

# Novel Influence Maximization Algorithm for Social Network Behavior Management

Dr. D. Sivaganesan,

Professor,  
Department of Computer Engineering,  
PSG Institute of Technology and Applied Research,  
Coimbatore, India.  
Email: [sivaganesan@psgitech.ac.in](mailto:sivaganesan@psgitech.ac.in)

**Abstract:** The users largely contributing towards product adoption or information utilization in social networks are identified by the process of influence maximization. The exponential growth in social networks imposes several challenges in the analyses of these networks. Important has been given to modeling structural properties while the relationship between users and their social behavior has being ignored in the existing literature. With respect to the social behavior, the influence maximization task has been parallelized in this paper. In order to maximize the influence in social networks, an interest based algorithm with parallel social action has been proposed. This is algorithm enables identifying influential users in social network. The interactive behavior of the user is weighted dynamically as social actions along with the interests of the users. These two semantic metrics are used in the proposed algorithm. An optimal influential nodes set is computed by implementing the machines with CPU architecture with perfect parallelism through community structure. This helps in reducing the execution time and overcoming the real-world social network size challenges. When compared to the existing schemes, the proposed algorithm offers improved efficiency in the calculation speed on real world networks.

**Keywords:** Behavior attributes, CPU architecture, parallel algorithm, influence analysis, social networks analysis

## 1. Introduction

There is an explosive increase in the number of users as well as data exchange in social networks on an everyday basis over the past few years. Information can quickly reach a large

number of people with the rapid growth of these networks [1]. Several schemes exist for analysis of this network data, of which, social influence analysis is a significant method. The voluminous social data is utilized in an efficient manner using this technique. Political movements, epidemic outbreaks, tracking news, designing marketing, and several such seed nodes of highly influential networks may be identified for several practical applications [2]. The seed set representing a small user group is targeted by companies for designing marketing in order to advertise and recommend new products to their friend circle for maximization of the product user base. In case of Covid-19 and other epidemic outbreaks, social networks can help identifying the infectious diseases, and its influence while helping the government to adopt control measures [3].

Users are influenced by spreading awareness and information via social networks by election candidates to vote for them in terms of politics [4]. Influence maximization (IM) through social networks is a new word of mouth information diffusion ideology. Within a specific spreading model, the total users to whom the message can reach is maximized over a network with a specific user count [5]. Identification of this user set is defined as the influence maximization issue. Under conventional diffusion models, IM is NP-hard. Linear threshold (LT) and independent cascade (IC) models are used for capturing the seeds and their dynamic influence over other users. Task parallelization is challenging in social networks due to the high degree of dependency [6]. Based on the influence spreading, the network may be properly divided using community detection strategy.

## 2. Related Works

In social network analysis, it is a significant challenge to discover influential individuals who create substantial impact [7]. Several researchers are directed towards addressing the influence maximization (IM) problem over the recent years. The existing literature on the IM problem may be generally categorized into sequential and parallel algorithms. Sequential determination of the nodes ability to influence is determined using the former algorithm [8]. On observing the influence of first  $n$  nodes, the  $(n+1)$ th node selection is performed. Due to the high space and time complexities, in very large scale networks, utilization of these networks lead to significant challenges in global and centralized control. Sequential algorithms are

classified based on greedy algorithms, based on network nodes topological localization and by considering influence maximization as an optimization issue [9].

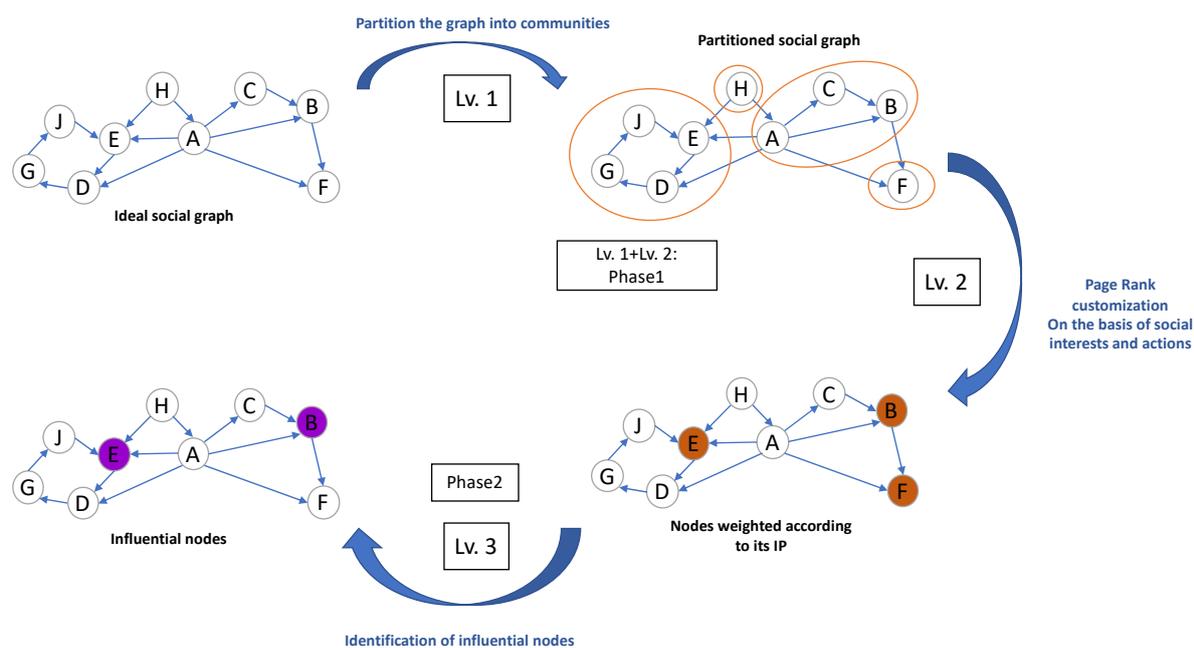
The influence elements may be calculated by converging the iteratively obtained optimum local choice at each stage into a global one in case of greedy algorithms [10]. The accurate spread of marginal influence is estimated based on Monte Carlo simulations using these algorithms. This, however, is a time consuming process. An optimal solution is provided with over 60% approximate guarantee for solving IM problem with a greedy solution [11]. Despite the effectiveness of this solution, it provides close approximation and has lesser propagation probability between links in large-scale social networks. When compared to simple greedy algorithms, there is a reduction by 700 times in the total calculation of node influence propagation using the Cost-Effective Lazy Forward (CELF) technique [4]. The unnecessary marginal gain re-calculation is avoided providing a more vivid and better evaluation by the improved CELF algorithm called CELF++.

The greedy algorithm - Practical Partitioning and Seeding (PrPaS), is focused towards maximizing the usually medium social network influences. Another scalable and fast greedy algorithm, the State Machine Greedy (SMG) acts as a single state machine while recording the influence propagation values of the final state as well as specific nodes [12]. Dominated Competitive Influence Maximization (DCIM) algorithm in combination with CELF demonstrates high speed in large scale networks. The maximum likelihood, heuristic clustering, Initial multi-spreader nodes selection (IMSN), degree distance centrality, k-shell centrality and degree discount centrality measures are used for identifying the influential nodes [13]. H-Index, clustering coefficient, companion behavior and degree centrality factors are the significant local metrics used for measuring the nodal influence.

### 3. Proposed Work

In order to maximize the influence in social networks, an interest based algorithm with parallel social action has been proposed. This is algorithm enables identifying influential users in social network. In order to address the issue of identification of influential users on social networks, the complete CPU architecture is taken advantage of, along with a combination of semantics and structure of the network [14]. The interactive behavior of the user is weighted dynamically as social actions along with the interests of the users. These two semantic metrics

are used in the proposed algorithm. An optimal influential nodes set is computed by implementing the machines with CPU architecture with perfect parallelism through community structure. This helps in reducing the execution time and overcoming the real-world social network size challenges. Influence power estimation and influential node generation are the two main phases in this model. The partitioned graph is applied with page rank at each node for weight calculation for series operation of the modules as represented in figure 1.

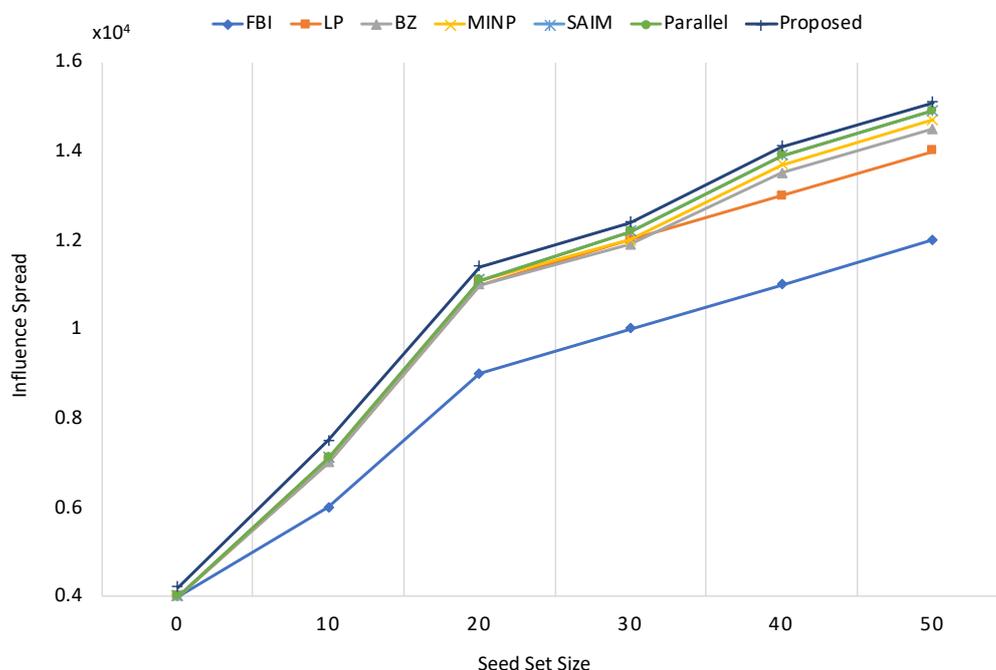


**Figure 1.** Flowchart of the proposed algorithm

For calculation of influence power, the social behavior based on the user interests and social actions is integrated with the proposed algorithm in the first module. Each users influence power is estimated with the page rank algorithm. In large graphs, when compared to other tools like degree-based centralities, the ranking resolution is higher with page rank algorithm. Large-scale network issues may be overcome by parallelization of the calculation of page rank [15]. High data dependency degree is observed in the calculation of page rank which is a significant obstacle. Estimation of other node ranks is essential for calculating the node rank. Based on the connectivity structure, the partitioning graphs principle is used by the proposed algorithm to overcome the issue of estimating node ranks.

#### 4. Results and Discussion

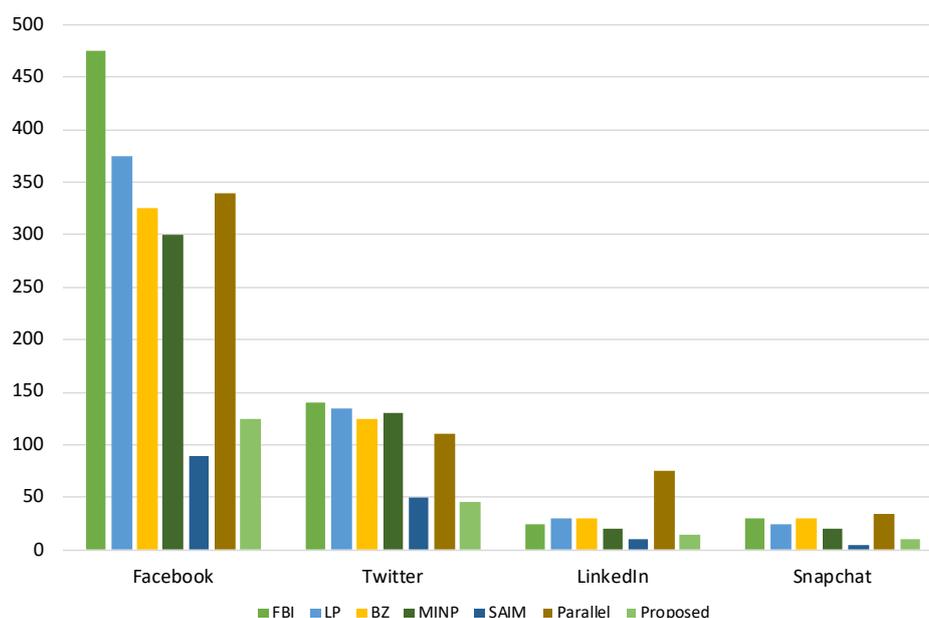
Five different algorithms are used for comparing the efficiency of the proposed algorithm. The five algorithms are chosen such that corenes centrality and k-sell type latest algorithm, page rank and degree type classic algorithms, parallel model, parallel version of serial model as well as community detection and heuristic strategy based greedy algorithm are included in the comparison. The superiority of the proposed algorithm is proven by the diversified comparisons. A 64GB memory 3 core Intel processor based 64 bit windows PC is used for executing all algorithms. Java language is used for implementing the proposed algorithm as well as the existing schemes.



**Figure 2.** Comparison of influence spread of various algorithms in specific network

Figure 2 provides the experimental results of the influence spread on specific network whose size varies between thousand to a million users. The memory size of the testing platform is limited to 64GB. The seed nodes can be determined efficiently from the candidate nodes using the influence breadth first search tree concept. This model begins with a black vertex in an empty queue. On extracting the first vertex from the queue, all the neighbors that are not visited are visited and included in the queue. A distance array stores the distance of each vertex

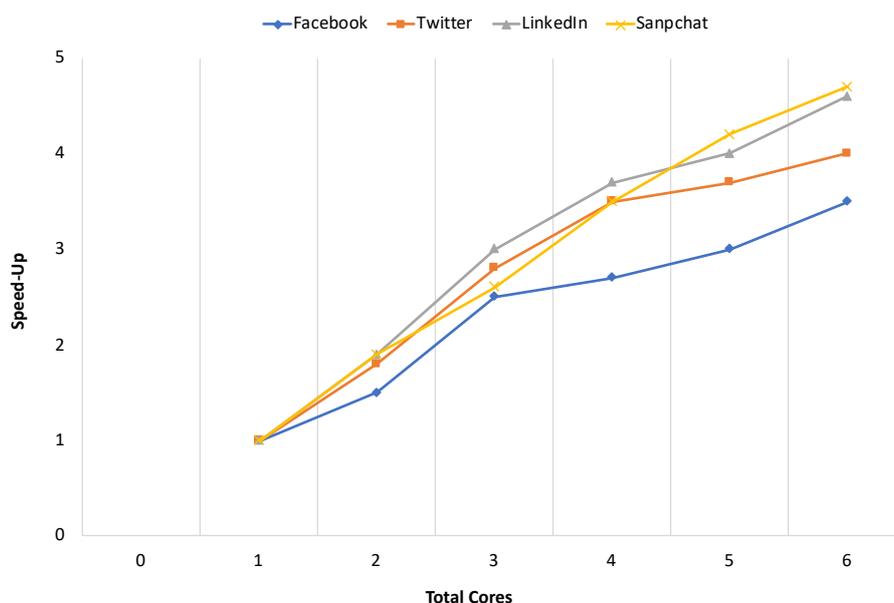
from its node. The maximum likelihood influence maximization (MLIM), parallel, linear programming (LP), coreness, mixed integer nonlinear programs (MINP), Bienstock-Zuckerberg (BZ), mixed integer programs (MIP), Social Action-Based Influence Maximization (SAIM), forensic based investigation (FBI) algorithm and so on may be compared with the proposed algorithm for information dissemination speed and efficiency. Longer black path and more similar seed candidates may be observed in the SAIM scheme on stringent observation.



**Figure 3.** Processing time for each algorithm on multiple datasets

The runtime complexity of the proposed algorithm is analyzed and compared with existing algorithms. Figure 3 presents the comparison of runtime of the proposed algorithm over multiple datasets and diverse networks. The efficiency of the proposed algorithm in terms of time complexity may be clearly observed from this image. Parallelism has been facilitated by inserting the source code with meta-programming expressions like API OpenMP for creating the parallel algorithm. Competitive computational efficiency is achieved by the proposed model using parallel computing for estimating the influence of each node independently in the community. On various social networks of different sizes, the proposed algorithm finishes first with shortest execution time without impacting the results. Page Rank algorithm has been found to be expensive which affects the computational efficiency of the

system. Parallel reduction, parallel scan, parallel sort and such efficient primitives may be transformed for improving the computational efficiency in a significant manner in the proposed algorithm. For multiple CPU cores, the speed-up growth factor of the parallel algorithm is as represented in figure 4. With the increase in the number of CPU cores, the speed-up factor increases in a sublinear manner.



**Figure 4.** Speed-up performance of the proposed algorithm

The proposed algorithm demonstrates improved performance using Higgs and Weibo network data sets. When compared to algorithms that use smaller memory, the execution time and influence spread of the proposed algorithm is higher. In comparison with the other algorithms, the time consumption is reduced significantly while using certain semantic attributes over twitter data set. For identification of influential nodes, parallelism requires minimal effort along with other significant advantages of parallel algorithm. In large scale networks, the concept of parallelism is applied to explore its operation. In terms of memory utilization and time consumption, there is a significant improvement due to the parallel CPU along with improved acceleration according to the experimental results. In order to identify the most influential node, the social semantics are exploited by the traditional parallel semantics based schemes. However, several disadvantages are faced by this as the social networks has high degree of data dependency and there is a need for increased effort in privacy protection.

## 5. Conclusion

An interest based algorithm with parallel social action has been proposed in order to maximize the influence in social networks. Potential candidate nodes are selected by excluding the less influential nodes by designing a novel parallel framework in the initial stage. The graph community decomposition and sampling scheme forms the basis of parallelism. Further, the social actions that are weighted dynamically as well as the user interests are the semantic properties of social behavior presented in the next stage. The social actions received by an individual from similar neighbors may be distinguished by the proposed model. The seed nodes can be determined efficiently from the candidate nodes using the influence breadth first search tree concept. Quick information spread may be ensured using this technique. The proposed algorithm is extremely fast, requires less memory, is time efficient and overcomes the tradeoffs seen in the existing algorithms. Future work is focused on implementing pruning algorithm in combination with parallelism in large networks. The dynamic evolution of networks in terms of time, social relationships, social actions and community change are also proposed to be addressed with adaptive algorithms.

## References

- [1] Wilder, B., Yadav, A., Immorlica, N., Rice, E., & Tambe, M. (2017, May). Uncharted but not Uninfluenced: Influence Maximization with an Uncertain Network. In AAMAS (Vol. 17, pp. 1305-1313).
- [2] Aral, S., & Dhillon, P. S. (2018). Social influence maximization under empirical influence models. *Nature human behaviour*, 2(6), 375-382.
- [3] Banerjee, S., Jenamani, M., & Pratihari, D. K. (2020). A survey on influence maximization in a social network. *Knowledge and Information Systems*, 62(9), 3417-3455.
- [4] ŞİMŞEK, A., & Resul, K. A. R. A. (2018). Using swarm intelligence algorithms to detect influential individuals for influence maximization in social networks. *Expert Systems with Applications*, 114, 224-236.
- [5] Wilder, B., Immorlica, N., Rice, E., & Tambe, M. (2018, April). Maximizing influence in an unknown social network. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32, No. 1).

- [6] Cui, L., Hu, H., Yu, S., Yan, Q., Ming, Z., Wen, Z., & Lu, N. (2018). DDSE: A novel evolutionary algorithm based on degree-descending search strategy for influence maximization in social networks. *Journal of Network and Computer Applications*, 103, 119-130.
- [7] Khomami, M. M. D., Rezvanian, A., Bagherpour, N., & Meybodi, M. R. (2018). Minimum positive influence dominating set and its application in influence maximization: a learning automata approach. *Applied Intelligence*, 48(3), 570-593.
- [8] Tang, J., Zhang, R., Wang, P., Zhao, Z., Fan, L., & Liu, X. (2020). A discrete shuffled frog-leaping algorithm to identify influential nodes for influence maximization in social networks. *Knowledge-Based Systems*, 187, 104833.
- [9] Han, M., Han, Q., Li, L., Li, J., & Li, Y. (2018). Maximising influence in sensed heterogeneous social network with privacy preservation. *International Journal of Sensor Networks*, 28(2), 69-79.
- [10] Caliò, A., Interdonato, R., Pulice, C., & Tagarelli, A. (2018). Topology-driven diversity for targeted influence maximization with application to user engagement in social networks. *IEEE Transactions on Knowledge and Data Engineering*, 30(12), 2421-2434.
- [11] Durairandian, M. (2020). Long Term Evolution-Self Organizing Network for Minimization of Sudden Call Termination in Mobile Radio Access Networks. *Journal of trends in Computer Science and Smart technology (TCSST)*, 2(02), 89-97.
- [12] Bashar, A. (2020). Sensor Cloud Based Architecture with Efficient Data Computation and Security Implantation for Internet of Things Application. *Journal of ISMAC*, 2(02), 96-105.
- [13] Sathesh, A. (2019). Enhanced soft computing approaches for intrusion detection schemes in social media networks. *Journal of Soft Computing Paradigm (JSCP)*, 1(02), 69-79.
- [14] Haoxiang, W. (2020). Emotional Analysis of Bogus Statistics in Social Media. *Journal of Ubiquitous Computing and Communication Technologies (UCCT)*, 2(03), 178-186.
- [15] Manoharan, S. (2020). Geospatial and social media analytics for emotion analysis of theme park visitors using text mining and gis. *Journal of Information Technology*, 2(02), 100-107.