IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Volume-5, Issue-12 E-ISSN: 2347-2693

A Survey on: Comparative Study of Minimum Spanning Tree Algorithm

R. Gomathijayam[1*], V. Santhi[2]

[1*]Department of Computer Applications, Bon Secours College for Women, Thanjavur, India
[2]Department of Computer Applications, Bon Secours College for Women, Thanjavur, India

[*]Corresponding Author: rgj.tnj@gmail.com, Tel.: 9944939110

Available online at: www.ijcseonline.org

Received: 25/Nov/2017, Revised: 02/Dec/2017, Accepted: 18/Dec/2017, Published: 31/Dec/2017

Abstract---- M*inimum spanning tree can be obtained for undirected connected weighted edges with no negative weight using conventional algorithms such as genetic, Prim's and Filter-Kruskal. This paper presents a survey on the conventional and the more recent algorithms with different techniques. This survey paper also contains comparisons of MST algorithm and their advantages and limitations.*

Keywords--- MST, Graph, GA – Genetic Algorithm, DCC_Trees, DWCM - Difference Weighted Circuit Matrix.

I. INTRODUCTION

Given a connected, weighted, and undirected graph, the minimum routing cost spanning tree problem seeks on this graph a spanning tree of minimum routing cost, where routing cost is defined as the sum of the costs of all the paths connecting two distinct vertices in a spanning tree.

Let G= (V, E) an undirected connected graph. A sub graph t = (V E') of G is a spanning tree of G iff t is a tree.

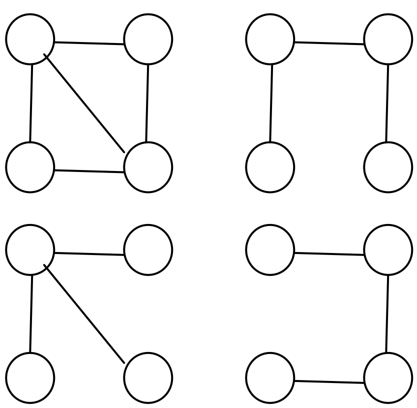

Figure1 Undirected graph and three of its spanning tree

There are two types of graph, Directed graph and undirected graph. In an undirected graph the pair of vertices representing any edge is unordered. Thus, the pairs (u, v) and (v, u) represent the same edge. In a directed graph each edge is represented by a directed pair (u, v); u is the tail and v the head of the edge. Therefore, the pairs (u, v) and (v, u) represent the different edges. The maximum number of edges in an undirected graph without a self-loop is n (n - 1)/2. In the case of a directed graph on n vertices, the maximum number of edges is n (n-1) [1].

A. Applications of MST [10]

Spanning tree has many applications. It can be used to obtain an independent set of circuit equations for an electric network, representing the connection between the cities, design of computer & communication network and telephone network, etc. It offers a method of solution to other problems to which it applies less directly, such as network reliability, clustering and classification problems and used to find the approximation solution for the NP hard problems[2].

i. Network design:

Minimum spanning trees are useful in applications like telephone, electrical, hydraulic, TV cable, computer, road network design. One standard application can be a problem like phone network design. If a number of offices are to be connected using leased lines and the phone company charges different amounts of money to connect different pairs of offices depending on the difficulties involved in laying phone lines. Then the telephone network that has to be installed can be a minimum spanning tree.

ii. Cluster Analysis:

K clustering problem can be viewed as finding a Minimum spanning tree and deleting the k-1 most expensive edges.

iii. Brain Network Analysis:

Analysis of the minimum spanning tree is useful in the comparison of brain networks. The advantage is that, it avoids methodological biases. Even though the minimum spanning tree does not utilize all the connections in the network, it still provides a, mathematically defined and unbiased, sub-network with characteristics that can provide similar information about network topology as conventional graph measures.

iv. Data Storage of Amino acids in Protein Structure:

A clustering method based on Minimum Spanning Tree-based was proposed by Karthikeyan *et al*. Firstly, the N x N distance matrix is constructed, where; N is the number of proteins in the dataset. Then the complete graph is constructed from the distance matrix. In the complete graph each node is associated with a single protein to be clustered. The distance of the protein from all the remaining proteins in the dataset is calculated using Euclidian distance method. Minimum Spanning Tree is constructed from the complete graph using Prim's algorithm. The weight of each edge in the Minimum Spanning Tree is the distance between the connected protein nodes. This minimum spanning tree is used for clustering the proteins.

v. Broadcasting in Computer Networks

In Ethernet network, information is broadcasted to all the nodes in the network using Spanning tree protocol. In large computer networks, it is useful in constructing trees for broadcasting information to all the nodes in the network.

vi. Other Applications

There are many other applications were information on the minimum spanning tree or the set of all possible spanning trees are useful of the input network data set is useful. Image registration and segmentation, Curvilinear feature extraction in computer vision, Handwriting recognition of mathematical expressions, Circuit design: Regionalization of socio-geographic areas, the grouping of areas into homogeneous, contiguous region, Comparing ecotoxicology data, max bottleneck paths , LDPC codes for error correction, learning salient features for real-time face verification, reducing data storage in sequencing amino acids in a protein, model locality of particle interactions in turbulent fluid flows , auto-config protocol for Ethernet bridging are some examples.

B. Objective of MST

a. To calculate the minimum cost of the spanning tree for both directed and undirected.
b. To reduce traffic load on the network.
c. To remove the cycle from the graph from the MST.
d. To enhance the time complexity of the MST.

The rest of this paper is organized as follows. Section 2 provides a summary of conventional MST algorithms. Section 3 describes a literature survey of recently devised efficient algorithm. Finally, we conclude the matters which are acquired through this study.

II. MST CONVENTIONAL ALGORITHM

There are various conventional algorithms available which are discussed below:

Genetic, Prim's and Filter-Kruskal algorithm are algorithms that used to find a minimum spanning tree for a connected weighted undirected graph. This means when the total weight of all the edges is minimized in the tree, at that time it finds a subset of the edges which forms a tree which includes every vertex.

A. Genetic algorithm [4]:

The GA is algorithm in performance of probability search in optimization and capable of dealing with any kind of objective and constraint functions. It starts from a set of candidates solutions of the problem, improves them step by step through biological evolutionary process and gets to the optimal solutions.

The self-adaptive GA approach for the MST problem is researched. In order to encode a corresponding tree structure solution for the problem, a tree-based permutation is developed that is able to represent all possible rooted tree such tree coding has the property that it inherits good properties from parent, and unlike traditional incremental approach to MST format, such approach directly performs a search on coding space and evolves them to the Pareto optimal solutions through the genetic operations.

Advantages:

Combined with the GA techniques and self–adaptive technique, the result of these methods shows its high effectiveness in dealing with the MST problem.

B. Filter-Kruskal Minimum Spanning Tree Algorithm [3]:

It is a modified and extended version of Kruskal's algorithm which avoids sorting all of the input edges. The algorithm runs Kruskal's algorithm on subsets of candidate edges repeatedly. To find these subsets, it chooses a pivot edge which separates unused edges into two sets. The method is called on these sets recursively. Lighter sets are more probable edge sets. If all the required edges are not found in these sets, then it explores the heavier edge sets. Another important feature of this algorithm is that it chooses only those edges whose end points are in two different components.

The algorithm selects $[V]$ number of edges each time it invokes Kruskal's algorithm, where c is a constant. It chooses the median of a random sample of size (k) where k is the size of an input segment as the pivot value. The running time of the algorithm is ($m + n \log n \log$) for not too sparse graphs, where $m = |E|$ and $n = |V|$.

Advantages:

Filter-Kruskal performs better than all other algorithms for random graphs with random edge weights. When the edge density is high, it is outperformed only by pJP. For difficult instances, JP and Filter-Kruskal perform in the same way, but Kruskal and qKruskal do better. Filter-Kruskal outperforms all the others for random geometric graphs with 216 nodes.

pJP and qJP run faster than Filter-Kruskal by a small margin for lollipop graphs. For real instances, this algorithm performs the best.

C. Prim's algorithm:

A greedy method to obtain a minimum-cost spanning tree builds this tree edge by edge. The next edge to include is chosen according to some optimization criterion. So, the edges are chosen that results in a minimum increase in the sum of the costs of the edges so far included [1]. Using a simple binary heap data structure complexity is O(|E| log |V|) where |E| is the number of edges and |V| is the number of vertices. Using Fibonacci heap in dense graph complexity is O(|E| + |V| log |V|), which is asymptotically faster prim's algorithm steps are given below[2]:

1) Create a set MST Set that keeps track of vertices already included in MST.

2) Assign a key value to all vertices in the input graph. Initialize all key values as INFINITE. Assign key value as 0 for the first vertex so that it is picked first.

3) While MST set doesn't include all vertices:

a) Pick a vertex u which is not there in MST Set and has minimum key value.

b) Include u to MST Set.

c) Update key value of all adjacent vertices of u. To update the key values, iterate through all adjacent vertices. For every adjacent vertex v, if weight of edge u-v is less than the previous key value of v, update the key value as weight of u-v.

Advantages are:

- i) Easy to understand.
- ii) Root node is selected so clear about the starting node.

Limitations are:

- i) Time taken to check for smallest weight arc makes it slow for large numbers of nodes.
- ii) Difficult to program, though it can be programmed in matrix form.
- iii) Same weight may increase the complexity when one of the weights is eliminated in a cycle.

III. LITERATURE SURVEY

In this section, lot of research works has been recorded from past few years. They are presented here:

A. Clustering Algorithm [5]:

The clustering algorithm is to discover clusters of unusual shapes and densities. Hierarchical and Density based ways are implemented for constructing minimum Spanning Tree; the MST can be divided into two segments. In the first segment, local density is guesstimate at every data point. In the subsequent segment, hierarchical ways are used by combining clusters according to the calculated cluster distance based to go-beyond in distribution of data objects. The recommended method improvises the efficiency of clustering result, where the data is distributed in different shapes and density; it leads to better clustering efficiency. This approach presents a clustering algorithm that is inspired by MST. In this algorithm, a new method for construction which reduces the computational complexity compared with traditional MST construction methods.

B. Cluster-on-Demand MST Clustering Algorithm [6]:

In this algorithm, the new idea has been introduced of forming clusters on demand as a means of checking the maintenance and stability of clusters which has been a major issue as far as V2V communications in VANETs is concerned.

The position of vehicles is obtained using GPS. A vehicle in a specific road segment is selected at random and using Prim minimum Spanning Tree algorithm, a list of vehicles are selected to form a cluster. The selected vehicle becomes the cluster-head for that specific cluster. In selecting the cluster members the criteria is to select vehicles with similar velocities, near the cluster-head and moving in the same direction as the cluster-head. Also the number of vehicles is regulated so that the maximum area and the number of vehicles per cluster are not exceeded. A stopping criterion is the area covered by the cluster since as increase number of vehicles under one cluster, the area increases until it reaches a threshold where QoS degradation is unacceptable. Prim algorithm is used to find a minimum spanning tree for a connected weighted graph. The algorithm finds a minimum cost spanning tree by selecting edges from the graph one-by-one as follows:

1) It starts with a tree, T consisting of a single starting vertex, x.

2) Then, it finds the shortest edge emanating from x that connects T to the rest of the graph (i.e., a vertex not in the tree, T).

3) It adds this edge and the new vertex to the tree T.

4) It then picks the shortest edge emanating from the revised tree T that also connects until a minimum spanning tree satisfying a specified condition is achieved. For example, consider a graph G formed by vehicles in motion in a road segment where vertices v is the total number of vehicles in this road segment and moving in the same direction. Edges

$$e = v-1 \quad (1)$$

is the number of edges in the graph G without forming a loop, and edge weights w is the metric resulting from combining similarity of velocity and minimum distance. Let the accepted QoS be represented by k. k is calculated from the distance D of any nearest vehicle from the cluster-head and the maximum range of the cluster-head R. They want to generate a graph T to form a cluster that satisfies their desired QoS. Starting with an empty graph T, randomly select a vehicle in a road segment. Let this vehicle be $v1$. Starting from $v1$, let vertex $v1$ and edge $e1$ have the smallest metric weight (w)in graph *G*, remove ($e1$,$v1$,$w1$) from *G* and add $e1$ and $v1$ to *T*. For the remaining edges (vi, ei), check if vi is not already in *G* and if so, find the minimum weight wi in graph *G* then add vi and ei to *T*. They then repeat this until QoS condition of k is achieved and then check if the number of vehicles so far selected exceeds the allowed value n. If so, it reduces the number of vehicles to satisfy this condition starting with the last vehicle to be added. The cluster is then formed with $v1$ as the cluster-head. Details of the algorithm are as depicted in Algorithm 1.

C. DCC_Trees Algorithm [7]

This algorithm proposes a unique solution to the problem of all possible spanning tree enumeration for a simple, symmetric, and connected graph. It is based on the algorithmic paradigm named divide-and-conquer. The algorithms developed for spanning tree enumeration are mainly targeted towards generating all trees in optimum time and space. Moreover, checking for the duplicate tree and non-tree sequences are two major issues in this area of research. A typical divide-and-conquer algorithm solves a problem using following three steps:

i) Divide: Break the given problem into sub problems of the same type
ii) Conquer: Recursively solve these sub problems.
iii) Combine: Appropriately combine the results of the sub problems.

The algorithm formulated is capable of computing all possible spanning trees of a simple, symmetric, and connected graph. The given graph has been divided into a number of partitions, which can be joined by a set of connectors. The selection of different connectors and the way they are being combined with the different partitions give rise to different spanning trees of the graph. It does not generate any duplicate tree and also minimizes the formation of the circuit in its tree generation procedure, which has also been taken care of eventually. Due to the limitation in a computing environment, some of the algorithms considered are taking time in weeks for execution.

Salient Features of DCC_Trees

Some of the key features of the algorithm as follows.

- The algorithm is based on a unique approach never attempted before.
- It guarantees no duplicate tree generation.
- It can generate the spanning trees with a particular edge(s) included, i.e., if find out those spanning trees of a given graph with a particular edge or a group of edges from a particular vertex being always included, then the algorithm can do so efficiently.
- The consistency of the proposed algorithm remains valid irrespective of the selection of starting vertex.
- It is also suitable for parallel processing. Once the primary partitions are ready, trees generated from any one partition are independent of those from another partition, and hence, can be processed in parallel.

D. Combinatorial algorithm [8]:

This algorithm deals with generation of all possible spanning trees in increasing cost of a weighted graph and uses one matrix called Difference Weighted Circuit Matrix (DWCM). To find a minimum spanning tree for a connected weighted graph with no negative weight can be obtained using classical algorithms such as Prim's and Kruskal. This algorithm performs two major tasks. 1) Cord: The edges that are not in the spanning tree of a graph are called the chord. That is the sub graph S is the collection of Chord of the graph G with respect to S the Spanning tree of the graph. 2) DWCM: The abbreviation is Difference Weighted Circuit Matrix. It is the little bit of modification of the FCM. A sub matrix in which all rows correspond to a set of fundamental circuits is called a Fundamental circuit matrix. If n is the number of vertices and e is the number of edges in a connected graph, then the matrix is an (e-n-1) *(n-1) matrix. Here the branch weights are present on the column head as branch mark. The chords (e-n-1) are for the row representation. Here each cell of the matrix is assigned difference weight of the chord and the branches participating for generating circuit. When the column head presented, this chord is joined to the spanning tree.

Advantages:

The minimal spanning tree represents the minimal path between the nodes of the graph. It may possible some times in real life that minimal path can't be reached due to some circumstances, in that case the next minimal spanning tree is useful.

Limitation:

Complication is much because of generating more than one spanning tree more.

E. Euclidean based MST algorithm [9]:

This algorithm is based in part on the well-separated pair decomposition, introduced by Callahan and Kosaraju. To compute the Euclidean Minimum Spanning Tree of n points in the space, one can naively link each pair of edges to build the complete Euclidean graph G=(P,E), where P stands for the vertex set, and E stands for set of edges, which consist of all pairs (p, q) ∀ p, q ϵ P. For such a graph, it has |E|=� nn 2�, which is Θ(nn2), and combining this with the O(ElogV)=O(nn2logn) running time for the execution of Kruskal's Algorithm, would result in an overall running time of O(nn2logn).

This algorithm for computing the Euclidean minimum spanning tree will be based on a number of components. More specifically, the program provides the construction of a point region quadtree, which used to store input set of data points. The algorithm then implemented the well-separated decomposition upon the constructed point region quadtree to store each well-separated pair of cells as cross links in the tree. The next procedure is to implement projections for each pair of well-separated cells in order to find (approximately) the closest pair of points and build the graph G. Finally, it implements Kruskal's Algorithm to obtain the Minimum Spanning Tree from G.

Advantages:

i) Well-separated decomposition theorem is widely applied in several fields such as astronomy, molecular dynamics, fluid dynamics, plasma physics, and even surface reconstruction.

ii) It greatly improves the storage space and running time efficiency over traditional approaches.

IV. CONCLUSION

This survey paper focuses conventional algorithms and advanced MST algorithm. In each study, the concept of existing algorithm, advantages and their limitations are discussed. The observation is that complexity is very high because of cycle in the graph and the edges with the same weight. In future, we hope that this paper provides guidelines and idea to generate new proposed algorithm to solve the minimum spanning tree problem.

REFERENCES

[1] Ellis Horowitz, Sartaj Sahni, Sanguthevar Rajasekarn, *"Fundamentals of Computer Algorithms (2nd Ed.)"*, University Press (India) Private Ltd. – 500 029, pp. 236.

[2] Nimesh Patel, Dr. K. M. Patel, *"A Survey on: Enhancement of Minimum Spanning Tree*", Nimesh Patel Int. Journal of Engineering Research and Applications, India, pp. 1, 2015.

[3] Abdullah-Al Mamun, Sanguthevar Rajasekaran, *"An Efficient Minimum Spanning Tree Algorithm"*, IEEE Symposium on Computers and Communication, (ISCC),pp. 1-2, 2016.

[4] Hong Liu, GenguiZhou, "*Minimum Spanning Tree Problem Research based on Genetic Algorithm"*, 2nd International Symposium on Computational Intelligence and Design, China, pp. 1-5, 2009.

[5] P.Praveen, B. Rama, T.Sampath Kumar, *"An Efficient clustering algorithm of Minimum Spanning Tree"*, 3rd International Conferences on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics, pp. 1-5, 2017.

[6] Jerry John Kponyo , Yujun Kuang , Enzhan Zhang , Kamenyi Domenic, *"VANET Cluster-on-Demand Minimum Spanning Tree (MST) Prim Clustering Algorithm",* in the proceedings of ICCP2013, China, pp. 2, 2013.

[7] Maumita Chakraborty, Ranjan Mehera and Rajat Kumar Pal, *"A Divide-and- Conquer Algorithm for All Spanning Tree Generation"*, Springer Nature Singapore Pte Ltd., R. Chaki et al. (eds.), Advanced Computing and Systems for Security, Advances in Intelligent Systems and Computing 567, pp. 6,14, 2017.

[8] Barun Biswas, Krishnendu Basuli, Saptarshi Naskar, Saomya Chakraborti and Samar Sen Sarma, *"A combinatorial algorithm to generate all spanning trees of a weighted graph in order of increasing cost*", CoRR, India, pp.2,3, 2012.

[9] Chaojun Li, *"Euclidean Minimum Spanning Trees Based on Well Separated Pair Decompositions"*, Dave Mount, pp.2-4, 2014.

[10] K. Lakshmi1, T. Meyyappan, *"Spanning Tree- Properties, Algorithms and Applications"*, International Journal of Computer Sciences and Engineering, pp-4, Vol.5(10), 2017.

Authors Profile

Ms. R. Gomathijayam pursed Bachelor of Science from Annamalai University, Chitambaram in 1998, Master of Science from Bharathidasan University in year 2001 and Master of Philosophy in 2006. She is currently working as Assistant Professor in Department of Computer Applications, Bon Secours College for Women, Thanjavur, Tamilnadu, India since 2016. Her main research interest is in the area of Algorithms, Database Systems, Big Data Analytics and Data Mining. She has 16 years of teaching experience and 8 years of Research Experience.

Ms. V.SANTHI pursed Master of Science from Bharathidasan University in 2005 and Master of Philosophy in year 2007. She did M.Tech from Manonmaniam Sundarnar University in 2009. She is currently working as Assistant Professor in Department of Computer Application, Bon Secours College for Women, Thanjavur, Tamilnadu, India. Her research interest is in the area of Steganography, Network Security, Data Mining and Big Data. She has 8 years of teaching experience and 4 years of Research Experience

