ISSN No:-2456-2165

Unpacking the Black Box: An Exploration of the Algorithms Driving Social Media Networks-A Review

Amey Thorat
Department of Computer Engineering
Vidyalankar Institute of Technology

Amulya Kura
Department of Computer Engineering
Vidyalankar Institute of Technology

Ansika Jaiswal
Department of Computer Engineering
Vidyalankar Institute of Technology

Sapana Survase
Department of Computer Engineering
Vidyalankar Institute of Technology

Amit Aylani
Prof., Department of Computer Engineering
Vidyalankar Institute of Technology

Abstract:- Social media algorithms are used for finding detailed information in large unstructured data by relevant keywords used by users. There are different algorithms used for social media from a searching point of view. One of the algorithms is the "Probability of Node's Degree" algorithm, which is based on the concept of breadth-first search, random walk, and the highest degree seeking algorithm. The algorithm involves selecting a source node and a target node, and then traversing the nodes in the network to find the target node. The algorithm checks if the target node is a neighbor of the current node and, if not, transmits a query message to other nodes based on their probability of being relevant to the search. Nodes with higher degrees are more likely to be searched, making the algorithm beneficial to nodes with higher degrees. In addition to this, there are other algorithms such as FP-FOREST, DSTree, UPTree algorithm, and KC-LA, which are used for finding frequent patterns, maintaining and mining frequent item sets, and finding K-Clique in complex social networks. These algorithms are useful in data-driven decision-making and in gaining insights into social media analytics.

Keywords:- *Social Media Algorithm, Social Media Analytics, Complex Social Network, Social Media, K-Clique, Learning Automation, Betweenness Centrality, Random Walk.*

I. INTRODUCTION

Social Media has become an integral part of our daily lives, allowing people to connect, share, and exchange information and ideas in virtual communities and networks.

Social media may help us connect with friends and family, learn about and pursue new hobbies, and have fun from a personal viewpoint. By engaging with other professionals in your field, you may utilize social media to advance your professional network and increase your expertise in a certain subject. Social media gives your business the opportunity to interact with customers, get their opinion, and build brand recognition.

Online social networks (OSNs) have recently gained considerable attention. For example, Twitter had over 300 million monthly active users in 2018. There are many studies that have analyzed a particular OSN as a graph with nodes of users and edges of relationships among users.

In this paper, we will discuss the findings of various researchers on different algorithms related to complex networks. The first paper by GU Yiran and ZHAO Wenwen introduces the concept of the Probability of Node's Degree algorithm that is based on breadth-first search, random walk, and the highest-degree seeking algorithm. The second paper discusses the development of a compact data structure named FP-FOREST that enhances the performance of an existing algorithm called INSTANT for frequent pattern mining in social media streams. The third paper by Mohammad Mehdi Daliri Khomami et al. proposes a distributed learning automata-based algorithm called KC-LA for finding K-Clique in complex social networks. This paper covers various algorithms proposed for solving maximal clique finding, which is an NP-hard problem, with practical applications in community detection in social networks. The fourth paper discusses the estimation of top nodes which has highest betweenness centrality which has a shortest path pass through the vertices. In the following sections, we will discuss each paper and highlight their contributions to the field of complex networks.

II. LITERATURE SURVEY

A. *Probability Of Node's Degree On Complex Networks*

In this Paper, author GU Yiran, ZHAO Wenwen proposed the social media's search algorithm "Probability of Node's Degree" which is based on the concept of breadth first search, random walk, highest degree seeking algorithm.

Before going on an improved algorithm of probability of node's degree first we will see the based algorithm of concept. Breadth First Search (BFS) begins as a level order traversal. Basically, in this algorithm, we can find the path between two people in level order. Random Walk algorithm is used for extracting information from the ensemble of paths between persons in a graph. Process of searching for the highest-degree neighbor in DS in each node is aware of the information about its neighbors. However, it also somewhat lessens the search process's query message flow, which can help prevent network congestion.

Now, based on the above concept, the probability of a node's degree algorithm is explained, which is given above: Each node knows the degree of its use neighbors, as shown by the image of the search by degree sequence, and searches for the neighbor with the highest degree at each step. The neighbor with the second-highest degree will be picked to broadcast a query message if the neighbor with the highest degree has already been visited. The search technique can thus get the best possible results when used on complex networks.

- *Algorithm of probability of node's degree search (PDS) works as follows:*

This algorithm involves selecting a source node and a target node, and then traversing the nodes in the network to find the target node. The algorithm checks if the target node is a neighbor of the current node and if not, transmits a query message to other nodes based on their probability of being relevant to the search. Nodes with higher degrees are more likely to be searched, making the algorithm beneficial to nodes with higher degrees. Nodes can be visited more than once, edges can be visited only once. The search continues until the target node is found or all the nodes have been visited.

B. *Finding Frequent Patterns In Social Media Streams*

A cutting-edge algorithm called INSTANT is being improved with the help of a small data structure called FP-FOREST, which demonstrates how to compress itemsets and count supports efficiently. The algorithm performs better in terms of memory use and execution time, according to the results. By utilising its appealing qualities, a novel tree structure known as DSTree (Data Stream Tree) gathers significant data from streams and can be readily maintained and mined for frequent itemsets as well as numerous other patterns such restricted

- *UPTree algorithm -*

Hash tables, disjoint sets, and graphs can all be implemented using the data structures and algorithms covered in the Unordered Data Structures course. For unordered data, these fundamental data structures are helpful. A hash table, for instance, offers instant access to data that is indexed by any key value, which could be a dictionary, a number (such as a memory address for cached memory), or a URL. Graphs are used to represent relationships between objects, and this course covers a variety of graph representational data structures and graph traversal algorithms, such as determining the shortest path between two nodes. This course will also address disjoint sets, another idea on which these graph algorithms will depend.data structure and associated algorithms.

C. *Finding K-Clique In Complex Social Networks*

The selected paper titled "Distributed Learning Automata-based Algorithm for Finding K-Clique in Complex Social Network" authored by Mohammad Mehdi Daliri Khomami et al. Proposes a new algorithm that utilizes distributed LAs to solve k-Clique called(KC-LA)in social networks.

The authors discuss the problem of maximal clique finding in graphs, which involves finding a subset of nodes in a graph where each node is connected pairwise. The maximum clique refers to the clique with the largest number of nodes. This problem is NP-hard, and various algorithms have been proposed to solve it, including k-clique algorithms, which involve finding cliques with a fixed size k. Many of these algorithms fall into one of three categories: deterministic, heuristic, and approximation. Maximal clique finding has practical applications in fields such as community detection in social networks. Some researchers have attempted to utilize the concept of cliques for community detection, and various algorithms have been developed for this purpose. Additionally, the authors mention LA-based algorithms that have been successful in solving graph problems such as positive influence dominating set, independent set, vertex cover, and community detection.

Now based on the above concept, a learning automaton is a type of learning model that learns to choose the best action from a set of actions in a random environment by receiving feedback in the form of rewards or penalties. It updates its probability vector for selecting actions based on whether the feedback is favorable or unfavorable. A DLA is a network of LA that cooperates to solve a particular problem. In this work, just a LA is deactivated at a ume. The number of actions performed by automata is equal to the number of LAS Connected to it [3].

- *KC-LA based for k-Clique -*

The KC-LA algorithm finds a k-clique in a social graph by using a network of learning automata. Each automaton is assigned to a vertex and chooses actions corresponding to the edges connected to the vertex. The algorithm selects an automaton, activates it, and adds it to the k-clique set. The automaton selects an action, and the algorithm checks whether it can be added to the k-clique set. The process continues until all automata are disabled. The algorithm evaluates the cardinality of the k-clique set and rewards or penalizes the selected action based on whether it increases or decreases the size of the k-clique set. The algorithm also

modifies the probability of selecting actions based on the degree of nodes to improve the selection process.

D. Estimation of Top-K Betweenness Centrality Nodes

In the paper "Estimating Top-k Betweenness Centrality Nodes in Online Social Networks", author proposed a new crawling-based algorithm to estimate the top-k nodes in online social networks which has the highest betweenness centrality. Privacy and security concerns prevent obtaining complete information in Online social networks hence Crawling based algorithm is used which considers limitation of request. This algorithm works in two stages: the first is a sampling method in which edges or nodes are sampled from a graph. They describe that they should choose a sampling method which has an overall high collection ratio. In the second stage the top-k nodes are estimated from a gram which has highest betweenness centrality. In this paper the author proposed a new crawling based algorithm which first estimates the ego betweenness centrality by random walk and then estimates the top-k betweenness centrality.

The betweenness centrality is the sum of the ratio of shortest paths between two nodes which pass through a node in the graph. The users or nodes which have high betweenness centrality in online social networks are the main key for the spread of information. Since these nodes have high betweenness centrality, hence they are present on many of the shortest paths in the graph.

III. COMPARISON

TOPIC	Probability of Node's Degree	Finding Frequent Patterns	Finding K- Clique	Estimation of Top-k between ness centrality nodes
WHY IS IT NEED ED?	To Send Message using highest degree node	Mining associations , correlations, relationships among data	Community detection, solving graph problems	Finding the nodes which plays bridge role in network
LIMITATIONS	Only works for undirected graph	Mining frequent patterns in large dataset	Identifying largest subgraph of size k	Ignores path and hops for longer than shorter path

Table 1 Comparison

IV. CONCLUSION

The improvements show an effect on search cost and the efficiency of the search. To extract frequent patterns, FP-tree algorithms combined with LRU structure, UPTree is exploited. It modifies the probability of selecting actions based on degree of nodes to improve the selection process. Top-k betweenness centrality nodes estimated to flow the information all over the network.

V. FUTURE SCOPE

In this paper, we discussed the algorithms which are mainly based on undirected graphs for finding required information which can be found by one side only. So we can upgrade the algorithm to use it in directed graphs and can suggest searching from both ends.

REFERENCES

[1]. GU Yiran, ZHAO Wenwen,"Improved Search Algorithm Based on Probability of Node's degree on Complex Networks" in 2012 Third International Conference on Digital Manufacturing & Automation

[2]. Suwook Ha, Yong Mi Lee, Kwang Woo Nam, and Keun Ho Ryu,"An Algorithm for Finding Frequent Patterns in Social Media Stream", in International Conference on ICT Convergence (ICTC) on 14-16 October 2013

[3]. Mohamad Mehdi Daliri Khomami,Alireza Rezvanian, Ali Mohannad Saghiri, Mohammad Reza Meybodi, "Distributed Learning Automata-based Algorithm for Finding K- Clique in Complex Social Network"2020 11th International Conference on Information and Knowledge Technology(IKT)

[4]. E.A. Akkoyunlu,"The enumeration of maximal cliques of large graphs, "SIAM Journal on Computing ,vol. 2,no. 1,pp.1-6,1973

[5]. M.Regneri,"Finding all cliques of an undirected graph,"2007

[6]. Peter Mika. "Ontologies are us: A unified model of social networks and semantics", In Proc.14th International Semantic Web Conference, 2005

[7]. Matsuo Y, Hasida K, Tomobe H, Ishizuka M. "Mining social network of conference participants from the web", IEEE/WIC International Conference, 2003

[8]. Kazuki Nakajima, Kenta Iwasaki, Toshiki Matsumura, Kazuyuki Shudo, "Estimating Top-k Betweenness Centrality Nodes in Online Social Networks", International Conference on. IEEE,2018.

[9]. L. C. Freeman, "A set of measures of centrality based on betweenness," Sociometry, vol. 40, no. 1, pp. 35–41, 1977

[10]. T. Matsumura, K. Iwasaki, and K. Shudo, "Average path length estimation of social networks by random walk," in Big Data and Smart Computing (BigComp), 2018 IEEE International Conference on. IEEE, 2018, pp. 611–614

[11]. B. Goethals, M. J. Zaki, Workshop Report on Workshop on Frequent Itemset Mining Implementations (FIMI), ACM SIGKDD Explorations Newsletter - Special issue on learning from imbalanced datasets, vol. 6(1), June 2004.