Influence maximization through user interaction modeling

David Oriedi, Cyril de Runz, Zahia Guessoum, Amine Ait Younes, Henry Nyongesa, Amine Younes

Influence Maximization Through User Interaction Modeling

David Oriedi
University of Reims
Champagne-Ardenne, CReSTIC
Reims, France
oriedi.do@gmail.com

Cyril de Runz
University of Tours, BdTln, LIFAT
Blois, Fance
cyril.derunz@univ-tours.fr

Zahia Guessoum
University of Reims
Champagne-Ardenne, CReSTIC
Reims, France
zahia.guessoum@univ-reims.fr

Amine Aït Younes
University of Reims
Champagne-Ardenne, CReSTIC
Reims, France
amine.ait-younes@univ-reims.fr

Henry Nyongesa
Dedan Kimathi University of Technology
Nyeri, Kenya
henry.nyongesa@dkut.ac.ke

ABSTRACT

A majority of influence maximization models in social networks in literature are based on a seminal work by Kempe *et al.*, in which two classic influence models were proposed i.e *Linear Threshold Model* and *Independent Cascade Model*. However, these two models use assumed values to model influence and influence propagation in social networks. This may lead to inaccurate approximation of influence. In this work, we model influence from actual social actions among members of a social network through a proposed algorithm - ***Selective Breadth First Traversal*** - that efficiently generates an optimal seed set for influence maximization. Experimental results on real data show that our approach provides an improvement over a number of traditional influence maximization algorithms.

CCS CONCEPTS

• **Networks → Online social networks**;

KEYWORDS

Social network, Influence maximization, Interactions, Social actions

ACM Reference Format:
David Oriedi, Cyril de Runz, Zahia Guessoum, Amine Aït Younes, and Henry Nyongesa. 2020. Influence Maximization Through User Interaction Modeling. In *The 35th ACM/SIGAPP Symposium on Applied Computing (SAC '20), March 30-April 3, 2020, Brno, Czech Republic.* ACM, New York, NY, USA, Article 4, 3 pages. https://doi.org/10.1145/3341105.3374080

1 INTRODUCTION

Online Social Networks are crucial in timely transmission of information to large groups of people. Furthermore, such networks have been used to model human relationships using graph concept in which people are represented using nodes while the relationships between them are represented as links. Typically members in a network share content of various types over the network from time to time and such contents attract reactions such as replies, retweets or favorites from other members. Some members attract more reactions from their neighbors than others and are therefore regarded as more influential. Determination of influential members in social networks has become an important research subject in the analysis of Social Networks.

Peng *et al.*[10] defines social influence as a relationship established between two entities(influencer and influencee) for a specific purpose. Influence maximization is the problem of finding a small subset of nodes (seed nodes) in a social network that could maximize the spread of influence optimally through the network [4]. At the core of social network analysis is the need to appropriately model both the entities interacting on the network and the form taken by that interaction.

Several works exist on influence maximization [5, 8, 10]. But Kempe *et al.* [7] showed that solving an influence maximization problem is NP-hard. As a result of this, most of the existing models in literature are based on extensions of either *Independent Cascade* (IC) model [13] or *Linear Threshold* (LT) model [6]. However, these models and their extensions rely on assumed probabilistic values to represent influence [9, 11]. Since the assumed values are mostly homogeneous, they may not accurately approximate influence as it occurs in real life [1]. This may lead to biased computation of influence. Some other models are based on Centrality measures, like in [12]. Nevertheless, even this category of models do not take into account user actions, and when they do, like in [2], they ignore the twin aspects of both centrality and spectrality.

In this paper, we address all these issues. We model social influence based on social actions (such as tweets, replies, likes or mentions) that take place among members of a social network. In doing this, we depart from the more common approach where prederdetermined influence threshold values are assigned to nodes or uniform probabilitic values are used to represent edge weights, as happens with LT and IC models respectively. To do this, we propose a new algorithm, called *Selective Breadth First Traversal* that generates an optimal seed set able to maximize influence by quantifying and assigning specific weights to social actions carried out among network nodes. Our definition of influence is based on both the relationship that a node has with its neighbors as well as the relationship that its neighbors has with it. To the best of our

Table 1: Symbols and their meanings

Symbol	Meaning
i,j	*Social network members i and j*
a_k	*A type of social action*
n_a	*Number of social action types*
α_{ak}	*Assigned weight of a social action*
$N_{ak}(i,j)$	*Social actions by i on posts of j*
v_i, v_j	*Represents nodes i and j on the social graph*
(i,j)	*Denotes a directed edge from* v_i *to* v_j
e_{ij}	*Weighted directed edge from* v_i *to* v_j
N_{pi}	*Total number of published contents*
$W_{Na}(i,j)$	*Total weighted social action value of* v_i *on* v_j

knowledge, this is the first work that brings all these issues together from the perspective of influence maximization.

This paper is organized as follows. In section 2, we explain our model and algorithms. Section 3 provides a summary of experimental results with comparison to several Centrality based approaches. Concluding remarks are in section 4.

2 PROPOSED APPROACH

The problem of influence maximization as put forward by Kempe *et al.*[7] involves generating a subset S of k nodes such that $|S| = k$ and the overall expected number of influenced nodes $\sigma(S)$ is maximized.

2.1 Preliminaries

We define a directed, weighted graph $G = (V, E, W)$ in which V is a set of vertices $V = \{v_1, v_2, ..., v_n\}$, E is an edge set $E = \{(v_i, v_j)|$ an edge exists from node v_i to $v_j\}$. W represents the set of edge weights $W = \{e_1, e_2, ..., e_n\}$. An edge weight e_{ij} indicates the strength of the relationship between node v_i and node v_j. The relationship strength is a reflection of how frequently a pair of nodes interact through likes, replies, mentions or retweets. Throughout this paper, the notations in Table 1 and the definitions that follow will apply.

As shown in equation (1), we have adopted the aggregation used in [2] to combine the different interactions according to their type. The total weighted social action value by a node v_i on the posts of a node v_j is the sum of all social actions each multiplied by its assigned weight. Given that there is a directed link from node v_i to node v_j, this value is given as:

$$W_{Na}(i,j) = \sum_{k=1}^{n_a} \alpha_{ak} N_{ak}(i,j) \tag{1}$$

2.2 Proposed Definitions of Influence Power

Our definition of influence is partly inspired by the work done by Azaouzi and Romdhane [2] in which they recognize the role played by the type, number and weight of social actions among network members as a major component of influence definition.

However, their index only expresses node $i's$ interaction with node j but does not give us a sense of how node i interacts with the rest of its neighbors. It is therefore necessary to find out what proportion of node $i's$ interactions with node j account for node $i's$ interaction with all its neighbors. There is also a need to express what portion of node $j's$ posts actually attracts reactions from node i in comparison to the other posts by all neighbors of node i. To address this, we propose to put together the influence that both nodes i and node j independently have on their immediate neihborhood.

We suggest that the influence of node j over node i is dependent on two things i.e:

(1) how much of node $i's$ social actions cover node $j's$ posts compared to how much node $j's$ neighbors react to $j's$ posts and;
(2) how much of node $i's$ social actions to the posts of its neighbors account for its actions on node $j's$ posts.

These two ideas form the basis of our definition of influence. Using I_1 to represent the normalized value of (1) gives:

$$I_1(i,j) = \frac{W_{Na}(i,j) - min_{k \in N'(j)}(W_{Na}(k,j))}{max_{k \in N'(j)}(W_{Na}(k,j)) - min_{k \in N'(j)}(W_{Na}(k,j)} \tag{2}$$

Similarly, we use I_2 to represent the normalized value for (2):

$$I_2(i,j) = \frac{W_{Na}(i,j) - min_{k \in N'(i)}(W_{Na}(i,k))}{max_{k \in N'(i)}(W_{Na}(i,k)) - min_{k \in N'(i)}(W_{Na}(i,k)} \tag{3}$$

In both cases, whenever $min = max$, the value is set to 1. In order to give a relative importance to each value we associate each of them with a dumping factor β and combine them in I_3 as follows:

$$I_3(i,j) = \beta \cdot I_1(i,j) + (1-\beta) \cdot I_2(i,j) \tag{4}$$

Formally, the value $I_3(i,j)$ is the *influence* of node j over node i or more specifically, the edge weight of a directed link from node i to node j.

We propose to get the Influence Power for a node i, $I_p(i)$, by applying equation 4 on each of the incoming edges of i, get the sum and divide the result with the maximum following among its neighbors with incoming links. This is shown in equation 5:

$$I_p(i) = \frac{\sum_{j \in follower(i)} I_3(j,i)}{max_{k \in N'(i)}(|follower(k)|)} \tag{5}$$

2.3 Influence Spread

The influence spread of a node is a commonly used metric for comparing the performance of influence maximization algorithms. It refers to the total number of nodes reachable directly or indirectly from a candidate seed node. Numerous models exist in literature for the computation of influence spread. Our approach provides four steps:

(1) We compute the influence power of each node according to equation 5.
(2) We generate a set of nodes each of which has an influence power value that is higher than the mean of its neighborhood. This set becomes the set of influential nodes.
(3) From step 2, we generate an ordered set of seed nodes i.e the top k most influential nodes.
(4) We apply our algorithm, *Selective Breadth First Traversal* algorithm, outlined in Algorithm 1, in order to determine influence spread for the k nodes given by step 3.

Algorithm 1 *Selective BFT for Influence Maximization*

Require: *a weighted directed graph* $G' = (V', E', W')$ *with IP of the top* k *significant nodes*
Ensure: *Influence Spread values for top* k *Significant Nodes*

for $i \leftarrow 1$ **to** $|S|$ **do**
  $Q \leftarrow \emptyset$ //initialize queue Q
  $S_p(i) \leftarrow 0$ //initialize spread for node i
  $visited(i)$
  $Q.enqueue(i)$
  while $(\neg empty(Q))$ **do**
    $i = Q.dequeue()$
    for each $j \in adjacencylist(i)$
    if $(edge.in()) \wedge (\neg visited())$ **then**
      $Q.enqueue(j)$
      $visited(j)$
      $j.parent = i$
      $S_p(i) = S_p(i) + 1$
    end if
  end while
  return $S_p(i)$
end for

3 EXPERIMENTATION

The model was developed in Java language on Netbeans IDE. The experiments were run on the real dataset *C-Elegans*[1] used in [12], on a desktop computer with Windows 10, 8GB RAM, 1TB Hard Disk and Intel Core i7 2.40 GHz processor. *C-Elegans* is a directed and weighted neural network of the nematode worm C.elegans. It is composed of 453 nodes and 2,025 edges. For our purposes, the synapse movement has been represented as a reply from one node to the other. In this dataset, there is only one type of action, therefore its weight is set at 1. The value of the dumping factor β was set at 0.85.

The performance of our model is compared with results from four centrality-based algorithms as presented in [12] (Degree, Top-k, MC-Greedy, IV-Greedy) and the well known PageRank algorithm [3]. The results in terms of influence spread, illustrated in figure 1, show that our approach obtains the best performance. Our approach identifies nodes that are better spreaders in the graph. For instance, for a seed set of 15, our approach is able to activate about 93% of the nodes while the best of the others (PageRank) is around 53%. This is because we select influential nodes based on their ability to engage the neighborhood through social actions. This selection chooses nodes that are locally pertinent (according to their neighbors' influence power), and not just the set of the best nodes in terms of the index, which can be closed into the graph. Secondly, the influence power takes into account several aspects of interaction that are not considered by the other models.

4 CONCLUSION

In this paper, we proposed the *Selective Breadth First Traversal* algorithm that computes a seed set for influence maximization. We have argued that the *IC* and *LT* models may lead to inaccurate

[1] source: https://snap.stanford.edu/data/C-elegans-frontal.html

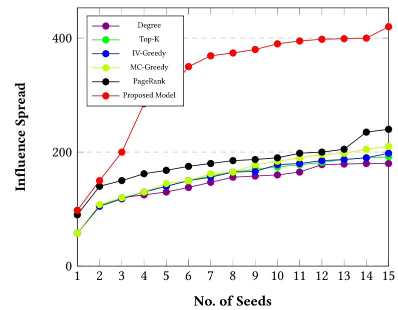

Figure 1: Results on *C.elegans* dataset.

approximation of influence. Our submission is that the real source of influence is from the social actions that take place during interactions among network members. For future work, we will propose a deeper study of our approach properties. We will also investigate the identification and separation of the effects of malicious artificial social user applications that mimick real user interactions thereby falsely increasing influence scores. Another interesting perspective of this research would be to work on the influence dynamics as well as a scaling mechanism to cope with very large social networks through distributed and parallel computing.

REFERENCES

[1] Sinan Aral and Paramveer S Dhillon. 2018. Social influence maximization under empirical influence models. *Nature human behaviour* 2, 6 (2018), 375.
[2] Mehdi Azaouzi and Lotfi Ben Romdhane. 2018. An Efficient Two-Phase Model for Computing Influential Nodes in Social Networks Using Social Actions. *Journal of Computer Science and Technology* 33, 2 (01 Mar 2018), 286–304. https://doi.org/10.1007/s11390-018-1820-9
[3] Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems* 30, 1-7 (1998), 107–117.
[4] Wei Chen, Yajun Wang, and Siyu Yang. 2009. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 199–208.
[5] Evren Güney. 2019. On the optimal solution of budgeted influence maximization problem in social networks. *Operational Research* 19, 3 (2019), 817–831.
[6] Furkan Gursoy and Dilek Gunnec. 2018. Influence maximization in social networks under Deterministic Linear Threshold Model. *Knowledge-Based Systems* 161 (2018), 111–123.
[7] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 137–146.
[8] Alan Kuhnle, Md Abdul Alim, Xiang Li, Huiling Zhang, and My T Thai. 2018. Multiplex influence maximization in online social networks with heterogeneous diffusion models. *IEEE Transactions on Computational Social Systems* 5, 2 (2018), 418–429.
[9] Yuchen Li, Ju Fan, Yanhao Wang, and Kian-Lee Tan. 2018. Influence maximization on social graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering* 30, 10 (2018), 1852–1872.
[10] Sancheng Peng, Yongmei Zhou, Lihong Cao, Shui Yu, Jianwei Niu, and Weijia Jia. 2018. Influence analysis in social networks: a survey. *Journal of Network and Computer Applications* 106 (2018), 17–32.
[11] N Sumith, B Annappa, and Swapan Bhattacharya. 2018. A holistic approach to influence maximization in social networks: STORIE. *Applied Soft Computing* 66 (2018), 533–547.
[12] Wenjun Wang and W Nick Street. 2018. Modeling and maximizing influence diffusion in social networks for viral marketing. *Applied network science* 3, 1 (2018), 6.
[13] Zheng Wen, Branislav Kveton, Michal Valko, and Sharan Vaswani. 2017. Online Influence Maximization under Independent Cascade Model with Semi-Bandit Feedback. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 3022–3032.