, organization, etc.) and the edges (i.e. links) correspond to ties/interactions/relationships between these agents such as friendship, authority, cooperativeness, collaboration, etc. With the boost in the internet community, communication, and cooperation between people becoming more convenient, a tremendous amount of data is generated and stored in the form of social networks with the characteristics of every agent in the social network. Applying analytic techniques to social media data supports better decision making progress. However, the problem faced with the data collected is that social networks are highly dynamic that might lead the miscalculation in the decision making progress. The decision making progress consists of an individual's decision making based on the information received by it from other surrounding nodes. The links and ties of individuals pass some information to the nodes which after being properly summarised the individual decides that information whether it's true or false and which further leads to the passing of that information to other nodes. Also based on the community of the nodes, the individual receives a similar type of information from surrounding leading to change in its decision over that particular community. That decision making can lead to the clustering of nodes and making it join the particular community. Therefore, predicting the future links or ties between these agents in current social networks is not only very important but also useful for the decision-making process. This problem is commonly known as link prediction.

Being one of the link mining and analyzing tasks, link prediction has many important applications[30]. Firstly, it can be applied to recommender systems in information retrieval and e-commerce, which can help people to find new friends and potential collaborators [23, 49], provide interesting items in online shopping [10], recommend patent partners in enterprise social networks and cross-domain partners [42], find experts or co-authors in academic social networks [34, 48], and predict cell phone contacts in large scale communication network. Secondly, it also can be used to infer the complete networks based on partially observed networks [31, 35], understand the evolution of networks better [1, 7, 27] and predict hyper-links in heterogeneous social networks [51]. Finally, the link prediction techniques can also be applied in bioinformatics and biology, for instance, in health care and gene expression networks [3], predicting specialists who are more likely to receive future referrals and finding protein-protein interactions. Even in other domains like security connected domain, it is often used to determine abnormal communications [21].

The communication between the agents is here proposed based on opinion propagation in a social network. These interactions among the users in social networks form the primary influence-diffusion pathway [37]. Influence propagation modeling has been extensively studied and widely applied in many research fields, such as the maximization of product adoption, the contagion of computer viruses, the spread of infectious diseases. The interactions can be regarded as a process of influence and opinions adaptation. These influence propagation can be further utilized in providing better decision making in the link prediction in the network. To explain the influence - diffusion pathway, the principle of influence propagation suggests that the more that people interact with each other, the more similar they become [11].

Among all nodes in a given graph model, it is important and interesting to discover nodes that can affect the behavior of their neighbors and, in turn, all other nodes in a stronger way than the remaining nodes. We call such nodes influential nodes or strong attractors. The primary source of information in any social network is strong influencers with the maximum number of interactions. The characteristics of different social networks can also be identified by looking at the characteristics of strong agents. Strong agents are the agents which facilitate the addition of new members into the social network. These new members will follow the same ideals and share a common opinion of the strong agents. Also, it controls the propagation of most of the information in the social network. These strong agents are based on quantifying the contribution of this agent to increase the size of the network by attracting new active members of the specific sub-community. The more a certain agent attracts new agents, the more that agent is important to the network i.e. the performance of our influencer measure is based on the information diffusion measure. So, we need to study their behavior and other parameters that will maximize their influence in the network.

In the past decade, many efforts have been made to solve the link prediction problem in social networks [4]. There are many link prediction algorithms and conclusions derived based on the degree measure of the common connecting neighbors, resource allocation of the agent, community detection of the nodes in the social network, learning-based approach over the past activity of nodes, and so on [24]. But there are several shortcomings in these approaches. One of which is that the common connecting neighbor's derivation discards the fact that the social network is not stable and dynamic over some time. Another learning-based approach depends upon the previous activity of the network and any anomaly can lead to false predictions of links. Also, these conclusions and derivations neglect the fact that information propagation is one of the key aspects in the activation of a particular agent towards other agents and the prediction of a future link between those two. Information propagation is of great importance in the social network as it controls the ties between the agents and their characteristics. It can also be termed as influence propagation because it also influences a set of agents in the network.

In our dynamic algorithm towards the link prediction problem, we used the concept of predicting link based on the influence gained by a particular agent because of the propagation of certain information created by another agent of the same network such that it creates a possibility of link in between both agents leading to a dynamic link prediction system. The method for this approach consists of a threshold for the amount of information gained by an agent to form a link. This approach solves the shortcomings of previous algorithms by taking care of the information propagation as an important feature in link prediction. Also, it does not depend on the previously collected behavior of the network for predicting the link which can lead to false results for a minor anomaly in the collected data. Our approach depends on the current dynamics of the social network where every information produced by the agent leads to the prediction of his new link with other nodes. For the implementation of this agent-based model, the information has to be propagated from generating node to the maximum coverage of the network depending upon the state of other nodes whether they will forward that information or ignore it. Each information received by a particular node leads to the activation of the node towards the generating node. As the information reaches a node, it repeatedly increases the curiosity of the node, activating it, and link is formed. The same principle works for propagating the information further. As a node receives the same information multiple times, it tends to forward it as it is triggered by the information and forwards it. These two principles are the base for our approach in a dynamic link prediction system to be implemented. For practical implementation, Stanford's Facebook network dataset has been taken and strong attractors were chosen as the source of the information. This can be combined with a real-time network system for better results of link prediction.

This paper provides a comprehensive view of social influence propagation and link prediction using decision making with strong agents. It is organized as follows. In section 1, firstly we identify these strong influencers in the network that are the primary source of information. This information needs to be propagated to other nodes through links present between them. So, weights are assigned to each edge present between nodes by using the resource allocation method in section 2. At any instant of time, each agent will be in either of the three states based on the information collected by it. In section 3, we categorize all agents in these states using doubt threshold value which is calculated by the associated weights. These states tell us if the agent is ready to transmit the information further. Lastly, a dynamic link formation algorithm is proposed in section 4 taking into account all of the above parameters. This is then simulated on a social media dataset and the results are verified with other standard link prediction algorithms. Our results also reveal the formation of new links over time which solves the problem of the incompleteness of network dataset collected at any time instant.

2 LITERATURE REVIEW

Agent-based modeling is common in the study of complex networks and social networks because of the computation simulation properties and simulating the action and interaction of nodes in the network as autonomous agents. Interaction between the nodes is purely based on the concept of influence propagation[9] path in a social network. In this literature, the Threshold model [38] is the fundamental model that says that an agent adopts the opinion or influence if the percentage of agents in the network have already adopted that opinion. Agents in the ABM are defined based on their action [44]. These actions further define the simulation of agent-based modeling. The threshold model has been proved mathematically in the context of the influence propagation in the social network.

There are numerous research works on information diffusion over social networks [17, 47]. For instance, Gruhl et al. [19] studied and modeled the dynamics of information diffusion on the blog's space environment. Yang et al. [22] proposed a model to capture the attributes of information diffusion which are related to speed, scale, and range. With spreading of information diffusion models and their variations, Vallet et al. [43] used graph rewriting to compare the different information diffusion models. There are many previous works for predicting the link between the new nodes; the link prediction of new nodes between the agents of the ABM which are not yet connected but received influence about a particular node tend to form the link between them. The similarity-based link prediction algorithms are mostly static in nature and follow similarity measure calculation, ordering pairs, and ranking them accordingly. Some of them are the Jaccard coefficient which normalizes the size of the common neighbors and provides higher rank to those having more number of common neighbors against the total of their neighbors. It was further improved by Adamic and Adar for computing similarity between two web pages at first [2], after which it has been widely used in social networks. Another improvement towards this is the cn_soundarajan_hopcroft score [41] which considers the fact that pairs belonging in the same community are likely to form links rather than other pairs. But all these are somehow static in nature and prediction of link for t+1 period totally depends on the network structure in t-time. The learning-based approach for the link prediction is based on the binary classification of node pair which requires a good time span of network simulation for providing promising results over the network, while other learning approaches like probabilistic graph model assign a probability depending upon some criterion. In our literature, we consider the link prediction based on the fact of influence propagation. A learning-based approach which is treating link prediction problem as binary classification task [20].

3 METHODOLOGY

In this section, we propose the agent-based model for social influence, taking into account the previously received information by each agent. The paper is organized as follows- Firstly, we identify the strong influencers in any social network using various standard algorithms. Secondly, we describe the individual behavior of each agent, find the weight of each link formed [50] between them, and finally form new links dynamically in the last section. The proposed algorithms are later simulated on a social media dataset and the results obtained are found to be in accordance with the previous non-dynamic link prediction algorithms.

3.1 Strong attractors identification

An agent is defined as a vertex v in an undirected weighted social network $G = (V,E)$ where $V = (v_1,\ldots,v_n)$ denoted the set of agents and E represent the set of edges.

Our first goal is to find the set of influential nodes which will be a subset of V. Initially, here we consider the information flow begins from these strong influencers to reach a maximum level of saturation in the network. Information diffusion refers to the spread of abstract ideas or technical information within a social system, where spreading denotes flow or movement from a sender to the receiver node, typically via a communication link. This process of information flow is discussed further in section 3.3. The attractiveness value of any node v can be given by using methods like Betweenness centrality [16], degree centrality, PageRank clustering [15, 39] or eigenvector value. The study of centrality i.e., determining the importance of different nodes, edges, and other structures in a network has widespread applications in the identification and ranking of important agents (or interactions) in a network. These applications include ranking sports teams or individual athletes, the identification of influential people, and much more. For these and many other applications, it is important to develop and improve mathematical techniques to extract concise and intuitive information from large network data. However, despite the fact that real-world networks change with time, most methods for centrality (and node rankings that are derived from them) have been restricted to time-independent networks. For time-dependent (temporal) networks, one such method is eigenvector based centrality measure. However, Betweenness [5] is useful for analyzing communication dynamics and finding the individuals who influence the flow around a system. A high betweenness count could indicate someone holds authority over or controls collaboration between, disparate clusters in a network; or indicate they are on the periphery of both clusters. Betweenness centrality of a node v is the sum of the fraction of all-pairs shortest paths [6] that pass through v.

$$C_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)} \tag{1}$$

where V : the set of nodes

σ (s,t) : the number of shortest (s,t)-paths

$\sigma(s,t|v)$: the number of those paths passing through some node v other than s, t

If $s = t$, then

$$\sigma(s,t) = 1$$

if $v \in s, t$, then

$$\sigma(s,t|v) = 0$$

.

The nodes with maximum value of C_B would act as strong attractors in the network. This algorithm has been implemented in many different libraries and can be used including NetworkX, Boost, MATLAB, GraphStream etc.

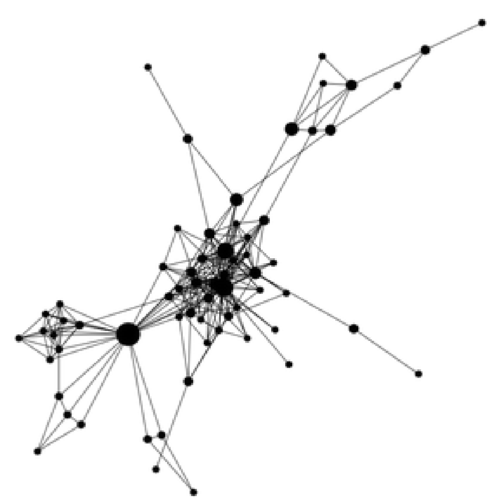


Figure 1: Strong agents in a social network showing nodes as agents where size of node is proportional to the betweenness centrality measure

In figure 1, the network consists of a set of different interconnected nodes as agents communicating through various links. On applying the above Betweenness centrality algorithm to the network, we get the shown segregation and small to big clusters formed around some nodes. The differently sized nodes have higher degrees of betweenness centrality value and so are the most influential nodes. Also, the centrality measure is directly proportional to the size of the cluster formed. Thus, we have identified strong influencers in the network. Among the rest agents, some of them may be more closely related or connected through a particular node than to others. This happens when that agent shares the same ideals, approaches, or opinions to that attractor. Thus, is more inclined towards it and is likely to propagate its information to others. In the further sections, it is well demonstrated that how this influence grows among the rest agents through these interconnecting links.

3.2 Edge weights assignment

The weights of the links in any social network depend on various factors and also it cannot be calculated totally. But in case of opinion propagation, resource allocation method can help in determining the edge-weights of the links approximately. Each link present between any two nodes would have an edge weight assigned to it.[32] This weight is assigned using the resource allocation method. It depends upon the number of common neighbours present and the degree of nodes.. The weight $w(i,j)$ i.e the edge weight between nodes i and j are given as

$$w(i,j) = 1 + \sum \frac{1}{N(U)} \tag{2}$$

where

$$U \in i \cap j$$

$N(U)$ is the function of input parameter node(U) which gives output as the degree of that particular node.

Equation 2 says that if there is only a single path present between any two nodes i and j, the value of U=0 i.e. There will be no common node between i and j, assigning weight to the link=1. Thus, the complete information flows through that single path. In this approach, it is decided beforehand, that what part of the information is transferred through a given link. The more the edge weight, the more reliable it is to transmit. It is similar to the strategy used by operating systems to allocate resources to user programs.

Algorithm 1 Edge weights assignment

for every edge (i,j) in graph G **do**
  $w(i,j) \longleftarrow 1$
end for
for every U $\in i \cap j$ **do**
  $w(i,j) \longleftarrow w(i,j) + \frac{1}{N(U)}$ // analogous to resource allocation
end for
return $w(i,j)$

In the example figure 2, node 1 is the sender node from where the information flow begins and node 2 is the receiver node. Information can reach the receiver node via three paths i.e. directly through the connecting link, or through neighbor nodes. Initially, assuming this an unweighted graph G, we implement the above edge weights assignment algorithm which uses the resource allocation method between nodes 1 and 2.

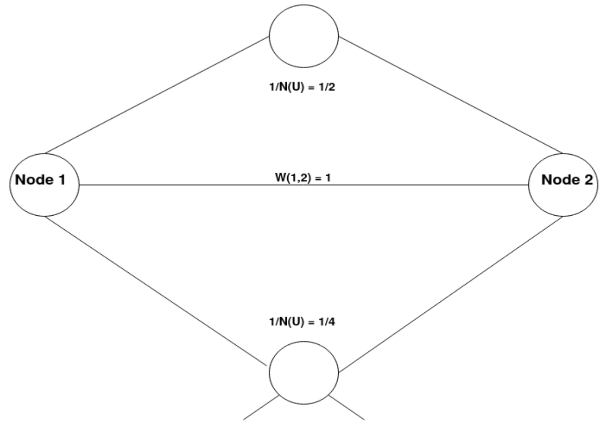


Figure 2: Example demonstrating assignment of edge weights between nodes

(1) Path 1 has a common node with degree 2. So we divide 1 by the degree and assign $\frac{1}{2}$.
(2) Path 2 has a single connecting link between nodes 1 and 2. As this is a direct link contributing weight as 1.
(3) Path 3 has a common neighbor, with degree 4. Thus, contributing weight as $\frac{1}{4}$.

Thus, after getting all the edge weights, the information is divided between the paths according to it's weight[12?]. The receiver node acquires the complete information after receiving and summing all it's connecting paths.

3.3 State of agents

The agents in the social environment are affected by the public opinion information [8] in many channels, and will have different actions based on the propagation effect. The public opinion information propagation characteristics are firstly analyzed, and different states of the people in the propagation environment are proposed. This is similar to the Linear Threshold model [18, 25] of information propagation where nodes get active in multiple stages. A node i is influenced by each neighbour j according to the doubt threshold value $D(i,j)$. The linear threshold model states that a node can make another node active only when it is in active state. Here doubt threshold value $D(i,j)$is given as-

$$D(i,j) = \frac{\sum_1^x w(i,j)}{\sum_1^n w(i,j)} \tag{3}$$

where x = no. of active neighbour nodes and n = total no. of nodes.

Based on the calculated doubt threshold, an agent will be in either of the following states at any particular instant of time-

Normal :Before the propagation begins, the node which has no information and thus, is in inactive state initially. After receiving some information, the agent then goes to the doubt state. The value of $D(i,j)$= 0.

Doubt :In this state, the receiver doubts the information received and so is not sure whether to propagate it further or not. This depends on the value of the doubt threshold. If $D(i,j) = (0, S_i]$, the agent remains in the doubt state. Here, S_i is called the spread threshold.

Spread :People in the spread state shall spread the information to its neighbours that they can affect, and they become the information source in the propagation process, thus becoming fully active. $D(i,j)$ = (S_i,1] for nodes in the spread state.

$$D_{i,j} = \begin{cases} 0 & \text{if in normal state} \\ (0, S_i] & \text{if in doubt state} \\ (S_i, 1] & \text{if in spread state} \end{cases} \quad (4)$$

where S_i=Spread threshold

Since the doubt threshold value always lie between [0,1], the above states can also be classified analogous to Continuous Prisoner's Dilemna used in game theory. [26, 45]

3.4 Dynamic link formation

Social networks[29] are dynamic in nature that may lead to the formation of new ties or links between the nodes or agents in the future. Therefore, predicting the new links formed with time is very important in understanding the network evolution [13]. In the past decades, many efforts have been made by computer scientists and economists to solve the link prediction problem in social networks. However, most of these prediction techniques are static in nature. These are the similarity based algorithms which take into account the number of common nodes initially in the network, and predict future links based on that data. The approach proposed in this paper is dynamic in nature. It considers the activation of nodes with time and threshold before predicting and forming new links in the network. This may solve the problem of dynamic behavior [36, 40] and incompleteness of the social network dataset up to a large extent.

Here in the proposed algorithms, each node maintains its priority heap which contains the description of the sender node i.e. the node which initially sends the information or opinion and information value. The information value is the hypothetical value of the information that might be used to trigger the receiving node. Initially, the sender node sends the information and we assume that information value is one, that propagates in the network by being multiplied with the edge-weight of the connecting links. It propagates the information further in the network based on its own state and threshold value. If the receiver node receives the same information from multiple paths [14, 28] over time, a link is formed dynamically between the sender and receiver nodes.

The exact method of information flow is given in Algorithm 2 for the nodes which change to spread state if the information received by it is greater than the spread threshold of the network.

send : For each neighbor of the node converted to the spread state, we send the information by calling the send function for that node. We assume that any node which is in spread state will send maximum information to its neighbors i.e. it will send $val = 1$ as assigned in line 4. $info\_id$ is the ID of the node whose information is going to be sent. In our case, the $info\_id$ will be the ID of any strong attractor. Line 5 onwards deal with the case when the strong attractor may receive information about itself. The input parameters for send functions are *node_id* which is going to send the information with $info\_id$ i.e. the node who has initiated the information propagation.

rec : Once the information with information value is sent by any node to its neighbor node, then rec function for that node is called. The parameters for the rec function are *send_ID* i.e. the node who sent information with $Info\_ID$,rec_ID is the node which is currently receiving the information. $info\_Val$ will be the information value received by the function. In lines 11-16, we take the case when the node is receiving information for the first time. The weights are calculated using the resource allocation method described in Algorithm 1. The *val* is set by diving it with total weight of it's surrounding nodes because according to Linear Cascade model [33], the probability of a node to be activated by another node depends upon the edge weight of the connecting node. We set a boolean variable *bool* to false which is used to check if the node has already sent the information once. This acts as the terminating case for our algorithm. In lines 17-21, if node has once received the information, then we just update the *val* and inf according to the further information received by it. Then we check for the edge threshold value according to the value of information received and again check the updated state of the node.

Algorithm 2 Information Propagation

```
1:  function SEND(nodeID,infoID)
2:      while eachv ∈ nodeID do
3:          inf ← edge_wt(v, nodeID)
        //info value set to edge weight
4:          val ← 1
5:          if v ≠ infoID then
        //check if sender node is not self node
6:              rec(nodeID, v, infoID, inf)
7:          end if
8:      end while
9:  end function

10: function REC(sendID,recID,infoID,inf,val)
11:     if infoID ∉ recID then
        //check if node is receiving info for the first time
12:         total_wt ← ∑ edge_wt(u, recID)
13:         val ← val∗edge_wt(sender,rec) / total_wt
        //val set to info received
14:         bool ← false
15:         link(th, val, recID, infoID)
16:         spread(inf, bool, recID, infoID)
17:     else if infoID ∈ recID then
18:         val ← val + edge_wt(sender,rec) / total_wt
        //val updated with new info received along with previous
    info
19:         inf ← inf + inf
20:         link(th, val, recID, infoID)
21:         spread(inf, bool, recID, infoID)
22:     end if
23: end function
```

Algorithm 3 Dynamic Link Formation

1: **function** LINK(th,val,rec,info_id)
2: **if** edge $\notin$ (info_id, r)& val $\geq$ thres **then**
3: $join(info\_id, rec)$
//check threshold and predict link
4: **end if**
5: **end function**

6: **function** SPREAD(inf,bool,rec,info_id)
7: $D_i \leftarrow \frac{inf}{total\_wt}$
8: **if** $D_i \geq S_i \& bool \leftarrow false$ **then**
9: $bool \leftarrow true$
10: $send(node\_id, rec\_id)$
//check spread threshold and send to all neighbors
11: **end if**
12: **end function**

The algorithm 3 forms the link between the sender and receiver nodes using the values calculated in these algorithms. If there is no existing edge between the sender and receiver nodes and the value of the node for particular info_id is greater than the edge threshold value, then we form a link between them. The function SPREAD in lines 6-12 checks for the state of the particular receiving node after it receives the information. If the *bool* value is false meaning that the node has never sent the information about that particular node and the information is greater than the spreading threshold i.e. $D_i > S_i$, then we assign *bool* as true, convert the node to the spread state and call the spread_info function for the receiving node as node_id and with the same $info\_id$. The range of the variable $info\_val$ will be [0,1] where 0 being the initial value when the node is not populated by any information from the strong attractor at a particular instance. As soon as it receives $info\_val$ from its neighbor node, it starts increasing the value of $info\_val$ based on the neighbor from which it received the information. The maximum value of 1 will be achieved when a node receives the same information from all of its neighbors.

The Big Oh complexity of the dynamic link formation algorithm will be $O(n^2)$. This will be the case at minimum threshold value which results in a complete graph formed between n nodes with the maximum number of links as- ${}^{n}C_2$.

Thus, by changing the spread threshold values, the maximum number of links are formed among all the nodes in the network, and the maximum level of saturation is reached when no more further links can be formed. This growing interaction between agents results in the increasing influence of the strong attractors in the network. [46]

4 RESULTS

This methodology was then implemented on a real-time social media dataset and the results were compared with the previous Adamic Adar index values to test the accuracy of our algorithm. The number of links formed by our opinion propagation algorithm was found to be more than those predicted by Adamic Adar because they were based on the information received by the nodes with time, that is, we have considered the dynamics of the network.

On varying the link threshold values, we get the different number of links formed between nodes. Our simulation results also show information outburst which is in accordance with the linear threshold model. It shows that a slight change in the threshold values brings a large change in the number of active nodes at a time. Thus, proving that an active node is more likely to make its neighboring nodes active.

4.1 Dataset

The method described in section 3 is implemented on a Facebook dataset consisting of nodes as users and edges as links [Dataset Web Link]. This dataset consists of 'circles' (or 'friends lists') from Facebook. Facebook data was collected from survey participants using this Facebook app. The dataset includes node features (profiles), circles, and ego networks. It consists of a total of 4039 nodes with 88234 edges between them. The data has been anonymized by replacing the Facebook-internal ids for each user with a new value. Also, while feature vectors from this dataset have been provided, the interpretation of those features has been obscured. For instance, where the original dataset may have contained a feature "political=Democratic Party", the new data would simply contain "political=anonymized feature 1". Thus, using the anonymized data it is possible to determine whether two users have the same political affiliations, but not what their individual political affiliations represent. Modeling and simulation are done on a subset of 1500 nodes of this dataset for verification.

The original dataset has an average clustering coefficient of 0.65 that is, the network is actually a closed network which is more preferred for the propagation of the information. The statics of this network is mostly compiled by combining the ego-networks with ego nodes themselves. Those ego nodes themselves act as strong attractors because of their sub ego networks. The ego circles are defined along with their features in the feat, but these features are not used in the implementation of agent-based modeling and directly implemented as edges and nodes. Simulation results are observed on these nodes and links by varying threshold and other parameters with time.

4.2 Comparing new links predicted

With the subset 50 nodes, firstly the strong attractors are identified using betweenness centrality and checked for a particular large D_i and S_i values. It is assumed that the nodes have IDs ranging from [1,50] and we observed that the 17^{th} node is the strong attractor according to the given graph. Further simulation is done according to the algorithm on S_i value of 0.3 and the link threshold value of 0.2. So for strong attractors, the Adamic Adar index predicted 3 future links for that instance that are (1,17=2.88),(17,41=1.63),(17,22=0.91). The range of Adamic Adar value is totally relative to the size of the network and total number of common nodes. For our algorithm to work properly, our implementation should cover these links initially than any other further links to be predicted. Our dynamic link prediction algorithm predicted 4 links initially for the above threshold value and they were (17,41),(1,17),(22,17),(17,31) covering original links along with prediction of one more link.

The loop exit condition for the algorithm is that once a particular node has sent the information of a node to its neighbor, then it would not send it again so that it does not go in an infinite loop and can finally converge.

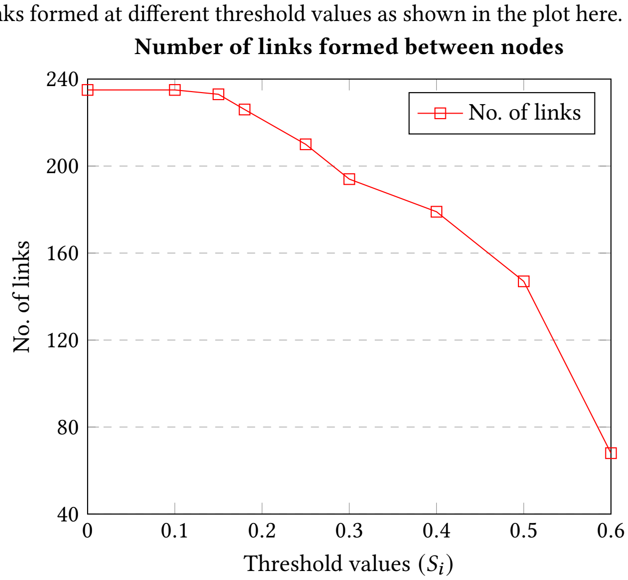
Number of nodes taken	Adamic Adar	Opinion Propagation
50	5	9
100	12	21
500	82	377
1500	266	402

Table 1: Comparison of links predicted by Adamic Adar index and Dynamic Opinion Propagation algorithm

We tested our algorithm similarly on the various number of nodes taken from the dataset and results are shown in table 1. Firstly, node ID 1888 was identified as the strong attractor by betweenness centrality measure. Then, at a link threshold value of 0.3 and the spread threshold value (S_i) of 0.01, links were predicted. Columns 2 and 3 show the number of links formed by Adamic Adar and Opinion Propagation algorithm respectively. The number of links formed by dynamic opinion propagation algorithm were more than as predicted by Adamic Adar. Moreover, our algorithm also covered all the previous links predicted by the Adamic Adar index. Thus, proving our dynamic algorithm more efficient.

4.3 Number of links formed

The variation of the doubt threshold D_i and spread threshold S_i are the key factors for determining the spread of information from the strong attractors and also the formation of future links to be predicted in the network. By varying these parameters, the interaction between agents in the network varies. At lower threshold values, information travels faster from one node to another, and links are formed quickly whereas, at higher values of threshold, less number of reliable links are formed in the network. Thus, we obtain a decreasing graph when plotted between the number of links formed at different threshold values as shown in the plot here.



4.4 Adamic Adar vs dynamic links

According to Adamic Adar index for this dataset, the number of links to be initially predicted for particular strong attractors is 266 for the dataset of 1500 nodes, while according to our algorithm a total of 402 links have info_val = 1, at the same threshold values taken as before, predicting these links to be formed first covering all the nodes initially predicted by the Adamic Adar index.

Serial no.	Adamic Adar (i,1888)	Opinion Propagation (j,1888)
1	1563	1383
2	1409	1563
3	1795	1539
4	1845	1395
5	1539	1521
6	1290	631
7	1359	1795
8	1688	1409
9	1280	1765
10	1644	1430
11	1040	584
12	1047	679
13	1765	616
14	1535	574
15	1029	1539
16	1528	1040
17	1056	682
18	932	622
19	1571	1029
20	1523	657

Table 2: Node IDs of top 20 links predicted by Adamic Adar index and Dynamic Opinion Propagation algorithm on a network of 1500 nodes.

Table 2 is the comparison between the top 20 links that are predicted by the Adamic Adar and our opinion propagation method. On applying the betweenness centrality measure over the Facebook dataset of 1500 nodes taken for obtaining strong attractors in that network for opinion propagation, we got node ID 1888 as the strong attractor. Column 2 in the table gives node IDs (i) who form a link with node ID 1888 by Adamic Adar algorithm whereas column 3 gives the node IDs (j) which form link with ID 1888 by opinion propagation algorithm. We notice that some of these IDs are common between the two algorithms and some are new in column 3. Now on applying the Adamic Adar over the strong attractor, we got the top 20 links predicted only on the basis of their common neighbor score, and if we continue to predict link according to that only, then we will end up with clustering of nodes along a node and no saturation of the data will be achieved. On the other hand, in the opinion propagation method, we can see that the top links are predicted according to the information propagated and it followed an iterative approach that is, first the information is propagated in the whole network after that the comparison for threshold is checked and links are predicted.

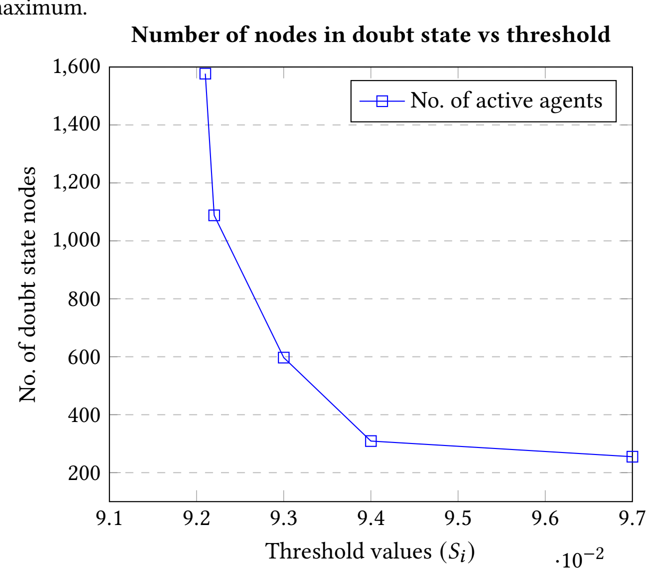
The accuracy of our algorithm can be checked as it does not miss any links that are predicted by Adamic Adar, and along with that we found different new links that are predicted. The Adamic Adar here is following breadth first approach and links predicted are more in the clustered form while in our opinion propagation method it is following more of depth first approach.

4.5 Convergence of algorithm

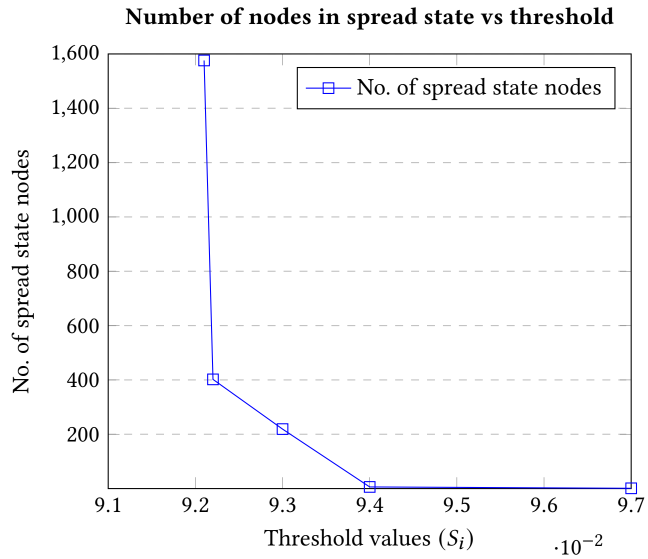
From these experimental results, it can be observed that the value of the spread threshold is inversely proportional to the size of the network. Also, the spread threshold is more promising in providing control over information flow and link prediction rather than the threshold value. The neighbor nodes could not cross the threshold value because of more node modularity and less value of the linear threshold. So, they simply propagate the information further and do not form any links. When this information is propagated to further nodes in the network, this threshold value is crossed and their probability of link formation increases.

4.6 Information outburst

Now, on applying this evaluation strategy on the whole dataset of 1500 nodes, the whole network is not totally connected and has 3 clusters, for instance resulting in the total number of links formed by the strong attractor to be less than 1500. The strong attractors can somehow control the flow of information up to some extent and that has been proved using the graph between the number of nodes in doubt state in information vs spread threshold value. As we can see from the graph that at a particular value of threshold, there is an outbreak of information and the reachability of information is maximum.



The same principle applies for the number of active nodes for the variation of the spread threshold value, as the outbreak of information happens, the number of nodes activated by that particular information from the strong attractor increases exponentially predicting the maximum number of link formations. This implies that the threshold value loses its generosity because everyone is in the active state sending information to one another. This is also in accordance with the Linear Threshold model used above which gives us the result that an active node holding some information is more likely to make its neighboring nodes active by spreading it further.



From such instances, it is observed that if a particular node has a common neighbour to the strong attractor and in the situation of information outbreak, that node will receive information from all of its neighbours and will go into spread state quickly.

5 CONCLUSION

In this research work, we have modeled the real-world influence propagation network as a complex system, where weighted bidirectional influence exists. The strong attractors in the network are identified and the clusters of the influenced nodes formed by them are studied. Information flow between nodes is analyzed by dynamic link prediction algorithms. The research achievements in this paper shall help to enhance the analysis and understanding of the strong influencers and how their influence can be maximized by altering various threshold parameters. This work can be applied to study the influence of major politicians in a community, in market analysis to increase the sale of a specific product, or for security purposes like to reveal new terrorist groups formed by taking into account their past links, etc.

Finally, this paper presented the use case of the Facebook dataset around strong agents. But in order to arrive at a more general model for social influence, the model needs to be tested on other use cases as well. The links predicted by the opinion propagation algorithm take the dynamics of social networks into account and so is more efficient than the Adamic Adar index.

REFERENCES

[1] 2010. Fast and scalable algorithms for semi-supervised link prediction on static and dynamic graphs. *Proceedings of ECML/PKDD'10, Barcelona* (2010), 131–147.

[2] Adar E. Adamic L A. 2003. Friend and neighbors on the web. Soc Networks. 25 (2003), 211–230.

[3] Jarada T N et al Almansoori W, Gao S. 2012. Link prediction and classification in social networks and its application in healthcare and systems biology. *Netw Model Anal Health Inform Bioinform* 1 (2012), 27–36.

[4] Jeong H Barab asi A L. 2002. Evolution of the social network of scientific collaborations. *Physica A* 311 (2002), 590–614.

[5] Ulrik Brandes. 2001. A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology* 25(2) (2001), 163–177.

[6] Ulrik Brandes. 2008. On Variants of Shortest-Path Betweenness Centrality and their Generic Computation. *Social Networks* 30(2) (2008), 136–145.

[7] Bonchi F et al Bringmann B, Berlingerio M. 2010. Learning and predicting the evolution of social networks. *IEEE Intell Syst* 25 (2010), 26–35.

[8] Longjie Li ; Shiyu Fang ; Shenshen Bai ; Shijin Xu ; Jianjun Cheng ; Xiaoyun Chen. 2019. Effective Link Prediction Based on Community Relationship Strength. *IEEE* 7 (March 2019), 43233 – 43248.

[9] Lakshmanan LVS Castillo C Chen, W. 2013. *Information and Influence Propagation in Social Networks.* Morgan and Claypool Publishers.

[10] Elena Ferrari Cuneyt Gurcan Akcora, Barbara Carminati. 2011. Network and profile based measures for user similarities on social networks. *Proceedings of the 12th IEEE International Conference on Information Reuse and Integration, Las Vegas* (2011), 292–298.

[11] V. M. Eguiluz D. Centola, J. C. Gonzalez-Avella and M. San Miguel. 2007. *Homophily, cultural drift, and the co- evolution of cultural groups Journal of Conflict Resolution* 51, 6 (2007), 905–929.

[12] Arenas A Rosvall M De Domenico M, Lancichinetti A. 2015. dentifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys Rev X* 5(1) (2015), 011027.

[13] Hially Rodrigues de Sá ; Ricardo B. C. Prudêncio (Ed.). 2011. *Supervised link prediction in weighted networks.* IEEE. https://doi.org/10.1109/IJCNN.2011.6033513

[14] Scala A D'Agostino G. 2014. *Networks of networks: the last frontier of complexity.* Springer.

[15] Rácz B. Csalogány K. Sarlós T. Fogaras, D. 2005. Towards scaling fully personalized PageRank: Algorithms, lower bounds, and experiments. *Elsevier* 2 (January 2005), 333–358. Issue 3.

[16] Linton C. Freeman. 1977. A Set of Measures of Centrality Based on Betweenness. *JSTOR* 40 (1977), 35–41.

[17] Jennifer Golbeck. 2013. *Chapter 3 - Network Structure and Measures.* Morgan Kaufmann. 25–44 pages. https://doi.org/10.1016/C2012-0-00171-8

[18] Mark Granovetter. 1978. Threshold Models of Collective Behavior. *Amer. J. Sociology* 83, 6 (May 1978).

[19] Liben-Nowell D Tomkins A Gruhl D, Guha R. 2004. Information diffusion through blogspace. *Proceedings of the 13th international conference on World Wide Web. London: ACM;* (May 2004), 491–501.

[20] Salem S et al. Hasan M A, Chaoji V. 2006. Link prediction using supervised learning. *In: Proceedings of SDM'06 Workshop on Link Analysis, Counter terrorism and Security, Bethesda* (2006).

[21] Lin D K J Huang Z. 2009. The time-series link prediction problem with applications in communication surveillance. *INFORMS J Comput* 21 (2009), 286–303.

[22] Scott Counts Jiang Yang. 2010. Predicting the speed, scale, and range of information diffusion in twitter. *Proceedings of the Fourth International Conference on Weblogs and Social Media ICWSM.* 10 (May 2010), 355–358.

[23] Hisashi Kashima Junichiro Mori, Yuya Kajikawa. 2011. Finding business partners and building reciprocal relationships - A machine learning approach. *IEEE* (2011). https://doi.org/10.1109/itmc.2011.5996005

[24] Budka M Juszczyszyn K, Musial K. 2011. Link prediction based on subgraph evolution in dynamic social networks. *Proceedings of the 2011 IEEE International Conference on Social Computing, Boston* (2011), 27–34.

[25] Heyan Huang Kan Li, Lin Zhang. 2018. Social Influence Analysis: Models, Methods, and Evaluation. *Elsevier* 4 (February 2018), 40–46. Issue 1.

[26] Timothy Killingback and Michael Doebeli. 2002. The Continuous Prisoner's Dilemma and the Evolution of Cooperation through Reciprocal Altruism with Variable Investment. *The American Naturalist* 160, 4 (2002), 421–438.

[27] Leskovec J Kim M. 2011. The network completion problem: inferring missing nodes and edges in networks. *Proceedings of the 11th SIAM International Conference on Data Mining (SDM'11), Mesa* (2011), 47–58.

[28] Goh K-I. Lee K-M, Min B. 2015. Towards real-world complexity: an introduction to multiplex networks. *Eur Phys J B* 88(2) (2015), 1–20.

[29] Lim E.-P. Lo-D. Weng J. Leung, C.W.-K. 2010. Mining interesting link formation rules in social networks. *In: CIKM* (2010).

[30] Sud Toulon-ROSSANO SCHIFANELLA CIRO CATTUTO BENJAMIN MARKINES LUCA MARIA AIELLO, ALAIN BARRAT and FILIPPO MENCZER. 2012. Friendship Prediction and Homophily in Social Media. *ACM Transactions on the Web* 6, 2 (2012), 1–33. https://doi.org/10.1145/2180861.2180866

[31] Priebe C E Marchette D J. 2008. Predicting unobserved links in incompletely observed networks. Comput Stat Data Anal. 52 (2008), 1373–1386.

[32] Albert Diaz-Guilera Alex Arenas Marián Boguñá, Romualdo Pastor-Satorras. 2004. Models of social networks based on social distance attachment. *Physical Review E (PRE)* 70 (2004).

[33] Nishith Pathak, Arindam Banerjee, and Jaideep Srivastava (Eds.). 2010. *2010 IEEE International Conference on Data Mining.* IEEE. https://doi.org/10.1109/ICDM.2010.153

[34] Ichise R Pavlov M. 2007. Finding experts by link prediction in co-authorship networks. *Proceedings of the 2nd International ISWC+ASWC Workshop on Finding Experts on the Web with Semantics (FEWS), Busan* (2007), 42–55.

[35] Hachen D et al Raeder T, Lizardo O. 2011. Predictors of short-term decay of cell phone contacts in a large scale communication network. *Soc Netw* 33 (2011), 245–257.

[36] Kleinberg J. Romero, D.M. 2010. The directed closure process in hybrid social-information networks, with an analysis of link formation on Twitter. *In: ICWSM* (2010).

[37] L. Muchnik S. Aral and A. Sundararajan. 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences* 106, 51 (2009), 544–549.

[38] T. C. Schelling. 1978. *Micromotives and Macrobehavior.* Vol. 17. Norton, W.W. Norton & Company, Inc. 500 Fifth Avenue, New York, N.Y. 10110.

[39] Masoud Asadpour Maziar Abbasi Mohammad Ali Tavallaie Shayan A. Tabrizia, Azadeh Shakery. 2013. Personalized PageRank Clustering: A graph clustering algorithm based on random walks. *Elseveir* 392 (2013), 5772–5785. Issue 22.

[40] Van de Bunt G.G. Steglich C.E.G. Snijders, T.A.B. 2010. Introduction to stochastic actor-based models for network dynamics. *Social Networks, Special Issue on Dynamics of Social Networks* 32(1) (2010), 44–60.

[41] Sucheta Soundarajan and John Hopcroft. 2012. Using community information to improve the precision of link prediction methods. *In Proceedings of the 21st international conference companion on World Wide Web (WWW '12 Companion). ACM, New York, NY, USA* (2012), 607–608. https://doi.org/10.1145/2187980.2188150

[42] Sun J M et al Tang J, Wu S. 2012. Cross-domain collaboration recommendation. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'12), Beijing* (2012), 1285–1293.

[43] Pinaud B Melançon G. Vallet J, Kirchner H. 2015. A visual analytics approach to compare propagation models in social networks. *EPTCS 181* (2015), 65–79. https://doi.org/10.4204/EPTCS.181.5

[44] Peter-Paul van Maanen and Bob van der Vecht (Eds.). 2013. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013).* IEEE. https://doi.org/10.1145/2492517.2492564

[45] T. Verhoeff. 1993. *A continuous version of the prisoner's dilemma.* Technische Universiteit Eindhoven.

[46] Wenjun Wang and W. Nick Street. 2018. Modeling and maximizing influence diffusion in social networks for viral marketing. *Springer* 3, 6 (2018), 1–26. https://doi.org/10.1007/s41109-018-0062-7

[47] J. D. Watts and P. S. Dodds. 2007. Influentials, networks, and public opinion formation. *Journal of Consumer Research* 34 (Dec. 2007), 441–458.

[48] Ichise R Wohlfarth T. 2008. Semantic and event-based approach for link prediction. *Proceedings of the 7th International Conference on Practical Aspects of Knowledge Management (PAKM'08), Yokohama* (2008), 50–61.

[49] Tang Jie Wu Sen, Sun Jimeng. 2013. Patent partner recommendation in enterprise social networks. *Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM'13), Rome* (2013), 43–52.

[50] Yang J Fang H Zhang QM Nie M Holme P Zhou T Zhao J, Miao L. 2015. Prediction of links and weights in networks by reliable routes. *Sci Rep* 5 (2015), 12261.

[51] Hughes J G Zhu J, Hong J. 2002. Using Markov models for web site link prediction. *Proceedings of the 13th ACM Conference on Hypertext and Hypermedia (HYPERTEXT'02), Maryland* (2002), 169–170.