Signed Page Rank Algorithm for Influence Maximisation in Signed Social Network

Report submitted in fulfillment of the requirements for the Exploratory Project of

Second Year B.Tech.

by

Sarthak Bansal(18075054) Saurabh Yadav(18075055) Sushant Ashutosh(18075059)

Under the quidance of

Dr. Lakshmanan Kailasam



Department of Computer Science and Engineering INDIAN INSTITUTE OF TECHNOLOGY (BHU) VARANASI Varanasi 221005, India June 2020

Dedicated to

Our parents, professors,
exploratory project convenor,
mentor and everyone who helped
and motivated us in successful
completion of this report.

Declaration

We certify that

1. The work contained in this report is original and has been done by us(group

members) and the general supervision of our supervisor.

2. The work has not been submitted for any project.

3. Whenever we have used materials (data, theoretical analysis, results) from other

sources, we have given due credit to them by citing them in the text of the thesis

and giving their details in the references.

4. Whenever we have quoted written materials from other sources, we have put

them under quotation marks and given due credit to the sources by citing them

and giving required details in the references.

Place: IIT (BHU) Varanasi Sarthak Bansal

Saurabh Yadav Sushant Ashutosh

Date: Department of Computer Science and Engineering,

Indian Institute of Technology (BHU) Varanasi,

Varanasi, INDIA 221005.

Certificate

This is to certify that the work contained in this report entitled "Signed Page

Rank Algorithm for Influence Maximisation in Signed Social Network"

being submitted by Sarthak Bansal(18075054), Saurabh Yadav(1875055)

and Sushant Ashutosh(18075059), carried out in the Department of Computer

Science and Engineering, Indian Institute of Technology (BHU) Varanasi, is a bona

fide work of our supervision.

Dr. Lakshmanan Kailasam

Place: IIT (BHU) Varanasi

Date:

Department of Computer Science and Engineering, Indian Institute of Technology (BHU) Varanasi, Varanasi, INDIA 221005. Acknowledgments

We want to express our sincere gratitude to the people who have helped us the most

throughout our project. We are grateful to our project supervisor Dr. Lakshmanan

Kailasam for providing us an opportunity to implement the paper entitled "Signed

Page Rank Algorithm for Influence Maximisation in Signed Social Network

" and his constant support for the project.

We wish to thank our parents for their support and attention. We would like to thank

our friends who encouraged us and helped us out in finalizing the project. At last,

we also thank the Almighty for his blessings.

Place: IIT (BHU) Varanasi

Sarthak Bansal Saurabh Yadav Sushant Ashutosh

Date:

Abstract

We have implemented the **Signed Page Rank** algorithm for influence maximisation in signed social networks which makes use of both friendly and hostile relationship between individuals to select handful of individuals which helps maximise the ad propagation thus keeping the cost of advertisement reasonable. This Signed Page rank algorithm is based on the traditional Page Rank algorithm which accesses the importance of web pages on a network based on their topological properties. Signed Page rank algorithm is a very efficient for ad recommendation in signed Social Networks.

Contents

List of Figures					
1	Introduction				
	1.1	Overv	iew	2	
	1.2	Motiv	ation of the Research Work	3	
	1.3	Organ	sisation of the Report	4	
2	Project Work				
	2.1 System Model		m Model	5	
	2.2 Algorithm		ithm	6	
		2.2.1	Information Propagation	6	
		2.2.2	Belief Update	9	
		2.2.3	Signed Page Rank Algorithm	12	
3	Dat	aset		14	
4	4 Results and Conclusion				
Bi	Bibliography				

List of Figures

2.1	Information Propagation	8
2.2	Belief Update	11
2.3	Signed Page Rank	13
3.1	Dataset	15
4.1	Result1	16
4.2	Result2	17

Chapter 1

Introduction

1.1 Overview

In today's world, almost everyone uses social networks in their life, and thus, it has become a very integral part of our society. To put it simply, life without social network is unimaginable in today's world. This evergrowing use of social networks attracts various commercial companies as they see it as a scope of advertisement for their products. Social networks are becoming very important for the propagation of advertisements as various groups of connected people are influenced by each other which can be used as an advantage for ad propagation. This new form of advertisement serves as a revenue for various social media sites. These company hire specialists or buy implementations of various ad propagation techniques and algorithms to fulfil the needs of the commercial companies interested in using their social networks for advertisement. It is important that most influential members of the social network are chosen for initiation of ad propagation as it is only practical and fruitful that while keeping the cost of advertisement minimal, the advertisement reaches maximum spread. This is called influence maximisation in social network. To achieve this the influence maximisation algorithm chooses some (say k) most influential members of the social network as seed node. (Each member of the social network is considered a node). So, a social network can be represented as a signed weighted directed graph in which an individual is represented as a node and the directed edge (say from node u to v) between two nodes represents the relationship between those two nodes. Weight of the edge denotes the influence of one individual(u) over other(v). Label of the edge is either positive or negative. Positive label is when two nodes share a positive relationship (u shares positive relationship towards v) i.e. they are friends and negative when two nodes share a negative relationship (u shares negative relationship towards v) i.e. they are foes. Sign (or the nature of influence of one individual over other) is considered an important factor in information propagation as an individual getting ad recommendation from a foe may have an adverse effect on ad propagation while from a friend is beneficial. So according to researchers, signed social networks give better results than unsigned social networks as signed social networks have richer information about the social relationship between two individuals and hence the influence of one inividual over other. We have implemented the Signed-Page-Rank (SPR) algorithm for influence maximization as all known algorithms (eg. SIR(Susceptible Infected Removed) model, Independent Cascade (IC) model, etc) are for unsigned social networks. There had been efforts to implement them in signed social networks but these implementations treated the negative and positive parts of the network separately which is not very efficient.

1.2 Motivation of the Research Work

We explored on social networks and research work on influence maximisation in social network. We studied about various influence maximisation algorithms for unsigned social network (eg. SIR (Susceptible Infected Removed) model, Independent Cascade (IC) model, etc) We came across the Novel Signed Page Rank Algorithm [1] as researched and published in [1]. Since this was a recent research work and also we realised that there has not been much work on influence maximisation in signed social

network, we, under the guidance of our supervisor decided to explore this algorithm more and to implement it with our own codes and test it with our selected dataset and gain better understanding of this algorithm trough this implementation.

1.3 Organisation of the Report

We have implemented the Signed-Page-Rank (SPR) algorithm for influence maximization. Signed-Page-Rank (SPR) algorithm ranks the individual (nodes) according to their influence. The traditional Page-Rank algorithm is used to sort the importance of web pages based on topological properties of web graphs. We have also used information propagation algorithm which is like a directing function for the influence maximization in signed social network i.e. it will call SPR(for selecting initial k nodes for maximizing influence) and will also update the belief of the nodes after every time increment by calling belief update function which will update belief according to a certain rule which is explained in later section of report. We start the report by defining the system model in section 2.1. We discuss our algorithm in the following section in three sub-section. In the first subsection we discuss about information/ad propagation in the social network. Following which we discuss Belief update of nodes and algorithm to do it. Finally we end the section by discussing the Signed Page Rank algorithm which is used to select the initial seed nodes for ad propagation. We End the chapter with this. In the following chapters we discuss the dataset that we have used and the result of our implementation of the aforesaid algorithm.

Chapter 2

Project Work

2.1 System Model

A social network having both friendly and hostile relationship between individuals where it is not necessary that these relationships are mutual (We will further say that this characteristic of the social network allows the weighted directed signed graph that we will suggest to be non-symmetric). The social network with the above mentioned characteristics can be well represented by a weighted directed (digraph) signed graph G(V,E) where V is the set of individual (or the nodes of the graph) and E is the set of directed edges between the individual (nodes). The number of nodes or individuals is **N**, i.e. |V| = N. Each edge in **E** is associated with a weight $w_{i,j} \in [0,1]$ which represents the level of influence of individual i on individual j, i.e. the probability of successful recommendation of ad by individual i to individual j. As mentioned earlier the graph is asymmetrical, i.e. $w_{i,j} \neq w_{j,i}$. Every edge also has a label $l_{i,j}$ which is equal to +1 for positive influence or friendly relationship and -1 for negative influence or foe relationship. The label is a representation of approval or disapproval of one another's comments on social platforms, trust or distrust one another's reviews about a certain product on social platforms. A positive directed edge indicates friendliness of one individual towards the other, while a negative link implies hostility of one individual towards the other.

Every individual in the social network has a belief, a subjective consciousness towards the advertisement which at any time t, of node v_i is represented by $x_{i,t} \in [0,1]$ which will evolve over time. also each node v_i is receptive to ad recommendations only within its time window $T_i = [T_i^l, T_i^u]$, called the recommendation cycle. No node will be receptive to the advertisement before or after its time window. $A_{i,t}$ represents the attitude of node v_i at time t. If it is 1 it will recommend the ad to its out-degree neighbours otherwise it will remain in susceptible state (if the current time is within in its recommendation cycle) or in removed state (if its recommendation cycle has elapsed). The set of in-degree neighbours is represented by H_i^{in} ($H_i^{\text{in+}}$ for positive indegree neighbours and $H_i^{\text{in-}}$ for negative in-degree neighbours). The set of out-degree neighbours is represented by $H_i^{\text{out-}}$ for positive out-degree neighbours and $H_i^{\text{out-}}$ for negative out-degree neighbours).

2.2 Algorithm

We implemented the influence maximisation algorithm in three parts as following:

2.2.1 Information Propagation

We can explain this function by dividing it into 3 sections that are initiation, propagation and termination. Let us define the following terms first:

- 1. Infectious individuals: All the individuals that have an attitude equal to one are termed as infectious as they will spread the infection or in other terms they will recommend ad to their neighbours, they are also termed as seed nodes.
- 2. Susceptible individuals: All the individuals whose attitude is equal to zero and within her recommendation cycle are susceptible. So they are the individuals who will receive ad recommendation from their neighbours and will update their beliefs.
- 3. Removed individuals: All the individuals whose attitude is equal to zero and

beyond her recommendation cycle, i.e., $t>T_j^u$, will be removed.

So, now the information propagation function works as follows:

Initiation: If t = 0, all individuals will have an attitude equal to zero so we will select exactly k nodes i.e. seed nodes for maximizing the influence using SPR algorithm.

Propagation: If t > 0 then we will propagate and select all the infectious individuals at that time, i.e. individuals or nodes having attitude equal to one. If there are no susceptible individuals at current time then we will increment the time and call the propagation function to propagate. Otherwise we will first update the belief of these susceptible individuals and check if individuals can recover i.e. individuals whose attitude become one after belief update and then recover to zero by a certain probability $p_{i,t}$ which is basically defined for every individual v_i as $1 - x_{i,t}$.

Termination: The information propagation terminates if all individuals are either infectious or removed.

Below is code snip of our implementation of information propagation.

```
def inpro(G,x,A,label,TP,t,k):
     if t==0: #Intiation
    S=SPR(G,x,label,k) #Selecting initial k seed nodes
          for i in range(len(S)):
    A[S[i]]=1
          S=set()
          for i in range(len(A)):
    if(A[i]==1):
        S.add(i) #In
     fa=[]
fb=[]
     fb=[]
for i in range(n):
    if i not in S and t>=TP[i][0] and t<=TP[i][1]:
        fa_append(i) #Susceptible individuals</pre>
          elif i not in S and t<TP[i][0]:
    fb.append(i)  #Susceptible individuals but not ready for receiving ad recommendation</pre>
     if len(fa)==0 and len(fb)!=0:
          t=t+1
          inpro(G,x,A,label,TP,t,k)
     elif len(fa)!=0:
          (x,A)=BUpdate(G,x,A,label,S) #Updating belief of susceptible individuals
           for i in range(n):
    if i not in S and A[i]==1:
                     if 1-x[i]>=0.7:#let the threshold on revoking the advertisement after accepting be 0.7
A[i]=0 #Recovered
          inpro(G,x,A,label,TP,t,k)
          print("The total time taken for propogation: ",t) #Termination
```

Figure 2.1 Information Propagation

2.2.2 Belief Update

When a susceptible individual receives ad recommendation from seed nodes (or infectious individuals) then their belief will change and will be updated according to the below-specified rules. We have assumed that the belief of the seed node i.e. individual who has accepted the ad will not be updated. As we have already discussed that positive and negative relationships between two nodes have different effects on information propagation. So their belief update will also be different as it is obvious that individuals are more likely to trust their friends than their foes. So negative relationships may reduce the belief of the individual whom it has recommended the ad.

There are three types of belief update rules:-

While calculating positive and negative updates we have assumed that this node receives ad recommendations only from its friends or only its foes but not both. In case an individual receives ad recommendation from both its friends and foes its belief is updated using the third rule of parallel recommendation.

1. Positive Belief Update: When node v_i receives ad recommendation at time t from its friends, i.e $l_{j,i} = 1, \forall v_j \in H_i^{in} \cap \mathcal{S}_i, v_i$ updates its belief as:

$$x_{i,t+1} = x_{i,t} + \alpha_i \cdot \sum_{v_j \in H_i^{in} \cap S_t} w_{j,i} \cdot (x_{j,t} - x_{i,t}), \qquad (2.1)$$

where α_i is the coefficient of positive embeddedness of individual v_i defined as

$$\alpha_i = \frac{\left| H_i^{in+} \right|}{\left| H_i^{in} \right|} \tag{2.2}$$

where $|H_i^{in}|$ is the number of in-degree and $|H_i^{in+}|$ is the number of positive in-degree neighbours, and we have $0 \le \alpha \le 1$.

2. Negative Belief Update: When node v_i receives ad recommendation at time t

from its foes, i.e $l_{j,i} = -1, \forall v_j \in H_i^{in} \cap \mathcal{S}_i, v_i$ updates its belief as :

$$x_{i,t+1} = x_{i,t} - \beta_i \cdot \sum_{v_j \in H_i^{in} \cap S_t} w_{j,i} \cdot (x_{j,t} - x_{i,t}), \qquad (2.3)$$

where β_i is the coefficient of negative embeddedness of individual v_i defined as

$$\beta_i = \frac{\left| H_i^{in-} \right|}{\left| H_i^{in} \right|} \tag{2.4}$$

where $|H_i^{in}|$ is the number of in-degree and $|H_i^{in-}|$ is the number of negative in-degree neighbours, and we have $0 \le \beta \le 1$.

3. Belief Update for parallel recommendation: When a node v_i receives ad recommendation from friends as well as foes at time t, the belief update rule is somewhat hybrid of positive and negative belief updates. It is given by

$$x_{i,t+1} = x_{i,t} + \alpha_i \cdot \sum_{v_j \in H_i^{\text{in}^+} n s_i} w_{j,i} \cdot (x_{j,t} - x_{i,t}) - \beta_i \cdot \sum_{v_j \in H_i^{\text{in}^-} n s_t} w_{j,i} \cdot (x_{j,t} - x_{i,t}) \quad (2.5)$$

where α_i and β_i are positive and negative embeddedness as defined above.

After the belief update, the node having sufficiently high value value of belief (say above a certain threshold (Here we have taken the threshold to be 0.7)) and do not have attitude 1 will change their attitude to 1.

The belief update function returns the updated beliefs of each node as well as the new Attitudes of each node.

Here is the code snip of our implementation of Belief Update.

```
def BUpdate(G,x,A,label,S):
    for v in range(n):
        alp=0.0
        bet=0.0
        pos=0.0
        neg=0.0
                      #Belief of seed nodes will not be updated
        if A[v]==1:
            continue
        for u in S:
            if v in G.neighbors(u):
                if label[u,v]==1:
                    pos=pos+G.edges[u,v]['weight']*(x[u]-x[v])
                    alp=alp+1
                else:
                    neg=neg+G.edges[u,v]['weight']*(x[u]-x[v])
                    bet=bet+1
        alp=alp/(alp+bet)
                                 #Positive embeddedness
        bet=bet/(alp+bet)
        if(alp):
            x[v]=x[v]+pos*alp
        if(bet):
            x[v]=x[v]-neg*bet
        if x[v] >= 0.7:
            A[v]=1
    return x,A
```

Figure 2.2 Belief Update

2.2.3 Signed Page Rank Algorithm

The traditional page rank algorithm marked the importance of web pages based on their topological properties. The same algorithm is used on each node to mark their influence maximisation properties. Nodes are ranked in non ascending order of their influence maximisation properties and the most influential nodes are chosen as seed nodes for initiation of ad propagation. The influence of individual v_j on individual v_i as suggested by equation(2.1) is given by $w_{j,i} \cdot (x_{j,t} - x_{i,t})$ and page rank of v_i indicates position and influence of individual v_i in network therefore the SPR should be calculated from presenter standpoint or its neighbours. First we calculate the normalised weighted matrix $\widehat{\mathbf{W}}$ and the matrix of labels L and then using the normalised weighted matrix and matrix of labels we calculate Signed Page-Rank adjacency matrix Y as

$$\mathbf{Y} = d \cdot (\widetilde{\mathbf{W}} * \mathbf{L}) \tag{2.6}$$

where $\widetilde{\mathbf{W}}^*\mathbf{L}$ is the element wise multiplication of matrix $\widehat{\mathbf{W}}$ and \mathbf{L} also known as Hamdard product and $d \in [0,1]$ is the damping coefficient to prevent the signed page ranks from increasing indefinitely.

 $\operatorname{Sort}_{\tau}$ denotes the Signed page-rank of nodes at time τ . We keep updating the Signed Page-rank of all nodes using the equation

$$SPR_{i,r+1} = \sum_{v_j \in H_i^{\text{out}}} (SPR_{i,r} - SPR_{j,r}) \cdot y_{i,j} + (1-d)/N$$
 (2.7)

where $y_{j,i} \in \mathbf{Y}$. This update is done until the **SPR** converges, i.e. $\operatorname{Sort}_{\tau}$ and $\operatorname{Sort}_{\tau+1}$ are equal at a iteration.

After the SPR algorithm converges the Signed Page Rank of nodes at this time is the final Signed Page Rank of the nodes. The most influential (say k) nodes are chosen as seed nodes for initiation of ad propagation.

Below is the code snip of our implementation of our implementation of Signed Page Rank algorithm.

```
def SPR(G,x,label,k=5):
    t=0
    W=np.zeros((n,n))
    L=np.zeros((n,n))#Matrix of labels
    d=0.85#taking value of damping coefficient to be 0.85
    for (u,v) in G.edges():
        W[u][v]=G.edges[u,v]['weight']
        L[u][v]=label[u,v]
    W=normalize(W,axis=1,norm='11')#Normalised weight matrix
    Y=np.multiply(W,L)
    Y*=d#Signed page rank adjacency matrix
    curr_SPR=x#SPR at time t
    next_SPR=[0 for i in range(0,n)]#SPR at time t+1
    curr_sort=[i for i in range(1,n+1,1)]#Signed Page Rank at time t
    temp=sorted(curr_SPR, reverse=True)
    next_sort=[(temp.index(curr_SPR[i]))+1 for i in range(n)]#Signed Page Rank at time t+1
    while next_sort!=curr_sort:
        for i in range(n):
            curr_sort[i]=next_sort[i]
        for i in range(n):
            for j in G.neighbors(i):
                next_SPR[i]+=(curr_SPR[i]-curr_SPR[j])*Y[i][j]+(1-d)/n
        temp=sorted(next_SPR, reverse=True)
        for i in range(n):
            next_sort[i]=temp.index(next_SPR[i])+1
        for i in range(n):
            curr_SPR[i]=next_SPR[i]
            next_SPR[i]=0
    S=[]#Set containing the initial seed nodes
    for i in range(n,n-k,-1):
        S.append(curr_sort.index(i))
```

Figure 2.3 Signed Page Rank

Chapter 3

Dataset

For generating a dataset, to test our program, we used the **Networkx** library of python to generate a directed graph. We used a function **gnp_random_graph** which takes the number of nodes, the probability for edge creation between nodes, the seed value for a random number generator, and a bool value to make it a directed graph as arguments.

Seed helps in maintaining the graph if we re-run the program with no other arguments changed. We have kept the probability for edge creation between nodes low (0.1), so as to see the maximum effect of selected k nodes on the graph.

Similarly for assigning initial belief of every node we again used random function.

At last for assigning a time period to every node, we randomly generated the starting point from 0 to 50 and then added that particular value to other generated value from 0 to 10 to get the endpoint of a particular node. Thus making the values scattered from 0 to 50 but at the same time, no node will have a time period greater than 10. We have provided sufficiently large time window to every node to ensure that every node receives at least one recommendation within its recommendation cycle.

For now, we have made 0.7 as the default threshold value of the belief after which that person gets influenced.

Values which are subject to change are:

- 1. Total number of nodes(n).
- 2. The number of initial seed nodes(k).
- 3. The probability for edge creation between nodes.

```
import random
import networkx as nx
n=100#The number of nodes
random.seed(10)
G=nx.gnp_random_graph(n,0.1, seed=10, directed=True)#generating the graph
label={}#dictionary of labels
c=0;
for (u,v) in G.edges():
    G.edges[u,v]['weight']=random.uniform(0,1)
   if(c%2):
        label[u,v]=1
        label[u,v]=-1
    c+=1
x=[]#Initial Beliefs of all nodes
TP=[]#Time window or recommendation cycle
for i in range(n):
   x.append(random.uniform(0,1))
   TP.append([random.randrange(0,50)])
    TP[i].append(TP[i][0]+random.randrange(0,10))
k=7#Value of initial number of seed nodes
```

Figure 3.1 Dataset

Chapter 4

Results and Conclusion

We generated the synthetic dataset as discussed in the Dataset chapter and then run our Signed Page Rank and Ad propagation algorithm to get the below result. We set the value of initial number of seed nodes to be 7.

When we take the probability of a directed edge between two nodes to be 0.1, below is the result.

Figure 4.1 Result1

Result 1 shown above shows that when the dataset is generated keeping the probability of a directed edge between two nodes to be 0.1 and rest of the variables constant as discussed in the chapter 3, when the initial number of seed nodes taken to initiate

the ad propagation equal to 7, the total reach of advertisement is 57 out of 100 individuals and the final attitude status of each node is shown by the list containing 0's and 1's. 1 attitude status represents that this node has accepted the advertisement and is now a seed node whereas, 0 attitude status represents that this node has not accepted the advertisement. The 0-1 attitude can be interpreted as the reach of ad. The more the number of 1's the more is the reach of the advertisement.

Now, when we take the probability of a directed edge between two nodes to be 0.4, thereby making the graph more connected and even more dense, below is the result.

Figure 4.2 Result2

Result 2 shown above shows that when the dataset is generated keeping the probability of a directed edge between two nodes to be 0.4 and rest of the variables constant as discussed in the chapter 3, when the initial number of seed nodes taken to initiate the ad propagation equal to 7, the total reach of advertisement is 73 out of 100 individuals (73 individuals have accepted the ad recommendation out of 100 in the social network) and the final attitude status of each node is shown by the list containing 0's and 1's.

As is quite clear from our result and also very obvious, that more connected or more dense the network more is the number of 1's in the list representing the final attitude status (attitude status of the nodes after the information propagation terminates) of the nodes and more is the spread of the advertisement and for every node the number of nodes it receives the recommendations from is increased to 4 fold as we increase the probability of an edge between two nodes from 0.1 to 0.4.

Bibliography

[1] X. Yin, X. Hu, Y. Chen, X. Yuan, and B. Li, "Signed-pagerank: An efficient influence maximization framework for signed social networks," *IEEE Transactions* on Knowledge and Data Engineering, 2019.