ABSTRACT

The purpose of our project is to detect k-size subset of nodes in a social network that can maximize the influence spread in the network using node connectivity, node activity and node recency. Node connectivity is the measure of the connection between the nodes and how does that connection contribute in influence spread. Node activity tells about the actions of the particular node and in the same way node recency tells about how recently the node has interacted in the past. Regarding recency, we observe that a node which is actively interacting now is more important than the one who is inactive in the present scenario. All these factors together helps in finding the influencer more effectively. In the recent past node influence has always been talked in terms of connectivity but in this project we will be including node activity and node recency as well. We have tried to use more efficient heuristics to incorporate the above three parameters. We have applied the UAC Rank algorithm for finding the top or seed nodes in terms of connectivity and activity. Further we have used the time factor to find out the nodes with the most recent interactions.

Here we are finding the influencers who can help in digital marketing in social media. One of the basic aspects of digital marketing is finding out such influencers who will provide a good marketing to their products on social media so that the product market grows exponentially. In our project we are contributing in finding out those influencers on social media.

INTRODUCTION

Influence maximization has attracted a lot of attention due to its numerous applications, including diffusion of social movements, the spread of news, viral marketing and outbreak of diseases. The objective is to discover a group of users that are able to maximize the spread of influence across a network.

During the last decade, online social networking sites (e.g., Facebook, Twitter, LinkedIn, Tumblr etc.) and relevant smartphone applications have caused a remarkable growth of research on social networks. This led to the development of many applications of social networks of which a rich body of studies has been classified as *the analysis of influence or information diffusion in social networks*.

Influence propagation studies have found applications in various fields. From studying human, animal or even plant epidemics to viral marketing, social media analytics, spread of rumors, expert finding, recommendation systems, etc. A key task in order to understand information and influence diffusion is the identification of vital nodes that play a significant role in these cases. Such nodes may allow us to control the spread of an epidemic, to predict successful scientists and scientific publications based on co-authorship and citation networks, to design influential advertisements for new products etc.

PROBLEM DEFINITION

A social network can be represented as a graph G(V,E), where the set of vertices V represents the users and the set of edges E represents the connections between the users. Social networks could be both directed, like Twitter, and undirected, like Facebook. Every edge in a directed network has a source and target. In case of undirected networks, edges have no direction associated with them and there is only one edge between a pair of nodes. So an edge between nodes u and v would be considered as a two-way communication channel between the two nodes. For successful viral marketing, the first step is to assess the influential capabilities of users, and finally select an initial set of adopters who can help maximize the spread. For a given network G, the aim is to find a k-node set S (such that $S \subset V$) that is likely to provide maximum spread of information over the network. The set S is known as the seed set, and k is the number of initial adopters. Let $\delta(S)$ denote the total number of individuals expected to adopt the information, then the goal of influence maximization is to find a k-node set S for which $\delta(S)$ is maximum. When using a social network for viral spreading of information, the fact that how fast a node can spread information is as important as the number of connections it has.
So we define our problem as finding those k node set S using three parameters node connectivity, node activity and node recency.

PROPOSED METHOD

We use a node’s connectivity,activity and recency to find out how influential it is.A weighted social network graph is generated based on each node’s activity and connectivity.

Node Activity Weights (NAWs) is assigned to the nodes based on the number of communications they have initiated in a given time span , and Edge Activity Weights (EAWs) to the edges based on the number of interactions carried out over them by their corresponding source node .Out degree of each node is also calculated.Then we rank (using UAC-rank) each node in the network based on it's NAW and outdegree.

Node Activity Weight(u) = $\sum$ Interactions from node u to v

Edge Activity Weight (u, v) = $\sum$Total interactions from node u to v over edge E(u, v)

UAC-Rank(v) = (α * Node Activity Weight(v)) + (β * Outdegree(v))

where, α=β=0.5

Algorithm 1: UAC-Rank(G,k)

Input parameters: Graph G = (V, E),

size of initial seed set k

Output parameters:initial_seeds[ ] : k-size set of selected nodes

Algorithm:

1: For each node,compute Node Activity Weight (NAW) using (1), out-degree (OD) value, and UAC-Rank using (3);

2: Compute Edge Activity Weight (EAW) for each edge in the network using (2);

3: Initialize seed_counter = 1;

4: WHILE seed_counter<= k DO

5: selected_node = MAX(UAC-Rank(vi));

6: initial_seeds[seed_counter] ← selected_node;

7 Find predecessors of selected_node and store them as neighbors_of_selected_node;

8 FOR neighbor in neighbors_of_selected_node

9: I F neighbor not in initial_seeds

10: N AW(neighbor) = NAW(neighbor) –EAW(neighbor, selected_node);

11: O D(neighbor) = OD(neighbor) -1

```
      E ND IF
    Recompute UAC-Rank for each node as per (3);
    E ND FOR
   seed_counter = seed_counter + 1;
END WHILE
```

Using the above algorithm we get k-seed nodes based on activity and connectivity.Now,we use recency on these k seed nodes to find the most influential initial seeds.

Formula used:

The formula used for finding out the recency factor is-

Let e1 be the maximum value in the time range

Let d1 be the value at that point

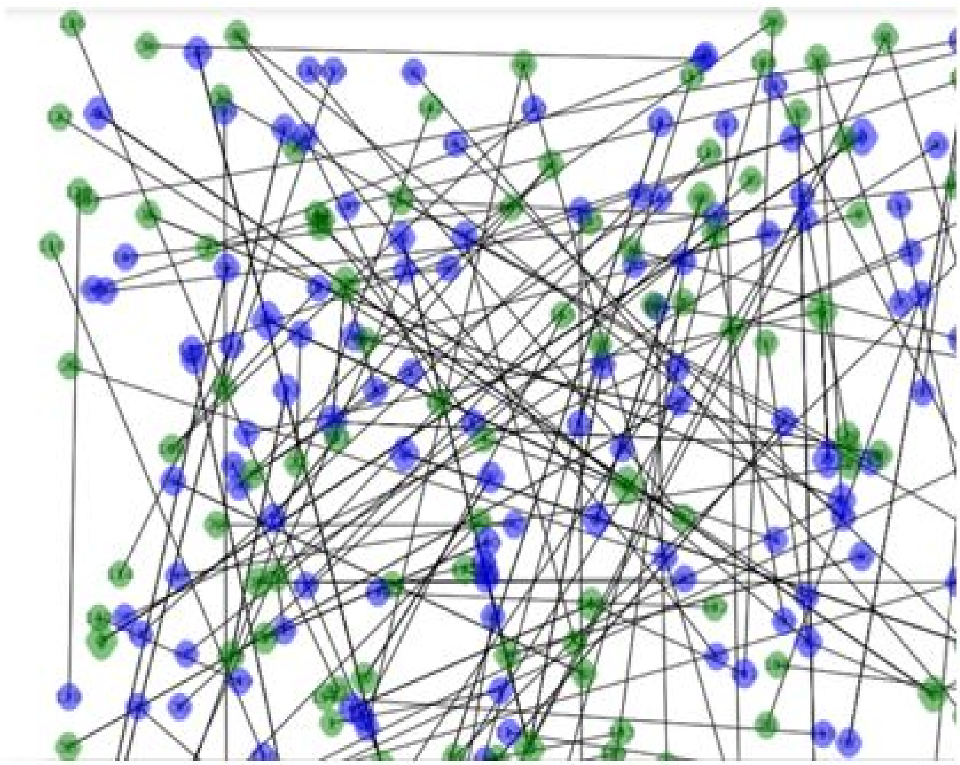
t(i)=1 if {e1-d1 <75% of e1}

t(i)=((e1-d1)/25% of e1)pow2+1 if {e1-d1>75% of e1 &&e1-d1<25%of e1}

t(i)=0 if {any other condition}

DATASET USED

Facebook wall posts—A network representing a subset of posts written by a user on another user's wall. There are 46,952 nodes with 274,086 static edges and 876,993 temporal edges wherein each edge represents a post. As a user may write multiple posts to the same user, multiple edges connecting same pair of nodes exist in the network (available at http://konect.uni-koblenz.de/networks/facebook-wosn-wall).

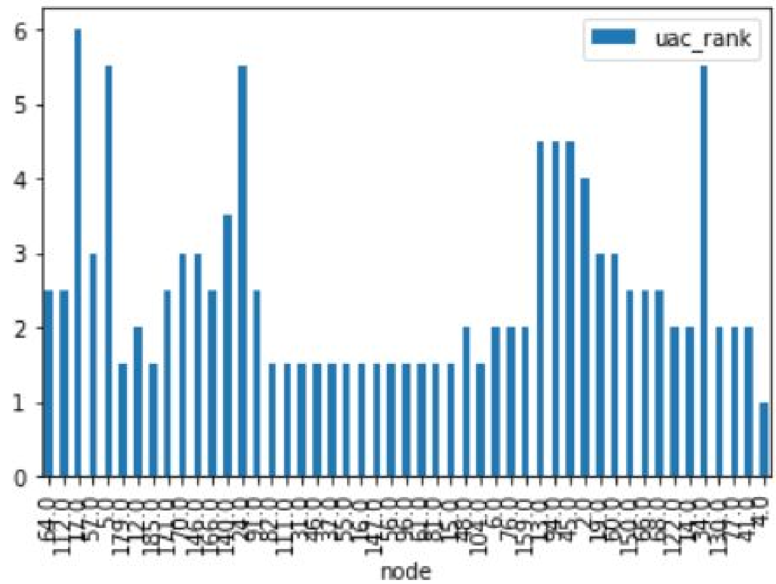


NETWORK GRAPH

EXPERIMENTAL RESULTS

	NAW	OUT	uac_rank	node
0	8.0	4.0	6.0	17.0
1	9.0	2.0	5.5	34.0
2	9.0	2.0	5.5	5.0
3	7.0	4.0	5.5	24.0
4	5.0	4.0	4.5	13.0
5	8.0	1.0	4.5	94.0
6	6.0	3.0	4.5	45.0
7	7.0	1.0	4.0	2.0
8	5.0	2.0	3.5	140.0
9	3.0	3.0	3.0	70.0
10	3.0	3.0	3.0	19.0
11	5.0	1.0	3.0	146.0
12	4.0	2.0	3.0	57.0
13	3.0	3.0	3.0	60.0
14	3.0	2.0	2.5	91.0
15	3.0	2.0	2.5	64.0

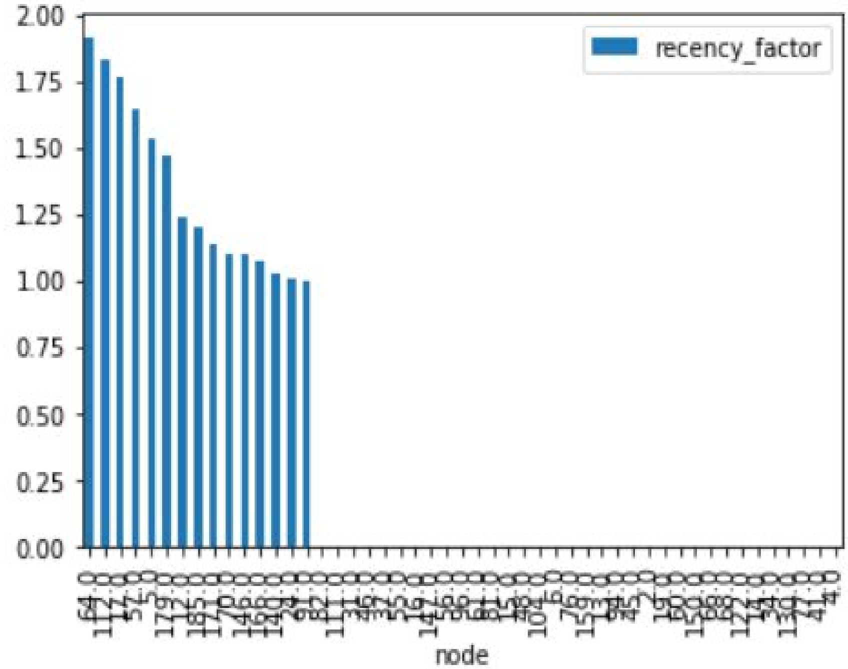
TOP 10 NODES AFTER UAC RANK



GRAPHICAL REPRESENTATION OF NODES AND UAC RANK

	NAW	OUT	uac_rank	node	recency_factor
15	3.0	2.0	2.5	64.0	1.913904
20	3.0	2.0	2.5	112.0	1.827684
0	8.0	4.0	6.0	17.0	1.762832
12	4.0	2.0	3.0	57.0	1.645144
2	9.0	2.0	5.5	5.0	1.539111
39	2.0	1.0	1.5	179.0	1.474897
29	3.0	1.0	2.0	12.0	1.239092
38	2.0	1.0	1.5	185.0	1.201117
21	3.0	2.0	2.5	171.0	1.137139
9	3.0	3.0	3.0	70.0	1.097780
11	5.0	1.0	3.0	146.0	1.097566
16	3.0	2.0	2.5	166.0	1.069796
8	5.0	2.0	3.5	140.0	1.023237
3	7.0	4.0	5.5	24.0	1.009859
14	3.0	2.0	2.5	91.0	1.000000

TOP 10 NODES AFTER RECENCY FACTOR



GRAPHICAL REPRESENTATION OF NODES WITH RECENCY FACTOR