A Study on Maximising the Influence of Social Network Nodes Based on a Comprehensive Influence Hierarchy Model: Structural Analysis and Optimisation of Information Dissemination

Zhongbao Zhou *

School of Management, Harbin Institute of Technology, Harbin, China, 150006

* Corresponding Author Email: zhouzhongbaomail@163.com

Abstract. The work of network power is related to the construction of public opinion and the management of information dissemination, which is crucial to the development of industry and information technology. It has been a key issue to study how to maximise positive information influence in social network information dissemination. This study focuses on network structure and node influence distribution, and proposes a comprehensive influence level model (CILM) and an innovative node influence factor. The factor integrates the influence level and node degree attributes to provide a basis for node influence assessment. Further, the Neighbour Influence Traversal Algorithm (NITA) and Influence Cascade Transfer Algorithm (ICTA) are designed to optimize the node influence calculation through the network structure and information dissemination characteristics. Experimentally, it is proved that CILM and related algorithms are significantly better than other methods in terms of accuracy and running time, and provide an effective tool for the optimisation of information dissemination in social networks.

Keywords: Social Networks; Influence Maximisation; Information Dissemination Management. CILM; NITA.

1. Introduction

The importance of digital construction is highlighted, and the prosperity of social media in information dissemination has triggered scholars to study its influence and dissemination [1]. Influence is the ability to change the thoughts and actions of others by being accepted, a key factor in assessing the importance of users in social networks. The influence maximisation problem originating from online marketing is dedicated to finding a set of seed nodes to maximise their influence spread. Due to the huge size of the network, the running time and convergence of the algorithm become considerations. To this end, this paper proposes a comprehensive algorithm that demonstrates good runtime and convergence in real network tests, which is valuable in applications.

In related work, the diffusion process of large complex networks has been the focus of research, initially studied from an algorithmic point of view by Domingos et al. The contribution of Kempe et al. pioneered the study of greedy algorithms, which led to a wide range of subsequent studies. Early studies were mainly divided into greedy algorithms, which are computationally expensive, and heuristic algorithms, which are fast but sacrifice accuracy. Node influence evaluation methods based on network structure involve centrality indicators, but have limited effect on large-scale networks. The models and algorithms in this paper are developed based on social network features and outperform traditional heuristic algorithms [2-4].

The main contributions of this paper include defining node influence factors, covering influence level and degree, as a way to quantify node influence level. The CILM model is proposed to assess node influence, and its accuracy is optimised by the new algorithms NITA and ICTA. Extensive experiments verify that CILM+ has better results in undirected network graphs, which is especially suitable for large-scale networks, and the results of simulation experiments are more realistic by comprehensively considering the basic structure of the network and the characteristics of information propagation.

The research idea includes modelling the social network data, obtaining the node influence factor, and obtaining the initial node influence level through the CILM model. In terms of network structure, the NITA algorithm is introduced to optimise the accuracy of CILM; in terms of information propagation characteristics, the ICTA algorithm is introduced to carry out iterative calculations through "simulated flow". Finally, IFR is used to get the final ranking of nodes and determine the most influential nodes. The structure of this thesis is consistent with the research idea, as shown in Figure 1 below:

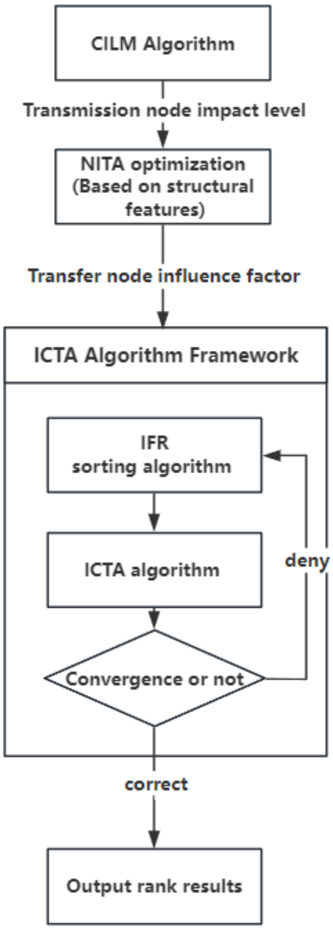

Figure 1. Technology road

2. Node influence modelling

2.1. Overview of the CILM model

Definition 1 [Impact factor] $I_i < l_i, d_i >$ is defined as the influence factor of node i and contains two attributes, l_i and d_i, which represent the influence level (influence level) and degree (degree) of node i, respectively. Where the influence level l_i is determined by the situation of neighbouring nodes [5].

Definition 2 [Influence Factor Update Rule] The update rule for a node's influence factor can be formalised as a multivariate function:

$$I = Update(I_{i1}, I_{i2}, I_{i3}, \dots, I_{ik}) \tag{1}$$

where I_{ik} is the influence factor of the k neighbour of node i and Update denotes the update rule for the factor.

2.2. The CILM model

The CILM model is essentially a model that obtains the true influence factor of each node based on the network structure. The model is inspired by LPA [6], i.e., the idea of factor propagation. In this paper, the idea is mainly reflected in the fact that the initial influence factor of each node is like a data source, and the new influence factor of the target node is introduced through the initial influence factor. Meanwhile, in the propagation process, the factor of each node will also be propagated through the propagation path of network information. In LPA, the update rule is defined to be calculated based on the similarity of neighbouring nodes. In CILM, this paper adopts the quality and quantity of neighbours to update the node influence factor.

Definition 3 [Hi-I Neighbour Nodes] Neighbour nodes whose influence level is not less than their own influence level are defined as Hi-I (High-Influence) Neighbour nodes. $K_i^{l_i}$ denotes the set of Hi-I nodes in the neighbouring nodes of node i that have an influence level not less than l_i.

$$K_i^{l_i} = \{v_j | l_j \geq l_i; v_j \in NB(i)\} \tag{2}$$

Where $NB(i)$ is the set of neighbours of node i.

Definition 4 [Com-I Neighbour Nodes] Neighbour nodes whose influence level is less than their own influence level are defined as Com-I (Common-Influence) Neighbour Nodes. $Q_i^{l_i}$ denotes the set of Com-I nodes in the neighbourhood of node i whose influence level is less than x.

$$Q_i^{l_i} = \{v_j | l_j < l_i; v_j \in NB(i)\} \tag{3}$$

In this paper, we set the condition for upgrading the influence level of node i: the influence level of node i can be upgraded when the number of Hi-I neighbouring nodes of node i is greater than $l_i(|K_i^{l_i}| > l_i)$. That is, the influence level l_i is determined by the number of Hi-I neighbour nodes. It can be found that when the influence level of a node is higher, the number of its Hi-I neighbours is less, and the upgrade is more difficult. That is, the expression "the higher the influence level, the more difficult the upgrade" is closer to the actual network situation [7-8].

The following is a graphical representation of the upgrade rule. As shown in Fig. 2, let the influence level of these four nodes in the network be 2, i.e., $l_i = l_j = l_k = l_o = 2$. For node i, we have: $l_i = 2$, so, $|K_i^{l_i}| = |K_i^2| = 3$. And because $|K_i^2| = 3 > l_i$, l_i can be upgraded to 3.

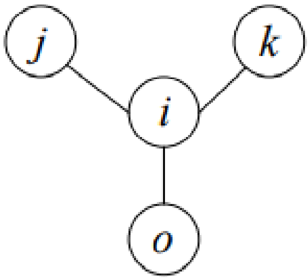

Figure 2. Demonstration of CILM escalation rules

Therefore there is the following definition:

Definition 5 [Iterative update rule for impact level] The iterative update rule for impact level l is shown in equation (5):

$$l_i' = \begin{cases} l_i + 1 & if(|K_i^{l_i}| > l_i) \\ l_i & otherwise \end{cases} \tag{4}$$

Definition 7 [Influence Factor Ranking] With node influence level as the primary attribute and degree as the secondary attribute, when two nodes have equal influence, the higher the influence level of a node or the higher the degree of a node, the more influential it is.

$$\begin{cases} l_i > l_j \\ l_i = l_j \ and \ d_i > d_j \end{cases} \Rightarrow I_i > I_j \tag{5}$$

With the above, the framework of the basic influence assessment algorithm corresponding to CILM has been roughly completed, which is described in Algorithm 1. The algorithm mainly consists of the following steps:

Step 1: Traverse the graph network once and initialise the node's influence level according to the degree of each node (lines 1 to 8).

Step 2: Determine the number of Hi-Is of each node by comparing each node's own influence level with that of its neighbours according to Definition 3 (lines 12 to 16).

Step 3: According to Definition 6, combine the number of Hi-Q neighbours of each node calculated in the previous step, and update the influence level of each node by Equation (6) (lines 17 to 19).

Step 4: Skip to Step 2 until the influence level of each node converges.

2.3. Optimisation of the algorithm

In this paper, the CILM model is optimised by Neighbour Influence Traversal Algorithm (NITA, Neighbour Influence Traversal Algorithm), which needs to be defined with the relevant formulas before formally introducing the NITA algorithm.

In order to further strengthen the influence of Com-I neighbouring nodes and the difference of Hi-I neighbouring nodes, the influence level of each node needs to be further adjusted:

$$l_i'' = (1 + \delta) * l_i' \tag{6}$$

Where l_i' denotes the original impact level of node i after calculation by the basic CILM algorithm, l_i'' denotes the optimised impact level of node i, and δ is the optimisation function.

Based on the characteristics of the network's own node distribution and the incremental and decremental nature of the optimisation process, this paper will use the logarithmic function for the construction of the optimisation function

$$\delta = log_\mu(1 + \lambda) \tag{7}$$

Where λ is the optimisation factor and for ease of computation in this paper the value of μ is set to 10.

Let NB(i) be the set of neighbours of node i. It is easy to know that for node i, Com-I neighbour nodes are complementary to Hi-I neighbour nodes. In this paper, we consider the influence of both for node i by summation, since the influence of Com-I and Hi-I on i is not exactly the same, so in this paper, we achieve the differentiation of Com-I and Hi-I neighbour nodes by the way of weights in summation. In addition, when updating the influence level of node i, it is still necessary to follow the rule of "the higher the influence level is, the more difficult it is to upgrade" mentioned in the previous section, therefore, in the later section of this paper, when constructing the optimisation factor, l_i is introduced into the denominator position of the adjustment formula. From the above, this paper constructs the optimisation factor as follows

$$\lambda = \frac{\alpha * \sum_{v_j \in Hi-I_i(j)} l_j' + \beta * \sum_{v_j \in Com-I_i(j)} l_j'}{\left|K_i^{l_i'}\right| * l_i'} \tag{8}$$

Where $Hi - I_i(j)$ is the set of high-influence neighbour nodes of node i, and $Com - I_i(j)$ is the set of general-influence neighbour nodes of node i. l_j' is the original influence level of node j after the calculation of the basic CILM algorithm. α and β are the weight coefficients, which reflect the influences of different types of neighbour nodes on node i, respectively.

3. Influence cascade transfer algorithm

The ICTA algorithm was developed based on the characteristics of information propagation in networks, through which a node is considered more comprehensively than just comparing among a node and its neighbours [9-10].

The algorithm defines the flow between nodes as $p(v_j, v_i) * l(i)$, where $p(v_j, v_i)$ is a fundamental property of the social network graph, i.e., the probability of propagation between node i and node j. For the lower ranked node, part of his influence level will "flow" to his neighbour nodes, while the influence level of his neighbour nodes will also partially flow to their respective neighbour nodes, and so on. If the flow is labelled according to the source of the node, it is eventually found that this process simulates the process of information propagation in the network well

Based on the above introduction, this paper proposes ICTA, which consists of the following steps:

Step 1, flow simulation is performed based on the input Rank(i) ranking, which ensures the flow direction of the node influence level (from low to high).

Step 2, IFR re-ranking is performed based on the newly generated influence factor.

Step 3, repeat the above steps until the ranking Rank(i) converges.

In summary, the core idea of this paper is:

The network data is modelled in the initial stage, and the initial node influence rank is obtained according to the CILM model and its algorithm.

On this basis, in terms of network structure, the Neighbour Influence Traversal Algorithm (NITA) is proposed to optimize the accuracy of CILM according to the situation and number of neighbouring nodes of each node. In terms of network information dissemination characteristics, the Influence Cascade Transfer Algorithm (ICTA) is proposed, which iteratively calculates through the idea of "simulated flow" of influence levels. The Influence Factor Ranking (IFR) algorithm is also incorporated into the algorithm iteration.

The final ranking of each node is obtained by relying on IFR.

4. Comparison of algorithms

In the time dimension, this paper selected LFA, greedy algorithms for comparison. In terms of the propagation effect of the algorithm, this paper chooses the common heuristic algorithms based on the network structure (such as Degree, Centrality, etc.), as well as the more classical heuristic algorithms based on the characteristics of the information propagation in the network (such as LFA, etc.), so as to reflect the advantages of this paper's algorithms in various aspects.

In this paper, two datasets are chosen to compare the algorithms, the number of nodes and edges in dataset 1 is relatively small compared to dataset 2, so the simulation results in dataset 1 will be regarded as a small network simulation, and dataset 2 is a large network simulation. The following are the simulation results in dataset 1 (Figure 3):

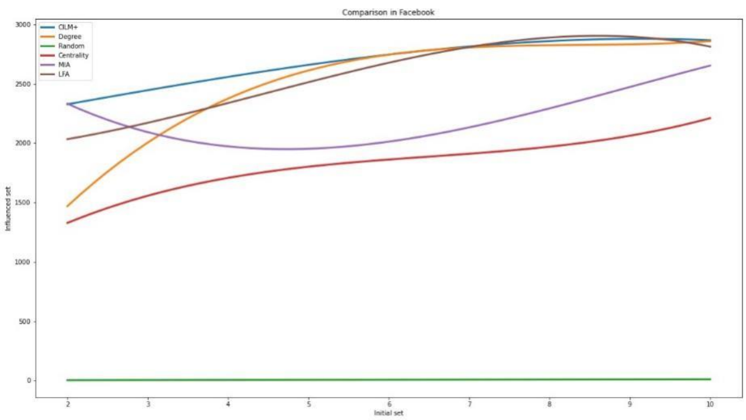

Figure 3. Facebook dataset comparison results

Here are the simulation results in dataset 2 (Figure 4):

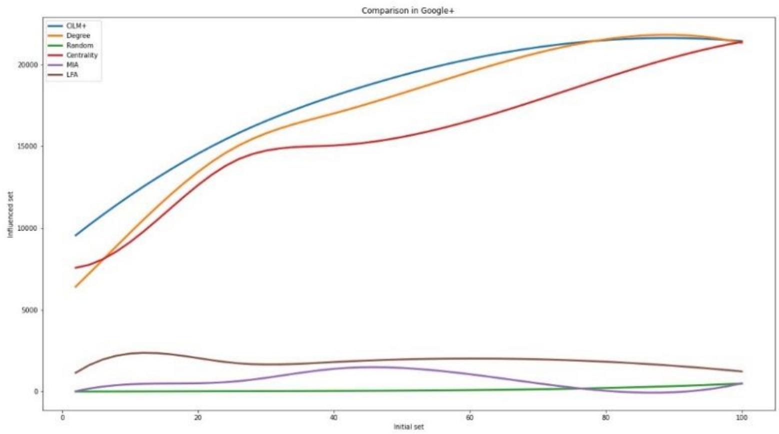

Figure 4. Google+ dataset comparison results

Comparing the above results, the following conclusions can be drawn:

The CILM+ model and algorithm proposed in this paper has better results than traditional heuristics and classical heuristics in undirected network graphs.

CILM+ is more suitable to be applied to large networks compared to the results of running in small networks.

The CILM+ algorithm well considers the basic structural characteristics of the network and the basic characteristics of information propagation in the network, so the results obtained in the simulation experiments are more realistic.

In addition, the convergence of the proposed algorithm in this paper is found to be faster in practical tests, and it can achieve convergence within 5 times for both dataset 1 and dataset 2.

The above is a comparison of the propagation effect of CILM+, in order to better prove the possibility of the algorithm to be applied to large networks in practice, this paper also does the analysis of the time aspect, taking dataset 2 as an example, the time aspect comparison is shown in the figure below:

In the actual running process, facing the dataset 2, the test result of the algorithm proposed in this paper is 199.78s, and the test result of the algorithm of LFA in the same environment is 522.68s, which is nearly three times difference. Meanwhile the test result of the greedy algorithm is too long, his time is 10^n seconds more than CILM+.

In conclusion, based on the results of the experimental part, it can be concluded that the CILM+ model and algorithm developed in this paper is accurate, realistic and usable.

5. Conclusions

In this paper, the network structure and information dissemination characteristics are considered comprehensively for the research of influence maximisation problem. Based on this, an effective model CILM with its corresponding algorithms such as NITA, CITA and IFR is developed, and the results are obtained through iterative optimisation. The overall strategy of this paper is to first determine its own influence level in the CILM model based on the number of Hi-I neighbours of each node, and then continue to consider the Hi-I and Com-I neighbours of each node through NITA, which in turn corrects the results of the CILM model. Eventually, the results of the priors are put into the CITA algorithm, which is iteratively updated according to the IFR sorting algorithm. In addition, this paper verifies the convergence, accuracy and running time of the overall algorithm. The results show that the model and algorithm established in this paper are feasible.

References

[1] Xinhua News Agency, "Outline of the Fourteenth Five-Year Plan for the National Economic and Social Development of the People's Republic of China and Vision 2035," http://www.gov.cn/xinwen/2021-03/13/content_5592681.htm, 2021 13 March 2021

[2] S. Cheng, H. W. Shen, J. Huang, et al, “IMRank: Influence Maximization via Finding Self-Consistent Ranking”, ACM, 2014.

[3] P. Domingos, M. Richardson, “Mining the network value of customers”, ACM(Knowledge Discovery & Data Mining), 2001, 57–66, 01.

[4] M. Richardson, P. Domingos, “Mining knowledge-sharing sites for viral marketing”, ACM(Knowledge Discovery & Data Mining), 2002, 61–70, 02.

[5] Kempe, et al, “Maximizing the spread of influence through a social network”, ACM(Knowledge Discovery & Data Mining), 2003, 03.

[6] J. Leskovec, A. Krause, C. Guestrin, et al, “Cost-effective outbreak detection in networks”, ACM (Knowledge Discovery & Data Mining), 2007, 420-429, 07.

[7] W. Chen, Y. Wang, S. Yang, “Efficient influence maximization in social networks”, ACM (Knowledge Discovery & Data Mining), 2009, 199–208, 09.

[8] Goyal, W. Lu, and L. V. Lakshmanan, “Celf++: optimizing the greedy algorithm for influence maximization in social networks”, The International Conference of World Wide Web, 2011, 47–48, 11

[9] S. Cheng, H. Shen, J. Huang, G. Zhang, X. Cheng, “Staticgreedy: Solving the scalability-accuracy dilemma in influence maximization”, ACM (International Conference on Information and Knowledge Management), 2013, 13.

[10] W. Chen, Y. Wang, S. Yang, “Efficient influence maximization in social networks”, ACM(Knowledge Discovery & Data Mining), 2009, 199–208, 09.